

第八屆
數位典藏與數位人文
國際研討會
論文集

*8th International Conference of
Digital Archives and Digital Humanities
Conference Proceedings*

8th International Conference of Digital Archives and Digital Humanities
第八屆數位典藏與數位人文國際研討會
Program Schedule 議程

Wednesday, November 29, 2017 第一天 11 月 29 日 (週三)

08:10-08:45	Registration 報到
08:45-09:00	Opening 開幕式 Jieh Hsiang 項潔, National Taiwan Univ. Shu-Heng Chen 陳樹衡, National Chengchi Univ.
09:00-10:00	Keynote Speech 專題演講 (1): The Psychology of Images: The Complex Relationship Between Digital Humanities and Visual Culture 圖像心理學：數位人文與視覺文化的繁複關係 Damian Schofield, the State University of New York at Oswego Moderator 主持人：Tsai-Yen Li 李蔡彥, National Chengchi University
10:00-10:20	Poster Introduction 海報發表宣傳 (1) Moderator 主持人：Jyi-Shane Liu 劉吉軒, National Chengchi University
10:20-10:35	Coffee Break 茶敘
10:35-12:15	Chinese Session 1 : Authorship and Statistical Analysis 著作權與統計分析 English Session 1 : Resource Construction 資源建構
12:15-13:25	Lunch 午餐時間
13:25-15:05	Chinese Session 2 : Social Network Analysis 社會網絡分析 English Session 2 : Context and Implication 脈絡與意涵
15:05-15:15	Coffee Break 茶敘
15:15- 16:55	Chinese Session 3 : Corpora and Text Analysis 語料庫和文本分析 English Session 3 : Panel: Eurasian Past Meets the Future: Digitalizing the Mongol Empire 論壇：歐亞史如何面對將來：數位化蒙古帝國

Thursday, November 30, 2017 第二天 11 月 30 日 (週四)

08:10-08:30	Registration 報到
08:30-10:10	Chinese Session 4 : Visual Presentation and Analytic Tool 視覺呈現和分析工具
	English Session 4 : Spatial and Visual Analysis 空間與視覺分析
10:10-10:25	Coffee Break 茶敘
10:25-11:25	<p>Keynote Speech 專題演講 (2): Towards Computational History: Databases and Agent-Based Simulations 計量史學：資料庫與行為者模式</p> <p>Siew Ann Cheong, Nanyang Technological University</p> <p>Moderator 主持人：Jui-Sung Yang 楊瑞松, National Chengchi University</p>
11:25- 11:50	<p>Poster Introduction Session 海報發表宣傳 (2) Moderator 主持人：Chih-Ming Chen 陳志銘, National Chengchi University</p>
11:50-13:20	Lunch 午餐時間
	Taiwanese Association for Digital Humanities Annual Meeting 台灣數位人文學會年會
13:20-15:00	Chinese Session 5 : System Design and Development 系統設計開發
	English Session 5 : Infrastructure and Practices 基礎設施與實作
15:00-15:15	Coffee Break 茶敘
15:15-16:55	Chinese Session 6 : Digital Culture 數位文化探討
	<p>English Session 6: Panel: Spatial Humanities Mapping in Taiwan 論壇：空間人文繪圖在臺灣</p>

Friday, December 1, 2017 第三天 12 月 1 日 (週五)

08:10-08:30	Registration 報到
08:30-10:10	Chinese Session 7 : Spatial and Temporal Analysis 時空分析
	English Session 7 : Information Extraction and Text Mining 資訊擷取與文本探勘
10:10-10:25	Coffee Break 茶敘
10:25-11:25	<p style="text-align: center;">Keynote Speech 專題演講 (3)</p> <p style="text-align: center;">Digital Humanities with Unison Pervasive Computing, Solving the Complex B (the Unbankable Billions): The Look of Past, Present, and Future 以數位人文與同步普及運算解決Complex B (流動貨幣): 過去、現在及未來</p> <p style="text-align: center;">Ted Huang, Yang Braxter Corporation</p> <p style="text-align: center;">Moderator 主持人: Shinn-Shyr Wang 王信實, National Chengchi University</p>
11:25-12:20	Poster Session 海報發表
12:20-13:20	Lunch 午餐時間
13:20-15:00	Chinese Session 8 : Digital Resource Management 資料庫建置管理
	English Session 8 : Corporal Analysis 語料庫分析
15:00-15:15	Coffee Break 茶敘
15:15-16:55	Chinese Session 9 : Subject Domain Application 藝術主題領域應用: 音樂、繪畫、文字
	<p style="text-align: center;">Chinese Session 10:</p> <p style="text-align: center;">Panel: The Digital Methods of the History of Metrical Patterns in Classical Chinese Poetry 論壇: 中國古代詩歌格律發展史的數位研究</p>

目錄 Contents

專題演講 Keynote Speech

The Psychology of Images: The Complex Relationship between Digital Humanities and Visual Culture (Damian Schofield)

圖像心理學：數位人文與視覺文化的繁複關係.....1

Towards Computational History: Databases and Agent-based Simulations (Siew Ann Cheong)

計量史學：資料庫與行為者模式.....3

Digital Humanities with Unison Pervasive Computing, Solving the Complex B (the Unbankable Billions): The Look of Past, Present, and Future (Ted Huang)

以數位人文與同步普及運算解決 Complex B (流動貨幣)：過去、現在及未來.....4

場次一：著作權與統計分析

Session 1: Authorship and Statistical Analysis

Digital Humanities for Authorship Attribution Problem of *Shi-Ji*

《史記》作者數位化研究初探 (邱詩雯)9

A Stylometric Study based on Quantitative Analysis of Chinese Function Words

漢語虛詞量化分析與作者寫作風格研究 (林欣穎、劉吉軒、李恭儀、李金瑛)11

Classification Statistics of Classical Poems in Humanities Research

傳統漢詩分類統計工具在人文研究上的應用 (許明珠)12

A Discussion on the Related Explanations of the Blacklist in Taiwan Association for Human-Rights: A Perspective of Digital Humanities Study

臺權會黑名單相關論述之探討：一個數位人文研究的視角 (薛化元、江子揚、蔡銘峰、李恭儀)15

場次二：資源建構

Session 2: Resource Construction

Collective Memory in the Digital Age: A Case of Beijing Drifters (Min Ren, Sherry Xie, Jin Zhang)

數位時代的集體記憶：以北京川流為例.....18

Research on the Web Design for Disseminating Culture Information on an Ethnic Minority (Guliziba Paierhati, Watanabe Hidenori)

少數族群文化資訊散布之互聯網設計之研究.....22

Digital Ricoeur: A Digital Research Portal (Ming Yeung Cheung)

數位呂格爾：數位研究之入徑.....23

場次三：社會網絡分析

Session 3: Social Network Analysis

Digital Humanities Study on the Daoist Method of *Fahai yizhu* and *Daofa huiyuan* Network Analysis for Daoist General in Song to Ming Dynasty

- 《法海遺珠》、《道法會元》中道法系統的數位人文研究：宋、元、明道教官將網絡分析（高振宏、鄭有容）.....27
- Development of a Text Retrieval and Mining System for Taiwanese Historical People
臺灣歷史人物文本檢索與探勘系統之建置（謝順宏、柯皓仁、張素玟）.....30
- Journal, Gender and Mobilization: Social Network Analysis of *Women's Resonance* (1929-1944)
期刊、性別與動員：《婦女共鳴》社會網絡分析（1929-1944）（葉韋君）.....31
- From CBDB to TBDB: *The Treatise of Historical Figures of Zhang-Hua Gazetteer* as a Touchstone
從CBDB到TBDB：以《新修彰化縣志·人物志》為試金石（張素玟、李宗翰、柯皓仁、李毓嵐）.....32

場次四：脈絡與意涵

Session 4: Context and Implication

CrossAsia: Services and Resources for Digital Humanities in Asian Studies (Seeger Miriam)

- 橫跨亞洲：亞洲研究中之數位人文服務與資源.....35
- A Critical Approach to the DH: The Digital Influence on the Subjective Sphere (Casani Paolo)
批判數位人文之路徑：主體性之數位影響力.....38
- Humanity, *Quo Vadis?* (Graydon Wetzler)
人文學，何去何從？.....46
- Embodiment, Becoming-animal, Nomadic Subject on Subject Issues of Rosi Braidotti's Situated Posthuman Ethics (Yan-Ping Gao)
具身化、生成動物與游牧主體：羅西·布拉伊多蒂情境化後人類倫理學中的主體問題.....49

場次五：語料庫和文本分析

Session 5: Corpora and Text Analysis

Text Analysis of Landscapes and Events in Qing Dynasty in Taiwan

- 台灣清治時期地景與事件的文本探勘（林淑慧）.....50
- Women's Studies in Taiwan: A Corpus Analytical Study in Taiwan Academic Institutional Repository
應用語料庫分析取徑探討臺灣女性研究主題：以臺灣學術機構典藏為例（張映涵）.....52
- Examining the Public Attitude toward the Implementation of Five-day Workweek Policy: A Corpus Linguistics Analysis of Four Major Online Newspapers in 2015-2017 Taiwan
以語料庫分析取徑探究社會對《一例一休》新法案之態度：以2015-2017年台灣四大報紙電子新聞為例（吳孟家、簡翊淇、葉克芸、陳品臻）.....55

From Authoritarianism to Democracy: Text Mining of Fourteen Inaugural Addresses of R.O.C.

Presidents

威權到民主：歷屆總統就職演說的文字探勘（鄧志松、周嘉辰、李宥霆）.....57

場次六：論壇：歐亞史如何面對將來：數位化蒙古帝國

Session 6: Panel: Eurasian Past meets the Future: Digitalizing the Mongol Empire.....58

Reconstruction the Commercial Networks of the Ilkhanate with Digital Database (Matanya Gill)

以數位資料庫重構伊兒汗國商業網絡.....59

The War between the Two Capitals Revisited (Vered Shurany)

再探兩類資本之抗戰.....60

Genealogy between History and Text: The Digital Humanities' Quest in Researching Multi-

Generational Chinggisid intermarriages in Mongol Eurasia (Qiao Yang)

歷史與文本之系譜學：數位人文研究探問元朝蒙古帝國之跨代聯姻.....61

Experts, Institutions, and Networks: An Uighur Astronomer in the Yuan Court (Ishayahu Landa)

專家、機構和網絡：元朝的維吾爾星象學家.....62

場次七：視覺呈現和分析工具

Session 7: Visual Presentation and Analytic Tool

Engaging Virtual Interactive Navigation Based on HMD and Mobile Platform: an Example of Dadaocheng Historical District

基於頭戴式裝置與行動裝置平台之虛實互動式導覽：以花蓮牛犁社區為例（周子凱、詹進發）.....63

Development of an Analysis Tool to Facilitate Check-in Research in Social Media

輔助社群媒體打卡研究之分析工具研發（李蔡彥、梁芷瑄、吳筱玫）.....64

The Application of Natural Language Processing Technology in Digital Humanities

自然語言處理技術在數字人文中研究中的應用（王東坡、葉文豪、梁繼文）.....65

The Meta- and Theoretical Meaning of Automatic Chinese Segmentation: A Pragmatism Perspective

中文斷詞器的後設與理論意涵：語用學與社群的觀點（劉慧雯）.....66

場次八：空間與視覺分析

Session 8: Spatial and Visual Analysis

History of Governmental Funding in Shenzhen: Analyzing the R&D Structure and Effects through the Spatial Analysis of Metadata of Research Literature (Yawen Zou)

深圳政府經費的歷史：通過研究文獻元資料的空間分析來分析研發結構和效果.....69

Walking into a Brave New World: Visualizing Character Paths through Shakespeare's Plays (Brian Kokenparger)

走入美麗新世界：莎士比亞戲劇中的視覺化人物路徑.....72

Revelation of Prehistoric Architecture through Augmented Reality: Representation of the Sanctuary at Tas-Silg (Shih-Chueh Kao, Vince Briffa)	
透過虛擬實境揭露史前建築：以Tas-Silg聖所的再現為例.....	73
Research on the Teaching and Learning of Using 3D Motion Capture Analysis on Chinese Opera Performance (Tai-Jui Wang)	
中國戲曲運用3D動作捕捉分析之教學研究.....	76

場次九：系統設計開發

Session 9: System Design and Development

Apply Network Analysis in the Teaching of Traditional Chinese Phonology	
網絡分析方法在音韻學教學中的應用：以《廣韻》反切系聯為例（胡佳佳）.....	78
Constructing a Full-text Database of “Pronunciation and Meaning of All the Scriptures (<i>Yiqiejing Yinyi</i>)” : A Collection of Ten Ancient Manuscripts in Japan	
《一切經音義》全文檢索資料庫的構建：以日本所藏十種古寫本為例（劉冠偉、李乃琦）...79	
Gaining Support from Non-eligible voters: Tactics of Impression Management on Social Media by Hong Kong’s Chief Executive Election Candidate	
爭取無票者的民意：香港特首候選人的社群印象管理戰術（江建聰、黃珮蓁、傅威）.....	80
Constructing a Digitization System for Chinese Classics with Various Annotations and Notes: A Case Study of “Discourse on the Perfection of Consciousness-only (<i>Cheng Weishi Lun</i>)” Edition	
經典與注疏之數位化系統設計與建置：以「《成唯識論》及其注疏編撰」為例（洪振洲、蔡伯郎）.....	82

場次十：基礎設施與實作

Session 10: Infrastructure and Practices

Institutional Infrastructure for Fast Project Development, Research Tools Deployment and Research Outputs Integration of Digital Research Lifecycle (Hou Ieong Ho, Florian Kräutli, Esther Chen)	
提供快速專案開發、數位工具應用、與研究成果整合之機構基礎設施.....	85
Aceh Tsunami Recurrence from Historical Records and Local Interviews (Nurjanah Jane, Inoue Hiroki, Hidenori Watanave)	
從歷史紀錄與地方訪談再現印尼亞齊海嘯.....	86
The Library as a Medium: Strategies for Tackling Challenges of Digital Humanities (Feng-En Tu, Sharon Yang)	
圖書館作為媒介：因應數位人文挑戰之策略.....	88
Public Practice of Anthropological Collections (archives) and its Digitalization in Museums of Taiwan (Hui-Tuan Chang)	
國立臺灣博物館人類學典藏之數位化與公共實踐.....	89

場次十一：數位文化探討

Session 11: Digital Culture

History, Current Status and Future Development of Sport Informatics

體育情報學發展歷程、現狀及發展前景展望（張鵬飛）.....92

A Preliminary Study on the Intercultural Communication of Subculture: A Case Study of Ben Hedges Makes Friends with Taiwanese Web Users by Empresses in the Palace

淺談次文化的跨文化：從郝毅博藉《後宮甄嬛傳》打入台灣鄉民圈來看（黃南穎）.....93

Big Data and Civil Society: From the Perspective of Digital Privacy

巨量資料與公民社會：從數位隱私談起（鐘孝宇）.....95

Humanoid Being and Digital Dignity: Do Digital Humanities Portend a Future Fate?

人形存有與數位尊嚴：數位人文要預示未來命運嗎？（蔡瑞霖）.....98

場次十二：論壇：空間人文繪圖在臺灣

Session 12: Panel: Spatial Humanities Mapping in Taiwan.....99

Spatial Humanities: An Integrated Approach to Spatiotemporal Research (David Blundell, Ching-Chih Lin, James Morris)

空間人文學：整合時間與空間之研究取徑.....102

Citizen Science and Volunteered Geographic Information (VGI) (Jihn-Fa Jan)

公眾科學與自願性地理資訊.....103

Expressing Dynamic Maps through 17th-Century Taiwan Dutch Manuscripts (Ann Heylen)

以十七世紀荷蘭臺灣手稿展示動態地圖.....104

Simultaneous Invention or Propagation of Cultural Practices? The Tang-Hao on Tombstones of Taiwan and Penghu (Oliver Streiter)

台灣與澎湖墓碑上的堂號：同時生成抑或文化的傳播？.....105

場次十三：時空分析

Session 13: Spatial and Temporal Analysis

The Digital Humanities Research on The method of division of historical periods : Focusing on the "Shenzhen" News Report in People's Daily

歷史分期的數位人文研究：以《人民日報》(1946-2016)中深圳主題報導為例（邱偉雲、梁穎誼）.....107

The Dynamic Presentation Design of the Multi-source in History and Culture-A Case Study of the Stone Culture of Penghu

歷史文化多源（元）脈絡之動態呈現設計：以澎湖石滬文化為例（吳如娟、陳至柔、王瑞琪、傅詩漪）.....109

Spatial Distribution and Displacement of Tang and Song Poetic Activities A Data Analysis based on Chronological Maps of Tang and Song Literature

唐宋詩歌活動空間的分佈與位移：基於《唐宋文學編年地圖》的數據分析（王兆鵬）.....110

場次十四：資訊擷取與文本探勘

Session 14: Information Extraction and Text Mining

DH Tools to Capture and Predict Literary Reading at City-Scale (John Shanahan)

以數位人文工具進行城市層級閱讀能力的蒐集與預測.....112

Quantitative Stylometry of Alice Bradley Sheldon: How Well Did Her Pseudonyms Hide Her True Identity? (Miki Kimura)

Alice Bradley Sheldon 計量文體分析：筆名真能隱藏其真實身份嗎？.....115

Digital Humanities Exploration—Text Mining on Modern Political Discourse and Chinese Classics (Hsaunlei Shao, Yu-Hsuan Tiffany Wang)

數位人文探索：當代政治論述與中文經典之文本探勘.....116

Taiwan Biographical Ontology (TBIO): A Working Annotation Scheme for Biographical

Information Extraction and its Application to Sociology of Literature (Alvin Cheng-Hsien Chen, Táňa Dluhošová)

臺灣傳記本體 (TBIO)：一個開發中可擷取傳記資訊之標註系統以及其在文學社會學之應用

.....117

場次十五：資料庫建置管理

Session 15: Digital Resource Management

Trend Analysis of Digital Archives Management in Big Data Era: Based on 5W1H method of Content Management

大數據環境下數字檔案管理趨向探析：基於以內容管理為核心的六何分析法 (郭先敏) ...118

Steps and Methods of the Transformation of Bibliographic Literature to Analytical Data Taking the 20th Century Tang Literature Research Retrieval System as an Example

目錄型文獻轉化為分析型數據的步驟與方法：以《20世紀唐代文學研究論著目錄檢索系統》為例 (邵大為)119

Study of Formation and Transformation of 'Individual' in Modern China based on Unsupervised Learning Analysis

中國近代「個人」觀念的形成與演變：以非監督式學習法為主 (鄭文惠、梁穎誼)121

A Preliminary Research on Digital Archive System: A Case Study of the National Central Library

數位典藏系統初探：以國家圖書館為例 (魏家惠、洪偉翔、鄭基田)123

場次十六：語料庫分析

Session 16: Corporal Analysis

Near Synonymous Body-Part Terms in Chinese: A Corpus-Based Study (Hsiao-Ling Hsu, Huei-Ling Lai, Jyi-Shane Liu)

中文的身體部位相近同義詞：一個以語料庫為基礎的研究.....124

Baidu News Information Flow and Return Volatility: Evidence for the Sequential Information Arrival Hypothesis (Dehua Shen, Xiao Li, Wei Zhang)	
百度新聞之資訊流與收益波動：作為資訊連續到達假說的證據.....	135
William Faulkner's Fictional Demographics: Mapping People and Places Using the Digital Yoknapatawpha Database (Johannes Burgers)	
William Faulkner小說人物統計分析：使用Digital Yoknapatawpha數位資料庫繪製人物與地點.....	136
Collocation Analysis in News Corpora and its Implication to News Values (Huei-Ling Lai, Chao-Lin Liu)	
新聞語料庫中的搭配詞分析及其在新聞價值上的意涵.....	143

場次十七：藝術主題領域應用：音樂、繪畫、文字

Session 17: Subject Domain Application

The Introduction and Influence of Western "Landscape Painting" in 1900-1924	
1900-1924年西方「風景畫」的傳入與影響（彭卿）.....	146
Digital Humanistic Analysis of Catalogue and the Theme: Charles Lang Freer's Chinese Painting Imagination	
圖錄與畫題的數位人文分析：弗里爾先生的中國畫想像（徐鶯、邱偉雲）.....	147
Using the Concept of Genome to Explore Modern Chinese: A Quantitative Study of Writing Style based on <i>Dream of Red Chamber</i> , <i>Anhui Suhuabao</i> , and <i>New Youth</i>	
現代白話文「基因組」初探：《紅樓夢》、《安徽俗話報》和《新青年》語言風格統計研究（鄭文惠、余清祥、何立行）.....	149
A Digital Humanities Research on the Music in <i>Quan Tang-Shi</i> : Focusing on String Instruments	
《全唐詩》音樂詩的數位人文研究：以弦樂器為主（鄭文惠、顏靜馨、梁穎誼）.....	151

場次十八：論壇：中國古代詩歌格律發展史的數位研究

Session 18: Panel: The Digital Methods of the History of Metrical Patterns in Classical Chinese Poetry.....

A Computer-Aided Study on the History of Metrical Patterns in Classical Chinese Poetry	
使用計算機工具軟件輔助研究中國古代詩歌格律演變史的案例和心得（杜曉勤）.....	154
The Development of the Analytical System: A Case Study of Digital Methods for Research of Chinese Poetry Metrical Patterns	
漢詩格律分析系統的開發：以格律研究為導向的數位工具案例（鄧賢瑛、王昱鈞）.....	155
A Preliminary Inquiry into the Ground Truth of Metrical Pattern in Classical Chinese Poetry	
中國古代詩歌格律的「自然合格率」理論初探（羅珮瑄）.....	157
The Concepts of the Development of Metrical Patterns in Classical Chinese Poetry under the Interaction between Digital Methods and the Humanities: From Early Tang to Early Modern	
數位方法模塑出的漢詩格律發展史觀：以初唐詩學為例（蔡瑜）.....	158

海報發表宣傳 Poster Introduction Session

Exploring the New Path of Optimizing the Emotional Experience of Digital Archives Users from Archival TV Programs

從檔案類電視節目探索優化數位典藏用戶情感體驗的新路徑（王露露、徐擁軍）.....161

An Individual-Use Text-based Historical Event Analysis Tool and its Application in Studying Ming Pirates Events

個人歷史文本事件分析工具及其於明代倭寇研究之應用案例（白璧玲、吳承翰、范毅軍、蔡融易、蔡宗翰）.....163

Development of a Wearable System for Evaluation of Meditation

評估冥想之攜帶式系統演進（梁珮琳、郭博昭）.....166

Customer Choices, Charging Policy and Social Welfare Analysis in a Duopoly Music Industry

多元定價，廣告策略與社會福利之分析：以雙占音樂平台為例（林怡芯）.....169

DocuSky: a DH Platform in the Making

DocuSky：一建構中的數位人文平台（杜協昌、林農堯、洪一梅、項潔）.....170

數位人文在清華大學之思行及兼述清大「圖書館即時偵測服務環境」系統之研發初探（楊叔卿、林怡瑄、林宏擘）.....171

The Practice and Revelation of the British Intangible Cultural Heritage Digital Humanities Projects: With art-humanities.net as the research core

英國非物質文化遺產類數字人文項目的實踐和啟示—以art-humanities.net 網站為調研核心（林佳萍）.....176

以單字切割與水平排列方法提升中文古籍影像的文字辨識能力（吳昱賢、王祥安、蘇冠宇）...178

Co-Word Analysis Tool Applied in Political Platform

共現字分析工具於競選政見應用之初探（戴浩光）.....179

Development and Evaluation of an Automatic Text Annotation System for Supporting Digital Humanities Research

支援數位人文研究之文本自動標註系統發展與使用評估研究（陳志銘、蔡明月、劉鎮宇）...180

The Psychology of Images: The Complex Relationship Between Digital Humanities and Visual Culture

Damian Schofield *

Abstract

An intrinsic connection exists between humans and the memories they create; they define who we are, where we came from and our accomplishments and failures. However, decades of research has shown how fragile human memory can be. The early linguistic experiments of Elizabeth Loftus demonstrated how through misinformation and suggestibility we could influence and change the memory of others¹. More recent work has led to many theories regarding behaviour, and many theories and guidelines are now available to show how human decision making can be influenced by external stimuli^{2,3}. A large volume of research output exists in this field, but the majority of the research work has focused on language (oral and textual) and its ability to influence readers and listeners.

The digital age has brought along an increased multimediality of communication. This concerns the source data in the field of humanities as well as the ways to perform research and represent the results. The creative nature of digital humanities allows researchers to explore, discover and develop new possibilities for data analysis. Visual representation of data significantly enhances the interpretation potential of the artistic, literary, musical and historical corpora. Visualisation methods developed in other disciplines (such as GIS mapping, graphs, charts and computer graphics based representations) have already been used in some branches of the humanities. However, there is a need to reconsider the assumptions that underlie the use of these visualisations

* Director of HCI, Department of Computer Science, State University of New York, Oswego, New York, USA.

1 Loftus, G. R., and Loftus, E. F. (1974). The influence of one memory retrieval on a subsequent memory retrieval. *Memory and Cognition*, 2(3), 467-471.

2 Young, W., Davis, M., McNeill, I. M., Malhotra, B., Russell, S., Unsworth, K., and Clegg, C. W. (2015). Changing behaviour: successful environmental programmes in the workplace. *Business Strategy and the Environment*, 24(8), 689-703.

3 Davis, R., Campbell, R., Hildon, Z., Hobbs, L., and Michie, S. (2015). Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health Psychology Review*, 9(3), 323-344

based on our ever improving understanding of human psychology, cognition and perception.

As technologies develop, we are seeing the emergence of the use of multiple modalities in our everyday interfaces (often based on natural language processing technology). However, almost all interaction with digital humanities based visualisations rely on vision as the primary means of passing information to the user⁴. Visual based interfaces have been used on digital devices and displays for many years and their use may be seen as ‘every day’ and having little impact on the viewer.

However, it is perhaps worth considering that this specific form of visual media interaction requires special care and attention due to its inherently persuasive nature, and the undue reliance that the viewer may place on information presented through a (potentially photorealistic) visualisation medium. There are a number of fundamental implications inherent in the analysis of visual interfaces over textual/oral mediums. Their influence on human memory and behaviour cannot be underestimated. It is important that a rigorous investigation and analysis of all the facets of these interfaces is undertaken.

This keynote speech will introduce research undertaken by the author over the past twenty years that has experimented with, and examined a range of visual based presentation technology. Digital humanities based visual presentation systems (including interactive displays, computer generated graphical presentations, animated graphics and immersive virtual environment technology) have already been used in many innovative contexts⁵. This talk will illustrate research undertaken to assess the effect of visual technology on viewers (in particular their memory and decision making abilities) and describes some of the issues raised by the experimental results. The talk will connect psychological research with human cognitive and perceptual processes, to allow the evaluation and optimisation of digital humanities visual interfaces. The talk will conclude with a discussion of the potential benefits and problems of designing interactive digital humanities displays when considering the impact on human cognition.

⁴ Powers, D. (2006). Vision in HCI: Embodiment, Multimodality and Information Capacity, Proceedings of the Workshop on the Use of Vision in HCI (VisHCI 2006), Canberra, Australia. Vol. 56. Australian Computer Society, Inc.

⁵ Schofield, D. and Mason, S. (2012), Using graphical technology to present evidence, in: Mason, S. (Ed.), Electronic Evidence, 3rd edn, LexisNexis Butterworths, pp. 217 – 253.

Towards Computational History: Databases and Agent-Based Simulations

Siew Ann Cheong*

Abstract

Reloading the treasure of human experiences from the perspectives of artificial intelligence and computation is the next frontier of the humanities. With a crucial focus on primary sources provenance and validation issues, history should be at the forefront of this revolution, but it is somehow the Cinderella of computational humanities. In the first part of this talk, we will present the reasons for this impasse and propose solutions. In the second part of this talk, we will share our vision of how databases can be the next generation of (machine readable) critical editions of primary historical sources. While it should be possible to apply machine-learning techniques to these databases to generate scientific hypotheses that historians can test, we aim to go one step further. In the final part of this talk, we explain how agent-based models (ABMs) can help in the historian's perennial quest for truth, by potentially explaining what happened in the past, predicting what can happen in the future, and exploring how the present could be different had things turn out differently in the past. Drawing examples from our own work, we will describe the challenges we have encountered thus far, and the tentative solutions we have adopted to move this ambitious research program forward.

* Associate Professor at Division of Physics & Applied Physics, School of Physical & Mathematical Sciences, College of Science, Nanyang Technological University.

Digital Humanities with Unison Pervasive Computing, Solving the Complex B (the Unbankable Billions): The Look of Past, Present, and Future

Ted Huang*

Abstract

The increasing popularity of mobile computing devices has allowed for new research and application areas. Specifically, urban areas exhibit an elevated concentration of such devices enabling potential operation and sharing of resources among citizens.

People, architecture and technology together provide the infrastructure for these applications and implementation of such infrastructure is important for effective design and development. 1950s brought us credit cards to ease the burden of carrying cash and 1960s ATMs to replace tellers and branches. In the 1970s, electronic stock trading began on exchange trading floors and in 1980s we saw the rise of bank mainframe computers and more sophisticated data record-keeping systems but in the 1990s, the Internet and e-commerce business models flourished. As of the result with the introduction of online stock brokerage websites aimed at retail investors.

These five decades of developments have created a financial technology infrastructure which most people never think about, but use almost every day. Fintech in the last 50 years developed the necessary liquidity into the market. Sophisticated risk management, trade processing, treasury management and data analysis tools at the institutional level for banks and financial services firms. While these systems are not apparent to retail banking customers, they make up a multibillion industry aimed at supporting the needs of the financial services sector. Now, in the 21st century, retail financial services are being further digitized via mobile wallets, payment apps, robo-advisors for wealth and retirement planning, equity crowd funding platforms for access to private and alternative investment opportunities and online lending platforms. These fintech services are not simple enhancements to banking services, but rather replacing banking services completely. So, fintech can be thought of in two broad categories, consumer-facing and institutional. It is these consumer-facing fintech services which are quickly gaining customers and competing with banks.

* Board at Yang Braxter, provider of Fintech solutions to the financial technology industry, USA.

Despite initial hurdles posed by financial regulations, fintech has revolutionized the financial world and will continue to do so long into the foreseeable future. Technology's role is to simplify lives and streamline daily operations—and that's exactly where fintech excels and delivers. And the understanding of Digital Humanities has now gone beyond simple data mining. With Fintech developing new methodologies, and to motivate both reflections on and paradigm shift of humanities with Unison platform.

Unison platform with pervasive computing to solve the most complex issue (The Unbankable Billions)

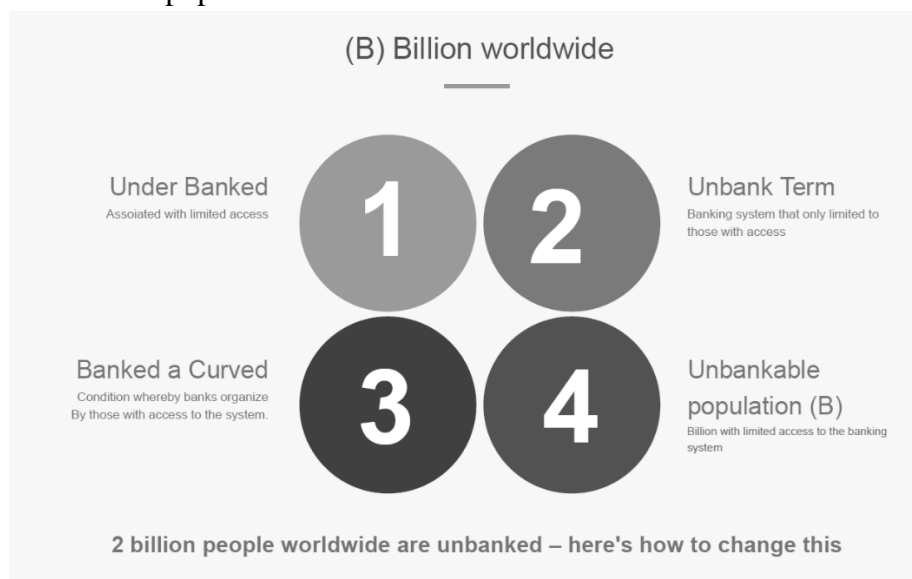
Most of the world's 2.5 billion unbanked adults live in developing economies. According to the most recent Global Fintech data, 89 percent of adults in high-income economies report having an account at a formal financial institution; in developing economies, 41 percent of adults do.

Regionally, according to this same database, the rate of unbanked adults is highest in the Middle East and North Africa, at four out of every five adults, followed by Sub-Saharan Africa and South Asia. In several developing economies, more than 95 percent of adults do not have an account at a formal financial institution.

Disparities also exist by gender. Among adults in developing economies living below the \$2-a-day poverty line, women are 28 percent less likely than men to have an account at a formal financial institution. Indeed, there is a persistent gender gap of six to nine percentage points across income groups within developing economies.

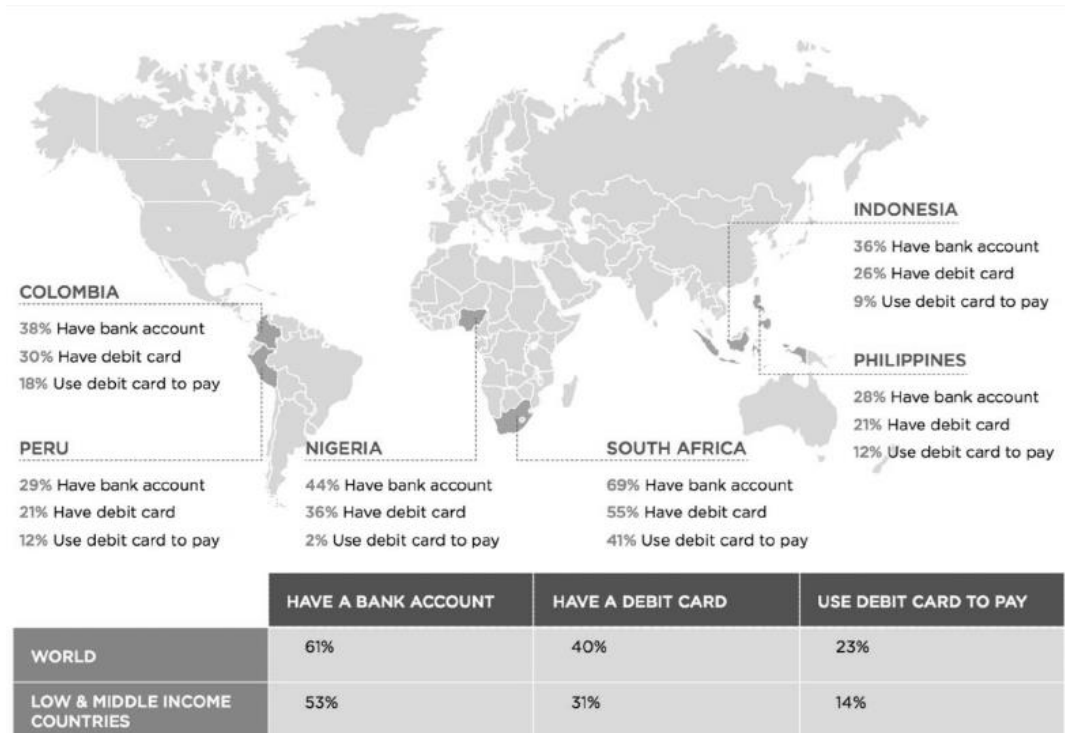
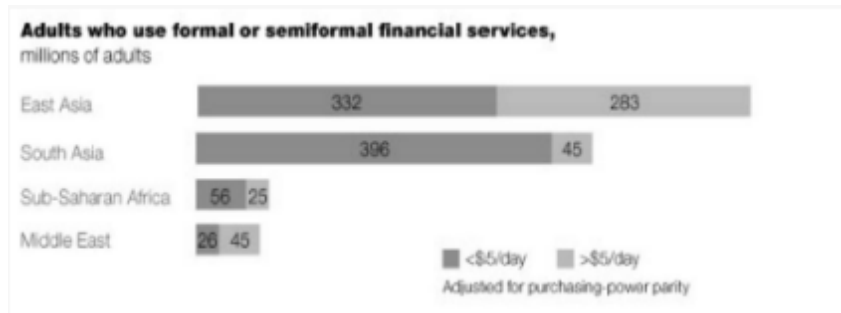
Four Important Facts: often known as a unbankable group.

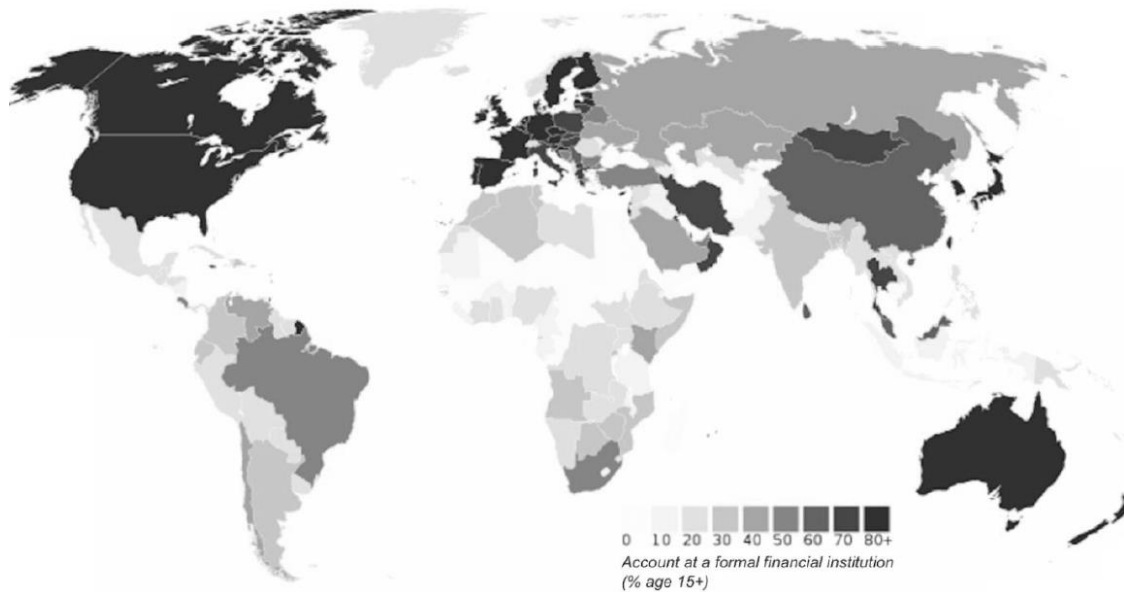
1. Under Banking
2. Unbanked term
3. Bank A Curved
4. Unbanked population



2.5 billion of the world’s adults unserved doesn’t mean unservable.

The world’s adults don’t use formal banks or semiformal micro finance institutions to save or borrow money, our research finds. Nearly 2.2 billion of these unserved adults live in Africa, Asia, Latin America, and the Middle East. Unserved, however, does not mean unservable. The micro finance movement, for example, has long helped expand credit use among the world’s poor—reaching more than 150 million clients in 2008 alone. Similarly, we find that of the approximately 1.2 billion adults in Africa, Asia, and the Middle East who use formal or semiformal credit or savings products, about 800 million live on less than \$5 a day. Large unserved populations represent opportunities for institutions that are able to offer an innovative range of high-quality, affordable financial products and services. Moreover, with the right financial education and support to make good choices, lower-income consumers will benefit from credit, savings, insurance, and payments products that help them invest in economic opportunities, better manage their money, reduce risks, and plan for the future.





Global Initiatives with Unison Platform

Global financial institutions invest significant resources into closing tremendous gap in access to banking. Mastercard and Visa, along with other financial institutions, shaped a coalition of partners last year to reach universal financial access by 2020.

The coalition will focus efforts on 25 countries (mostly African), where 73% of all financially excluded people live. India and China have the largest share of unbanked people. Together, they account for some 32% of them. The rest of the top-priority countries include Bangladesh, Brazil, Colombia, Cote d’Ivoire, DRC, Egypt, Ethiopia, Indonesia, Kenya, Mexico, Morocco, Mozambique, Myanmar, Nigeria, Pakistan, Peru, the Philippines, Rwanda, South Africa, Vietnam, Tanzania, Turkey and Zambia.

Unison platform with pervasive computing to solve the most complex issue (The Unbankable Billions)

Pervasive computing, is a growing trend of embedding computational capability into everyday objects to make them effectively communicate and perform useful tasks in a way that minimizes the end user's need to interact with computers as computers. Pervasive computing devices are network-connected and constantly available. The goal of Unison platform using pervasive computing to make devices smart to create a sensor network capable of collecting, processing and sending data, and communicating as a means to adapt to the data's context and activity; in essence, a network that can understand its surroundings and improve the human experience and quality of life. Thus provide the unbankable billions into the next frontier of financial needs.

✧ Understanding of Digital Humanities has now gone beyond simple data mining. With Fintech developing new methodologies with Unison platform.

the key thing is to provide tools that allow quick and easy access for security and longer term planning, preventing unnecessary spending, and emphasizing the security and certainty that banking services can provide.

Unison 4.0, let your most key access available at all time, any time.



✧ Very poor people in emerging economies not only have a surprising degree of interest in financial services but also, when possible, use them enthusiastically.

when it comes to assist is to provide mobile options beyond basic needs. Is to offer ways to improve usability and functionality:

- Creating a mobile userexperience more similar to interface.
- Leveraging advanced biometric security systems to increase customers' comfort with activities such as paying bills.
- Allowing greater flexibility in policies, such as higher limits or currency conversions;
- Creating a superior omnichannelexperience by integrating click-to-call, click-to-chat and video banking options.
- Online account openings should include a personal component

Objective is to be data-centric strategy, an opportunity for differentiation in the marketplace but also a powerful driver of customer engagement."

《史記》作者數位化研究初探：以三十世家為例

邱詩雯*

摘 要

《史記》被譽為「史家之絕唱，無韻之離騷」，兼具史學、文學價值。根據〈太史公自序〉，知其成書主要歷經司馬談、司馬遷父子二人之手。經司馬遷外孫公布之後，幾經流傳，今日所見《史記》，共五十二萬餘言，除司馬談、司馬遷之外，亦包含褚少孫的續補。褚少孫續補文字，在段落之前，皆明示「褚先生曰」，與司馬談、司馬遷父子文字明顯有別。然而除褚少孫文字外，由於司馬遷是繼承其父司馬談太史公的職志，根據司馬談部分舊稿，完成《史記》，因此《史記》中司馬談、司馬遷二人所寫文字，除〈自序〉中的「論六家要旨」一段，其餘並無標示，難以區別究竟何為司馬談所述，何為司馬遷所寫。

歷來《史記》研究學者亦曾針對《史記》作者，進行一番探討。最早如清方苞〈書史記十表後〉、〈又書太史公自序後〉，曾考論司馬談作十二本紀及〈十二諸侯年表〉、〈六國年表〉、〈秦楚之際月表〉、〈惠景間侯者年表〉等篇。又如王國維〈太史公行年考〉亦有論及列傳篇章。而顧頡剛〈司馬談作史〉及李長之〈史記中可能出自司馬談手筆者〉，更是大篇幅的討論《史記》作者問題。近來還有賴長揚〈司馬談作史補證〉、趙生群〈司馬談作史考〉等，亦關注此論題。據張大可統計，上述考論，認為今本《史記》中司馬談作史篇目共有 37 篇，包括本紀 12 篇、表 4 篇、書 4 篇、世家 8 篇以及列傳 9 篇。

然而，前人的研究成果，多以傳統的文獻研究法為主，從《史記》文字稱引的內部證據，以及司馬遷行年的外部證據，進行篇章作者的分析。然而，不同的作者，儘管父子家業傳承，在書寫文字使用的詞彙頻率，應當有別。因此，如若將前輩學者懷疑司馬談所做篇章，以數位人文詞頻統計方法，進行歸納，應該可以透過篇章間使用文字詞彙風格的區別，繼而照出司馬談、司馬遷二人所作篇章的證據。

本研究嘗試運用 DocuSky 平臺中「文本詞彙頻率統計工具」，以《史記》三十篇世家文本為對象，探討學界推測〈齊太公世家〉、〈魯周公世家〉、〈管蔡世家〉、〈陳杞世家〉、〈衛康叔世家〉、〈宋微子世家〉、〈晉世家〉、〈趙世家〉八篇，與其他二十二篇世家文字使用詞頻的異同。先參考《史記》學界虛詞研究成果，選出量化標的物，再就其分布進行統計檢定。再以褚少孫續補段落文字作為覆核的對照組，比較司馬談、司馬遷、褚少孫三人運用虛詞字頻是否有所差異。則透過此

* 國立成功大學中國文學系助理教授，Email: sw.chyu@gmail.com

數位人文工具的輔助，能迅速從事文本詞彙頻率的統計，較為客觀佐證《史記》世家篇章作者的異同。日後能再就世家的研究成果，推而廣之，擴及《史記》全部篇章，以數位人文研究方法，補強當代《史記》學研究成果。

關鍵字：史記、司馬談、司馬遷、文本詞彙頻率統計、數位人文

漢語虛詞量化分析與作者寫作風格研究

林欣穎*、劉吉軒**、李恭儀***、李金瑛****

摘 要

透過語言特徵的計量來進行風格辨別的研究，是近來語言風格學發展的趨勢。語言特徵的計量包含高頻詞、詞類、詞組、句子長短、段落長短、標點符號和虛詞等。虛詞作為一種以語法功能為主的詞彙，可以提供跳脫於文章主題、文體限制之外的視角，是探討作者風格的重要工具。本文首先以中研院平衡語料庫 4.0 版為基準，呈現出現代漢語常用虛詞，討論符合平衡語料庫詞類標記與否的計量差異，以《紅樓夢》、張愛玲及錢鍾書的文本作為漢語虛詞的實驗文本，討論虛詞應用於文學風格研究的可能性，並以變異數、T 檢定等方式檢驗統計結果。

關鍵字：漢語虛詞、量化分析、語言風格學

* 國立政治大學資訊科學系研究助理，Email: singingcindy102151010@gmail.com

** 國立政治大學資訊科學系教授暨創新與創造力研究中心主任，Email: jyishane.liu@gmail.com

*** 國立政治大學資訊科學系研究所碩士生，Email: james88385542@gmail.com

**** 康寧大學應用外語科副教授，Email: cylec@g.ukn.edu.tw

傳統漢詩分類統計工具在人文研究上的應用

許明珠*

摘要

由國立臺灣文學館策劃，委託國立成功大學執行的「《全臺詩》蒐集、整理、編纂計劃」到目前為止已執行 15 年，蒐羅的範圍從 1650 年代鄭氏時期以降，到 1895 年日本統治臺灣之前出生的古典詩進行蒐集，目前已整理出版的詩人數量多達 860 多位，大抵依出生時序排列（部份詩人之出生年未可考），詩作字數約 1100 萬字，已經編輯出版 40 冊紙本的《全臺詩》。

《全臺詩》是國內最重要的古典文學史料，這份文獻橫跨明鄭、清至日治時期，有助於了解台灣早期漢人遷移來台、在台生根、外族殖民的生活史，是一批內容非常重要，形式也非常獨特的文獻。開發這批文獻各種被解讀的可能，也就成為臺灣文化領域學術研究一個可期待且必須前進的方向。

本人執行科技部整合型計畫「數位全臺詩：知識庫開發與人文數位工具應用對台灣古典詩主題研究的加值與創新」下之子計畫「數位全臺詩延伸應用」已進入第三年度。在「數位全臺詩資料庫」全文檢索的基礎架構上（網站名稱：台灣古典詩），增加「分類統計」功能。雖然這個功能數位技術難度並不高，卻是現下數位平台所未有。

該計畫為增益「數位全臺詩資料庫」的可應用性，在「數位全臺詩資料庫」架構下，延伸「詩詠的物類聯想」、「文人的愉悅詩史」主題。「詩詠的物類聯想」將詩人們引用入詩之日常生活器用之物類進行分類，以詩中之「物」為主題，築構一個個研究範疇，甚或是研究型數據專題，而「文人的休閒娛樂」則是延續「詩詠的物類聯想」，將相關關鍵字詞歸入「物類」的檢索綱目之中，將相關「物件」（名詞）提取出來，分類整理，藉字詞使用頻率的計數及文本材料的兩個屬性去回應人文研究的問題：

- 一、詩歌作為一種史料
- 二、詩歌作為一種文學研究材料

就「詩歌作為一種史料」這層次來看，所有被歸類的「物」（字詞）可被納入時間、空間、人物等限制條件中，進行統計、分析，去看人類與其生產、使用或變更之物質的關係，可成為文學、符號學、語言學、人類學、社會學研究之寶藏。高桂惠曾提及「中文系的物質文化研究除了文本闡釋與感知／感性的掘發，對於文獻與圖書的文化史、禮俗研究，不僅可以測繪文化象徵、結參照體系，更

* 國立臺中科技大學助理教授，Email: mingchu@nutc.edu.tw

進一步深化傳統漢學的研究進路。」(〈物質文化研究在政治大學〉,《跨界對話——漢學、比較文學與物質文化研究》,2014)這一層次,有非常多的問題可以追問,決不限於中文系研究領域:詩人全部詩作與物品被寫及之數量比例、某詩人寫作某物在其寫作生涯曲線圖、某特定時期描寫某物品與全時期被寫作比例、某物品與更大類項被寫作的比例、某物品與他類物品被寫作比例、某物品在不同詩人詩作中的比例差異。如:某詩人以「竹」入詩的詩作數與其整體詩作比例?在台詩人以船入詩,以車入詩的比例?明清在台詩人以酒入詩,以茶入詩的比例?結合時間、地理資訊,功能將更為強大,甚可旁及物產、地景等資料。比如檳榔 88 首、糖 46 首、樟 23 首,檳榔被文人關注的次數大高過其他經濟物產,又如同是植物,竹 3045 首、梅 2015 首、松 1687 首,這數字差異究竟單純為文人寫作偏好、時代風尚,亦是經濟、地理因素,個中問題便值得深入探索,進而解答文化、社會、經濟、歷史學等問題,如探究「烏紗」的跨斷代發展,「烏紗」在傳統語境中,代表「官職」,未必指頭頂上的烏紗帽,進入日治時期,官員不再戴烏紗帽,「烏紗」還會入詩嗎?入詩的比例如何?在「分類統計」功能介面中,可分取現有明代及清初期(~1895 年)與 1875 年後出生詩人作品(詩人 20 歲)在明清詩人中為:31/8977,以「烏紗」入詩比例為 0.35%(共有 8977 首詩,31 首寫及烏紗),另取 1875 年後出生之詩人(詩人進入日治後已 20 歲)43/22885(共有 22885 首詩,43 首寫及烏紗),比例為 0.18%,與明清詩人寫入「烏紗」比例為 0.35%相較,比例差近一倍。可旁證時代改制改變已影響生活與書寫。

就「詩歌作為一種文學研究材料」,詩歌研究一個重要且核心的切面,就是「意象」。在中國詩歌傳統,「抒情傳統」是理解詩歌一個非常重要的切入點(甚至是唯一切入點),所有詩中出現之「象/物」,乃有「感」而發,因此,理解中國詩歌傳統就不能將之視為史料,切割其「情/意」而純粹論「物/象」。從這角度,也可留意鄭毓瑜所提出的反思:「眼前我們仍然缺乏由『物』的角度,而不只是『情』的先決優位,去重新討論與詮釋『抒情傳統』。」可知抒「情」傳統雖是中國詩歌之核心精神,「物」卻是詩歌創作、文學研究中不可忽缺的重要角色。鄭教授曾以黃遵憲為例,說明詩歌創作中,「物」象的去取,事實上飽含人與物關係的體現、個體在文化中的情感體現,「透過傳統感知與表述世界的模式,更容易對照出晚清古典詩人如何集體地體現了傳統『物』世界」。(〈類與物——古典詩文的「物」背景〉,《清華學報》41(1),2011,P.3-37)

數位全臺詩分類統計如何整理文獻資料、架構系統。其文獻整理的工作重點有三:

一、建立分類表:首先須針對《全臺詩》「詩詠的物類聯想」、「文人的休閒娛樂」此二面向建立資料表,規劃「器物」、「休閒活動」分類表(近 3000 細項)。

二、資料轉譯與校對:依據分類表,將詩作標記、歸類,將原有 word 格式之詩作資料,轉以 excel 格式呈現,於特定欄位填入關鍵之索引字詞。這需要詞夾子一類的數位工具協助之外(由子計畫三設計、開發),進而由校對人員進行人工一次校對、二次校對。

校對的過程中，除修正 excel 表格歸類之正誤，同時回過頭增修分類表的項目與字詞。工作時程參下節說明。

三、統計平台的建置：此「全臺詩分類統計平台」以「階層式分類」方式提供索引，以滑鼠點選，可將大類、中類、小類詞彙加進計算式中。

台灣古典詩分類統計系統是一個新的文史研究工具，對學術研究預期之貢獻有二：

(1) 本計畫橫跨歷史、文學與資訊工程等學科，結合各科專家學者，互助合作，針對台灣古典詩文本資料進行分析，以發現明鄭、清代、日治文人的詩歌語言中，所蘊涵的世界觀、社會文化與個人身體感等層面的文化意涵。在文史研究方面，可供做「物產經濟」、「消費文化」、「娛樂活動」、「意象經營」、「語言風格」、「生活美學」、「物質文明」等議題研究之工具，涉及文學、史學、美學、社會學、人類學、經濟學等層面。

(2) 以《全臺詩》這一筆 1200 萬字的資料，供給各學科皆可從中抉發議題與資料的統計分析平台，如若成功，這一技術與思維可納入「全唐詩」、「全宋詩(詞)」、「全清詩(詞)」，擴充其應用範圍。

關鍵字：數位全臺詩、臺灣古典詩、分類統計、物類

臺權會黑名單相關論述之探討：一個數位人文研究的視角

薛化元*、江子揚**、蔡銘峰***、李恭儀****

摘 要

1992年7月7日，立法院三讀通過國家安全法修正案，在原第3條第2款「有事實足認為有妨害國家安全或社會安定之重大嫌疑者」後加上「但曾於臺灣地區設籍，在民國三十八年以後未在大陸地區設籍，現居住於海外，而無事實足認為有恐怖或暴力之重大嫌疑者，不在此限」之但書，在設下該但書條件後，使得讓絕大多數在戰後出國留學、工作的海外異議人士得以獲准入境，「黑名單」的人數大幅限縮。1999年5月21日公布之「入出國及移民法」第5條規定「居住臺灣地區設有戶籍國民，自本法施行一年後，入出國不需申請許可」，自2000年5月20日起，國民入出境需先經申請的制度正式走入歷史。大法官於2003年4月18日做出釋字第558號解釋，指出國安法未區分國民是否於台灣地區設有住所而有戶籍，一律非經許可不得入境，違法憲法第23條規定之比例原則，「侵害國民得隨時返回本國之自由」而違憲。一般認為，海外「黑名單」制度至此才正式走入歷史。

回溯黑名單在台灣的歷史發展，《中華民國憲法》第10條保障人民有居住及遷徙之自由，然而，戰後台灣，卻有很長一段時間在動員戡亂及戒嚴體制之下，政府基於諸多政治、國家安全與社會安定的理由而限制國人入境或出境，所謂的「黑名單」係指在此一時空背景下，因被限制入境或出境的狀況形成。「黑名單」制度的存在與運作，是非常體制下對人民基本人權的侵害，也是強人威權統治時期的重要人權課題，而此一制度正式走入歷史，甚至延續到第一次政黨輪替之後。1949年2月21日，因國共內戰局勢影響，臺灣省主席陳誠公布「臺灣省准許入境軍公人員及旅客暫行辦法」，規定入境臺灣前須先經申請獲核准後持相關證明文件方能入境，且「絕對禁止無正當職業者入境」，無證入境者則予以遣返，開始針對本國人民進行常態性的入境管制。同年5月20日起臺灣全省戒嚴，戒嚴令中便明訂「無論出入境旅客，均應遵照本部（按：臺灣省警備總司令部）規定，辦理出入境手續，並受出入境之檢查」。後國民政府遷台，除持續對入境採取嚴格管制措施外，也開始緊縮國民出境的空間。1951年6月13日行政院

* 國立政治大學台史所，Email: hyh5595@gmail.com

** 國立政治大學雷震研究中心助理研究員，Email: wallace0510@hotmail.com

*** 國立政治大學資訊科學系副教授，Email: mftsai.mftsai@gmail.com

**** 國立政治大學資訊科學系碩士生，Email: james88385542@gmail.com

會議通過「戡亂時期臺灣省准許人民入境出境暫行辦法」(同年 6 月 15 日國防部公佈施行),對於申請出入境者規定必須提出「保證書」,保證人需保證被保證人「出境後,不得為匪工作」,被保證人「如有匪諜行為」,保證人應受連帶處分。至於入出境的管理機關則由臺灣省保安司令部(後由臺灣警備總司令部)負責,無疑將入境管制視同情治保安工作的一環,逐漸成為政府監控人民的手段。戰後臺灣的入出境管制不僅針對居住於臺灣的國人,旅居海外的國人同樣受到相當的限制,尤其是在海外從事臺灣獨立運動或其他不見容於執政當局的異議人士都面臨「有家歸不得」的處境,也就是被政府列為入境臺灣的「黑名單」。

「黑名單」的管制措施,主要是政府為了避免這些異議人士將其異議言論行動帶回臺灣,對於政府統治的正當性造成衝擊。「黑名單」的管制措施到了 1980 年中後期,隨著民主進步黨的成立、解除戒嚴、社會運動風起雲湧的背景下,開始在臺灣內外遭到強烈的質疑甚至實際行動的挑戰。在相關條文規範下,長期名列「黑名單」的海外異議人士,其返鄉之路仍是充滿不確定因素。海外台獨人士便發起返鄉運動,直接以闖關方式,以身試法正面迎戰政府的「黑名單」政策,而國內的社運團體及中央民意機構也透過各式聲援及質詢等方式要求政府檢討、解禁「黑名單」,而部分先前「非法」闖關回台的海外台獨人士,除被以違反國安法起訴外,也涉及內亂罪,直到懲治叛亂條例、刑法第 100 條陸續廢止及修正後,這些人才獲得釋放,其中,刺蔣案主角黃文雄則到 1996 年 5 月回台,至此,官方「黑名單」歸零。

台灣人權促進會(以下簡稱台權會)作為長期關注台灣人權事務且發展最久的民間組織,黑名單的形成與在其架構下人權所受到的侵害與約限,自然為其所關注的重點議題之一,特別是解除黑名單更是台權會長期以降所運作的重要課題。針對黑名單存在之時,台權會除屢屢透過發表聲明表示對黑名單的不滿外,更藉由街頭運動進一步標誌著黑名單的存在對於台灣人權情況的侵害,此外,亦透過如人權報告書等專文撰寫,對黑名單對台灣社會發展乃至更深入的影響傷害進行論述探討,其中,如獨台會、台獨聯盟等案皆是在台權會長期運作與交涉下,逐漸為世人所理解,也從中獲悉台灣在非常體制下,政府對異議人士有系統的壓迫與人權斷害。因此,透過對台權會相關文書對於黑名單的記載與討論,將有助於進一步理解當時國民黨政府對人權的侵害情況,而對黑名單課題的探討,對於戰後台灣歷史發展,特別是人權史脈絡,將更有其意涵與價值。

1949 年 5 月 20 日台灣實施戒嚴,戒嚴令明訂出入境旅客均應辦理出入境手續,本國人民必須比照外籍人士辦理簽證才能出入境,執政當局對主張台灣獨立、共產主義或否定中華民國憲法的異議人士,藉由簽證措施限制其入出境。1987 年 7 月 15 日解除戒嚴後,入出境規範改適用同日施行的國安法,仍採申請許可制,1992 年 7 月 7 日立院修正國安法第三條,將曾設籍台灣,後居住海外,而無事實足認為有恐怖或暴力之重大嫌疑者,規定不得禁止其入出境,黑名單人數才大幅縮小。2000 年 5 月 20 日國民入出境不再需要申請許可,人民遷徙自由始獲得真正的保障。台權會早期關注海外台灣人的返鄉權,聲援闖關被捕的黑名

單人士，同時要求政府應廢除違反人權的黑名單；之間亦論及「外省人」回中國探親的返鄉權，以及金馬人民來台後無法再回去故鄉的問題。台權會與國外的台灣人黑名單處理小組及國內的打破黑名單返鄉運動協調中心合作，揭發調查取得的黑名單之名單，戳破政府宣稱沒有黑名單的謊言，讓世人知道黑名單制度的不合理。在歷經聲援許信良、羅益世、李憲榮、蔡正隆、鄭自才、郭倍宏、李應元、陳榮芳與林明哲等世台會或台獨聯盟成員闖關被驅逐出境，或者尚未入境即遭原機遣回後，台權會積極主張廢除國安法與刑法百一條，待刑法一百條修正後，主張台獨的言論已經不再是叛亂犯，海外台灣人才能順利的返鄉。二千年後，則轉而關心國內人民的居住遷徙權。

本文主要將透過數位人文研究途徑，應用資訊工具並就台權會所藏的人權報告書、專刊、會議記錄等珍貴文書進行探查。首先，本文將利用詞頻的篩選，找出在文本中相對次數較多的詞彙與概念，其次，透過詞頻權重方法(TFIDF)進一步將所篩選出的重要內容透過加權的重調，擷取其中權重質高者作為分析標的，再者，應用社會網絡分析法(SNA)將文本中的重要概念與事件形成關聯，最終，將實驗產出的相關數據透過視覺化圖表方式進一步呈現在本文中。

關鍵字：台權會、黑名單、人權、數位人文研究

Collective Memory in the Digital Age:

A Case of Beijing Drifters

Ming Ren^{*}, Sherry Xie^{**}, Jin Zhang^{***}

Abstract

The advent of Web2.0 has greatly changed the way information is created, shared and used. People write freely about what they see and hear, personal feelings and opinions, resulting in an explosive growth of user-generated content. This has created an ever-growing virtual space that is parallel to the real world and even richer in some aspects, which provides exciting opportunities to understand the world and human behaviors. Especially for the sub-group of people sharing common culture or interest, they may find their own collective memory in this space. The user-generated content is from multiple people and perspectives, with much details and diversity, is keeping a vivid record of the people. Though it may be very personal and fragmented, a collective image will emerge from a large amount of data, which is supplement and verification of the traditional records.

Despite of the potential value of user-generated content, it is not easy to exploit it for the collective memory. The immense volume of user-generated content but intensifies the information overload problem; besides, it contains much redundancy, and noise. As a preprocessing step, the noise texts and unrelated ones needed to be filtered out, resulting in a clean dataset that is possibly much smaller in size. This can be done based on pre-defined rules, or using classification techniques. To do so, a part of texts needs to be labeled as 2 category (i.e., related, or unrelated). A classification model is trained on the labeled data, and then is used to classify the other texts automatically.

To exploits large set of user-generated content (e.g. texts), an approach is needed to extract the representative information from the large set of data, which covers the greatest amount of information and maintains diversity. Specifically, the first representative is the one that has the highest average similarity to the other sentences; and a subsequent representative, if any, are selected incrementally such that it is the

* School of information resource management, Renmin University of China, Email: renm@ruc.edu.cn

** School of information resource management, Renmin University of China, Email: sherrylx@gmail.com

*** School of business, Renmin University of China, Email: zhangjin@rbs.ruc.edu.cn

most different from the representatives already selected. In this way, the representative subset of sentences will cover the greatest amount of information while maintain diversity as well as low redundancy. Beyond this, the sentiment analysis as well as other techniques (like word-cloud) can be used to analyze and visualize the texts.

It is interesting to exploit the texts at the document level, or more specifically, at the theme level, so a collective image can be generated at different levels. To do so, a set (or taxonomy) of themes needs to be identified, so that the analysis can be conducted by theme. To identify the themes, qualitative analysis and text clustering (e.g. LDA) will be combined. Qualitative analysis based on literatures provides theories and domain knowledge, and reveals the themes commonly addressed in the studies (e.g. sociology, anthropology). Quantity analysis can find the quantity characteristics in the texts and greatly reduce the human labor needed. Only the combination of the two can generate meaningful text clustering results in an efficient way. According to the text clustering results, we know the keywords for each theme, as well as the distribution of the texts over the themes. Thus for each theme, a set of related texts are selected and the sentences containing the keywords of this theme are selected. From this set of sentences the representative sentences is to be selected.

In this study we attempt to understand a special group of people, the Beijing Drifters, using the proposed approach. Beijing drifters generally refer to a group of non-native people who work and live in Beijing, holding no Beijing Hukou. Originally, it refers to people working in show business without an affiliation in Beijing, but it has now expanded to almost all the people who seek opportunities in Beijing, including college graduates, and foreigners. It is difficult to tell the exact scale of Beijing drifters, and a common saying is that there are around 8 millions of them. They share common similar experiences and common objectives, such as a decent job, recognition and social inclusion. A collective image of this group of people is not only important for themselves, but also is critical for policy makers in reforming the policies, maintaining justice and fairness in Beijing and elsewhere. It is not easy to carry out such a research using traditional surveys, which require much manpower and resources. However, the web text makes it easier.

This study presents a pilot study using data from Sina Weibo, including a collection of short interviews of people in Beijing since 2014. The respondents include college students, white collar, workers, foreigners who work and live in Beijing. Each blog, with an average of 252 characters and a photo, tells in the tone of the respondent his demographics (e.g., hometown, job, etc.), the feeling of the life in Beijing, and/or the thinking of the past and the future. It is a good source for generating a collective image of Beijing drifters.

On the whole, the Beijing drifters are satisfied with their life in Beijing. The mostly

discussed term are: Beijing, work, like, friend, dream, graduation, hope, persist, college, choice. And they have shown a high percentage (86.05%) of positive attitude, which means they have a good or hopeful life. We also look into the texts with negative sentiment, and find the frequent terms as: leave, work, hospital, graduation, child, hometown, one year, back home, three years, family, from which we can see partially why they are negative towards the life here. In a word, a positive image of Beijing drifters has emerged. They pursue their dreams with passion and persist.

In order to conduct a theme-level analysis, we firstly formulate the taxonomy of themes based on literatures, which includes three levels, i.e., economic, social, psychological. The economic level, as the basis of the survival conditions, reflects how an individual's basic needs are met. The social level reflects interpersonal connections of an individual among the social groups, which give an individual a set of belonging and acceptance. The psychological level reveals how their psychological needs (e.g. understanding, trust, self-esteem) are met in the urban life, which gives an individual a sense of contribution or value. Compared with the text clustering results, the taxonomy is made more specific. The economic-level themes are: housing, job, and transportation; the social-level themes are: social connections (i.e., friendships, intimacy, family), after-working hours, political participation; the psychological-level themes are: the motives of staying in Beijing, the feelings of the life in Beijing, the spirit outlook. We also know the keywords for each theme, e.g., work, salary, resign for job, subway for transportation.

Next, according to the themes above, each piece of the texts is analyzed. The psychological level is discussed most, with a higher ratio of 92.09% than economic (79.53%) and social(77.67%). The mostly mentioned themes are: job(67.44%), the feelings of the life in Beijing(64.65%), the motives of staying in Beijing (63.26%). The representative sentences are extracted for each theme, which provide general knowledge of how their life in Beijing is. In short, their economic conditions are ordinary. Although they are satisfied with their jobs, they have the pressure is mainly from the expenditure of renting or buying an apartment. Beijing drifters would like to learn new skills and are active in sports (e.g. running, climbing), and gathering. They tend to make friends with common interests. Beijing drifters come to or stay in the city due to social relationships, to find employment opportunities, or to pursue future prospect. They like this city that is so inclusive and attractive, and they enjoy the struggle life. Although they may not have a feeling of belongings, they would not regret to devote their youth to Beijing.

This paper has proposed an approach to generate the collective image of a specific group of people and presented a case study in generating a collective image at the document and theme levels respectively. Although this paper focuses on the text, the

approach can be extended to deal with other types of data.

Keywords: Collective Memory, User-Generated Content, Representative Information, Text Clustering, Beijing Drifters

Research on the Web Design for Disseminating Culture Information on an Ethnic Minority

Guliziba Paierhati*, Watanabe Hidenori**

Abstract

This research aimed to practice a project for disseminating information about an ethnic minority, in this case the Uighur people Ethnic minority cultural information that tends to be buried, especially compared with other information, including the international situation, was placed to inform the general public about the Uighur. In this research, investigation about the tendency of a web design is conducted and the present condition about a Uighur's cultural information dispatch is grasped. The design contents of the website about the existing Uighur were analyzed As a result of analyzing the number of characters and photograph number of sheets of each website, the design with little photograph number of sheets with many characters was complicated, and subjects, like that it is hard to look for information and the contents which are not right appear here and there were found .Subsequently, based on the result, an activity is practiced about the information dispatch by SNS.

Keywords: Ethnic Minority, Cultural Information, Web Design

* Tokyo Metropolitan University, Email: gulzibaparhat@gmail.com

** Tokyo Metropolitan University, Email: hwtnv@tmu.ac.jp

Digital Ricoeur: A Digital Research Portal

Ming Yeung Cheung*

Abstract

Background: about Paul Ricoeur

The French Philosopher Paul Ricoeur (1913–2005) is generally considered one of the most influential thinkers of the twentieth century. Having a long and prolific life, Ricoeur's publication in English alone amounts to over 40 books and 230 articles. Besides English, his works has also been translated into many other languages in the world. The international scholarship of Ricoeur's work has been under constant growth for over a decade. In Asia, international conferences on Ricoeur has been held in Soochow University in Taiwan in 2014, and in the Phillipines in 2015. The Paul Ricoeur Center was established in 2015 in Soochow Univeristy by Prof. Cristal Huang. International conferences on the thought of Ricoeur are regularly held in Europe and North America. Moreover, four Ibero-American conferences on Ricoeur have been held in Latin America since 2010 (Guadalajara, 2010; Rio de Janeiro, 2011 and Santiago de Chile, 2013 and La Plata, 2015).

As the interest in Ricoeur grows, the need of an archive providing access to Ricoeur's complete works, including some resources which are difficult to locate, becomes more and more important and urgent. Such an archive will serve a large community of researchers. A digital archive equipped with search and analytical tools can not only help answer some of the questions that scholars have, but even help them to discover new relations and ask new questions.

In this paper, I will report about the Digital Ricoeur digital portal project, an on-going project of providing digital research tools which involves international cooperation to overcome different difficulties, some may be common to most DH researchers, and some are quite specific for this particular project. While the project is still going on, the project team can demonstrate some insights into Ricoeur's works through some experiments with analytic tools. The international acceptance of Ricoeur's work also means that the future prospects of the project should be the possibility for a non-French speaker to be able to search and even understand texts in other languages, including the works of Ricoeur in their original French, with the help

* Fu Jen Catholic University, Department of Religious Studies, Email: 137027@mail.fju.edu.tw

of machine translation.

As a project initiated by George Taylor and Fernando Nascimento, Digital Ricoeur is both a response to the need raised by the large amount of texts that a scholar faces when s/he wants to research over the evolution of thinking across the long span of Ricoeur's lifetime, or to find out everything that Ricoeur has discussed about a topic, or an author, in his work.

Goal of the Project: Portal vs. Edition

Our project aims at providing scholarly research tools for exploring Ricoeur's work to researchers who are interested. Instead of providing one edition of his works, we would like to make available text analysis tools that will allow the researcher to explore thought a complete digital archive of Ricoeur published work in English. The results of these tools should not dictate what questions may be asked, but enable the researcher to use these tools to locate possibly interesting topics in which he or she may choose to explore further, with some other digital tools.

Brief History of the Project

According to Taylor and Nascimento, they started to apply text analysis computer programmes to a large body of English translation of Ricoeur's work in 2015. By using DH tools, they were able to confirm some of the scholarly insights about Ricoeur's thought, but they have some unexpected findings as well (Taylor and Nascimento 2016). As they present their initial work in different international conferences, the project team has gradually grown to six members, of which the author of the current paper is one. In their paper, they talk about the project in five aspects: digital access; copyright; text preparation for digital searches and analysis; the digital searches and analyses; and an archive portal interface. Since the publication of this paper and the expansion of the project team, all five areas have made significant progress.

In this paper, we will also report on the digitization process, especially our decision on how to markup the text resulted from passing the scanned images of books through OCR. Regarding the copyright issue, in order to comply to the "Fair Use" convention according to the U.S. law, we have put limits on the number of words that a user can see when she search for a term in the texts. Therefore, our term search tool returns the page numbers and a limited keyword-in-context. In fact, some testing and calculation have to be done in order to find a suitable limit to the text that we are showing. The decision process and its implications to our project will be discussed in the paper.

In their paper, Taylor and Nascimento report some of the experiments that they have done by applying several DH text analysis tools to a set of Ricoeur's writings. We will discuss some of them in this paper: their implications, how to refine them, and how

to extend them in the tools that we are testing and planning to use.

Challenges in Specific Areas

As we are setting up the portal, there are some difficulties that we encountered which are seldom discussed in DH. This might be due to the particular material that we are dealing with. We will discuss our current solutions to these difficulties, which may help in future projects of a similar nature.

Copyright Issue

One of the challenges of digitizing the works of a recent philosopher is the issue of copyright. Most of the digital archiving projects are dealing with texts that are not copyrighted, mainly because of their antiquity. Our project differs from many projects in DH which handles ancient texts, in the sense that we cannot show the complete texts to the readers because the texts that we are archiving are copyrighted. At the current stage, the scanning of the books and articles, OCRing and storing the digital form of the texts are rights that are protected under U.S. laws. (Taylor and Nascimento 2016, 126-127)

Digitization Process

Another challenge that we are facing is the digitization process. While the books and articles have to be scanned and digitized using optical character recognition (OCR). We further markup the texts resulting from OCR with TEI specifications. We have decided to apply a minimalist version of the TEI specification for this process. As philologists will know, tagging texts is a laborious work. Since we are scanning books that are printed in a modern time, the OCRed texts have some regularity that allow us to automate part of this process by programming. In this way, the workload of, e.g. tagging the page breaks, has been significantly reduced.

Ongoing Work with Analytical Tools

Voyant (www.voyant-tools.org) is a tool, or rather, a set of tools for text analysis developed by Stéfán Sinclair and Rockwell of McGill University in Canada. The tools that Voyant provides may be used individually, or they can be put together as widgets on the same page. Voyant tools provide the visualization and user interaction for different analyses, such as term search, or frequency lists, correlation of terms. Analytical data is provided not only for individual documents but also throughout a corpus consists of multiple documents.

Voyant reads documents in different formats, including the TEI format. Therefore, we are able to directly feed our TEI tagged files to the Voyant. The advantage of feeding a text to Voyant in TEI (XML) format rather than plain text is that we can specify the

document title so that it will be shown in tools such as Document, or Trends. This is very important for the analysis of a multi-document corpus such as the one we are using in this project. User will be able to see and select the document(s) that they would like to shown in graphs and tables, etc. instead of the whole corupus.

In this paper, we will discuss some of the tools Voyant provides that we plan to use on Digital Ricoeur portal, at least as a first stage. Furthermore, we would like to address how Voyant accomodates our need for protecting the complete text to be accessible for users. This feature of Voyant is important for our project, because the texts of Ricoeur are under copyright. Voyant provdes a limit-access mode during the document / corpus upload process, which allows us to create a limit-access corpus when we set a password through the access management option.

While Voyant may be suitable as a ready-made tool-set, the project team is also considering migrating in the long run to a R/Shiny environment for the flexibility of a programming language and simple web visualization and deployment. Shiny provides web based user interface for R. In the future, we can implement the tools provided by Voyant using R, while tuning and customizing the tools according to our specific need.

Test Feedbacks and Future Work

In this paper, we are also going to share some ideas on the desgin of the user interface. There are things that one has to bear in mind when making design decisions about the user interface, or better, the user experience. We would focus on the ease of use, the impression that a design would give to a scholar who is not very tech-saavy, and the motivation of trying the new (but also more complicated) tools to explore new questions.

What kind of analytic tools will be more easily understood and used by philosophers and humanists who are not tech-saavy? We will discuss the decisions that we have made regarding this question, for a the pilot run of our web portal, through which we get some feedbacks from potential users.

Last but not least, we will discuss about the more long-term goal and prospects of our project, including a multi-lingual archive and support.

References

Taylor, G. H., & Nascimento, F. (2016). Digital Ricoeur. *Études Ricoeuriennes / Ricoeur Studies*, 7(2), 124–145.

《法海遺珠》、《道法會元》中道法系統的數位人文研究：

宋、元、明道教官將網絡分析

高振宏*、鄭有容**

摘要

南宋金元時期被視為道教第二次「宗教大突破」時期，各種新興的道法及經典出世因應世變而創立，諸如神霄雷法、靈寶大法、元帥法、內丹學、飛鸞開化等，而道士與行使各式法術的法師相當程度的互滲與影響，與近世的宗教文化有密切的關聯，深受目前道教學界重視。其中《法海遺珠》(CT1166)與《道法會元》(CT1220)是備受關注的兩部法術彙編，記載了當時新興的神霄、天心、清微等道派的思想與內涵，涵括了考召、祭鍊、附生童法、元帥法等各式流行的儀式和法術，近年中外道教學者諸如司馬虛(Michel Strickmann)、施舟人(Kristofer M. Schipper)、勞格文(John Lagerwey)、鮑菊隱(Judith M. Boltz)、常志靜(Florian C. Reiter)、康豹(Paul R. Katz)、戴安德(Edward L. Davis)、梅林寶(Mark Meulenbeld)、Lowell Skar、橫手裕、松本浩一、二階堂善弘、酒井規史、劉仲宇、李遠國、李志鴻、李豐楙、謝世維、高振宏、吳瑞明等，陸續透過個案研究，逐步建構相關儀式及傳行情況，已能較清楚地把握《法海遺珠》與《道法會元》中的道法系統，同時也對整個宋元以來的道教法術、道士與法師、道法互動等課題有更清楚的研究。

然而，目前研究多以個別儀式或單一道派傳統為主，諸如新神霄派的雷法、酆岳法、火府法、馬元帥法、溫元帥法、趙元帥法等，相關儀式或法術間的傳衍或互動主要透過傳法祖師(譜系)或主法官將來連結，討論其間的可能關係，可是這樣類關鍵詞的串接方式仍是依從傳統人文研究「索引派」的方法，但就目前道教研究來看，宋元以來的道法系統中傳法祖師有不少託名或虛構的情況，其主要目的在於藉此建構該法的神聖性與權威性；而道法傳承也並非全然是單一線性的師徒傳授關係，許多道士與法師可能同時兼習不同的道教儀式與法術，也又透過同道間的參法、換法(同道之間的交流或交換不同的法術)或法本刊刻來習得其他法門，若僅就祖師或官將來討論這些道法間的關係，邏輯與理論基礎稍嫌薄弱，因此方法論上需要重新調整，同時也需要有更全面性或整合性的討論。準此，本文擬透過數位人文範疇中的社會網絡分析技術，以各法術系統中的主法官將為線索，嘗試重新建構《法海遺珠》與《道法會元》中法術系統間的關係，也兼及

* 國立政治大學中文系助理教授，Email: kch@nccu.edu.tw

** 國立臺灣大學圖書資訊系暨研究所博士生，Email: ron0816@gmail.com

《藏外道書》、《清微元降大法》(CT223)等相關道教經典，期望能對宋元以來新興的道法儀式能有更清楚的把握。

社會網絡分析結合社會計量學(Sociometrics)和圖論(Graph Theory)的概念，從社會學借用結構分析(Structural Analysis)觀點，便於宏觀的檢視大規模社會系統中的關係結構，運用社群圖(Sociogram)方法標示社會系統中的個體與屬性，釐清複雜脈絡關係下的微觀網絡模式。近年來，哈佛大學費正清研究中心(Fairbank Center for Chinese Studies, Harvard University)等團隊開發中國歷代人物傳記資料庫(The China Biographical Database)，將宋、元、明時人往來紀錄以社會網絡分析呈現，已取得很好的成果。然而目前數位人文方式進行道教研究的學者甚少，徐源(Micheal S. Baker)曾嘗試以此討論六朝上清經派的藥物地理分布與內涵，而日本學者松本浩一(Koichi Matsumoto)則試圖透過數位文字判讀技術，運用符籙間語意使用關連來討論宋元道法系統。除此之外，其餘可參照的相關研究相當有限，所以本文為一嘗試性與創造性的研究作法，聚焦於道將官將網絡分析，比較《法海遺珠》與《道法會元》的道法系統關係，以釐清道教科儀系統的傳行情形。

考慮到前文所述問題，因此本文在方法論上盡量避免本源說或因果論式的判斷——僅透過文獻時間、傳法祖師等斷定該道法系統的時限，繼而視其為最原始的權威，之後的法門皆是由其傳衍而出。因此本文盡量將這些道法系統視為各種官將、法門組構後的結果，透過這些更細微的元件分析來建構這些道法系統的網絡關係，進而在這種比較明確的網絡關係中來討論各儀式、法術間的可能傳行情況。當然，無可避免的相關資料中仍會有各種時間、區域等線索，因此也還是必須參照傳統研究方法，透過專業研究者的判讀來斷定該道法系統的流行時間、地域等，在這兩種研究方式的參照下可更清楚、全面地釐清兩書中的道法系統。其次，之所以以官將為中心來建構道法系統的社會網絡，係因宋代以來新興的道教法術中大量出現各式元帥、將軍，日本學者將其稱為「元帥神」、道教學者則稱為「官將」，其法門多是由行法者驅遣這些官將執行驅雷降雨、斬邪除祟等任務，在科儀法本中大體都有「祖師、宗師—主法官將—副將—法術內容」的結構，有時未必列出祖師、宗師或是副將，但幾乎都有主法官將，而相同的官將有時會出現在不同的道法系統或是轉為其他道法的副將，由此可知，這些官將位居整個道法系統的核心，依此為主軸來建構社會網絡將能更精準把握當時的道法系統。此外，為避免前述傳統「索引派」方法的缺失，因此除了以官將為軸外，也會對「法術內容」進行更細緻的討論，分析出其中更基礎的各種法門（諸如罡、呪、符、訣等），在社會網絡中也加入祖師、宗師及這些法門的變因。

所以在實際操作上，在個人網絡(Ego-Centric Network)層次，本文將透過社會網絡分析中的中心性(Centrality)測量方式，釐清宋元以降道教官將在道法系統中所扮演的角色與影響力；在整體社會網絡(Socio-centric network)層次，進一步透過聚類分析(clustering approach)方式，比較《法海遺珠》與《道法會元》所組構的道教官將、法式的關係紐帶(Relation ties)，以釐清其如何透過不同的元件來形構出具該法特色的科儀法式，期能由此更清楚看到宋、元、明道法網絡關係的

關連性與多變性。

關鍵字：法海遺珠、道法會元、官將、數位人文、社會網絡分析

臺灣歷史人物文本檢索與探勘系統之建置

謝順宏*、柯皓仁**、張素玟***

摘要

「人物」是歷史學研究重要的實體類型之一，因此，對人物傳記的深入了解有助於歷史事件的相關研究。值此大數據時代，以人力從大量人物傳記爬梳、彙整資料頗為曠日廢時，宜妥為運用資訊科技協助歷史學家。此外，儘管臺灣過去已建置眾多資料庫，也有各種人物傳和可資應用的資料文獻，卻較少進行歷史人物資料庫勘考、分析工具的開發。有鑑於此，研究者乃組成研究團隊，以《新修彰化縣志·人物志》為文本來源，發展資料庫檢索、全文檢索、文本探勘與社會網絡等分析工具，協助歷史人文學者進行研究，長期目標為建置「臺灣歷史人物資料庫(Taiwan Biographical Database, TBDB)」。

本文主旨在於描「臺灣歷史人物資料庫」所收錄之人物特性，闡述系統架構，以及說明目前初步成果。此外，本文將提出一套演算法辨識《新修彰化縣志·人物志》中的命名實體(named entity)，並以詩社名稱辨識為例說明。該套演算法的召回率達 96%，精確率則為 65%。最後，本文將說明建置「臺灣歷史人物資料庫」過程中習得之經驗和未來發展方向。

關鍵字：臺灣歷史人物資料庫、文本檢索、文本探勘、社會網絡分析、命名實體辨識

* 國立臺灣師範大學圖書資訊學研究所，Email: mayh@ntnu.edu.tw

** 國立臺灣師範大學圖書資訊學研究所教授，Email: clavenke@ntnu.edu.tw

*** 國立臺灣師範大學臺灣史研究所教授，Email: 109682@ntnu.edu.tw

期刊、性別與動員：

《婦女共鳴》社會網絡分析（1929-1944）

葉韋君*

摘 要

本文透過建立《婦女共鳴》作者的群體傳記，進行社會網絡分析，藉此彌補歷史傳記資料女性資料缺失的問題，並進一步分析國民黨資助的婦女刊物特質，解釋女性菁英的社會網絡與媒體經營關係，解釋組織連結的情形。

本文首先分析《婦女共鳴》與商業刊物《婦女雜誌》作者群體的差異，進一步解釋內部的分化與合作情形，並以主要編輯李峙山、談社英、陳逸雲分析社會網絡、媒體 關係與政治資本交換的型態。

關鍵字：期刊、作者、性別、國民黨、社會網絡分析

* 中央研究院近代史所博士後員，Email: videllya@gmail.com

從 CBDB 到 TBDB： 以《新修彰化縣志·人物志》為試金石

張素玟*、李宗翰**、柯皓仁***、李毓嵐****

摘要

一、前言

人物是歷史學研究重要的一環，從人物的各種面向與牽引出的出身經歷、社會階層、人物流動、婚姻網絡、政治網絡，都是研究歷史的關鍵議題。過去對人物相關網絡、家族譜系的研究主要依循傳統史學方法，以人工方式從資料文獻加以分析排比，對單一人物、單一家族的研究或許可以，若涉及到龐大的資料與複雜的網絡，往往力有所未逮。因此哈佛大學以累年的努力和集體力量，積極經營中國歷史人物傳記資料庫（The China Biographical Database CBDB），期待透過建立基本檢索系統和文本分析的功能，讓人物傳記資料庫同時具備資訊服務和研究資料文獻的功能。日本與韓國近年也在發展自己的 BDB。台灣史研究成果長久累積，又有為數眾多的資料庫，現在應是發展臺灣歷史人物資料庫 TBDB（The Taiwan Biographical Database）的時機了。

目前臺灣的學界對人物研究主要遵循傳統史學的文獻研究法或輔以口述歷史，台灣史學界對單一人物或家族的研究，兩者都有不少成果。但是傳統的研究方式，每一主題皆需經年累月，一一耙梳文獻，比對資料，處理極為費時。儘管各研究者對某地區的家族、士紳或農村菁英有所深究，卻難以梳理眾多人物的網絡關係；筆者進行彰化縣人物分析時，以人力處理已經左支右絀，更何況全臺灣的人物？因而深深瞭解到利用數位工具處理龐大資料的必要性。

臺灣史的研究經過幾十年的積累，研究主題由小區域、細緻，慢慢推向更寬廣與更長時距的議題。方志的纂修，人物家族、社會網絡的研究，都有可觀的成果。過去 20 年政府積極推動資料庫建置工作的完成，也使資料的流通與檢索更方便。儘管群體人物傳研究的主觀、客觀條件日臻成熟，臺灣卻一直沒有具有勘考、文本分析功能的歷史人物資料庫，因此建置臺灣歷史人物資料庫的目標便是在這樣的背景下啟動，先以《新修彰化縣志·人物志》為勘考對象，建構一個優質的人物傳記資料庫與相關軟體服務，作為數位化時代所不可或缺的人文基本建設。

* 國立臺灣師範大學臺灣史研究所教授，Email: 109682@ntnu.edu.tw

** 國立臺灣師範大學歷史學系副教授，Email: tsonghanlee@hotmail.com

*** 國立臺灣師範大學圖書資訊學研究所教授，Email: clavenke@ntnu.edu.tw

**** 國立中興大學歷史系副教授，Email: leeyulan@dragon.nchu.edu.tw

二、以《新修彰化縣志·人物志》為建置資料庫起點的原因

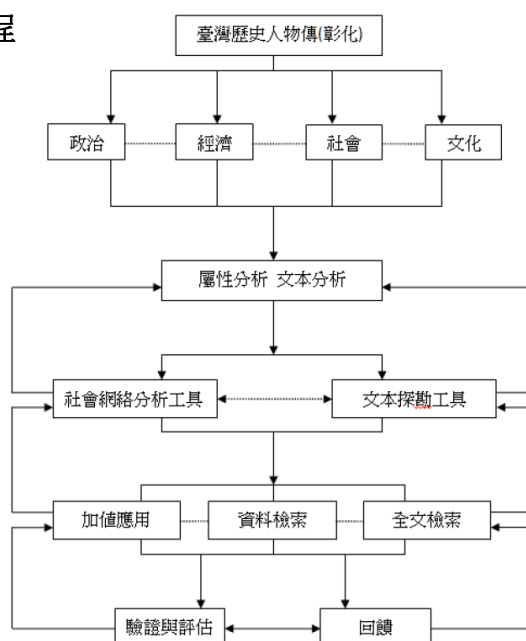
《新修彰化縣志·人物志》共分經濟、社會、政治、文化四篇，篇幅超過百萬，收錄 886 人，以之作為勘考、分析工具研發的文本資料來源，其具備的有利條件為：

1. 體例統一、考訂嚴謹、資料來源豐富，歷次審查皆獲肯定，是一本水平甚高的志書。
2. 四篇人物志撰稿者皆參與 TBDB 的建置，能確實掌握人物屬性分析與工具需求。
3. 作為勘考對象的文本來源「乾淨」，無版權、著作權問題，且能取得全部文字檔、圖檔，不需進行全文輸入的工作，省卻資料庫前置作業的時間與經費。
4. 彰化縣從清代以來便是具有政、經、文化的重要性，包括了多種屬性的重要人物，臺灣的一級大家族和為數甚多的地方仕紳，在全臺灣也相當有代表性，以彰化縣的人物傳作為建置臺灣歷史人物資料庫的文本資料來源為適切之選擇。

三、以 CBDB 為參考 以 TBDB 為目標

本系統在中國歷代人物傳記資料庫（CBDB）的基礎上，修訂並擴充 CBDB 的類別架構，首先需依照臺灣近現代人物與時代的特色，分析臺灣歷史人物與中國歷史人物的同質性與異質性，以調整 CBDB 原有的人物屬性架構。又因 TBDB 已收入現代人物，而必須找出時代變遷下人物多層性，和社會條件、物質文明發展帶來的新屬性，而需擴充 CBDB 的架構。在這套資料庫中必需協調知識本體的分析、研究功能的需求，控管資料探勘工具、分析工具功能的開發，完成 TBDB 資訊檢索與文本探勘前的處理機制，檢視政治、經濟、社會、文化四類人物所建立的基本屬性與關係屬性，是否能進行工具性分析。

四、TBDB 建置的流程



五、TBDB 未來的擴充與發展

TBDB 一開始先應用文本探勘與社會網絡分析工具於擬定之主題，產出人物類別、基本屬性與關係屬性、關聯等知識本體。並評估工具之效度和信度，使 TBDB 能確實符合歷史人文學者的研究需求。

以《新修彰化縣志·人物志》完成 TBDB 初步成果後，可望開放學界利用查詢，並進行除錯。日後繼續擴充文本資料，加入其他縣市之人物志、日治與戰後的各類人物辭典，逐漸延展到全臺灣的人物傳記。在擴充過程也因新資料、新人物類型的加入，對工具從事滾動式修正，期待逐漸成為研究臺灣歷史者不可或缺的資料寶庫和勘考分析的有效工具。臺灣學界多年來的人文研究成果，將透過數位科技新的研究方法和工具，增廣研究主題和研究效益。

在國際上，也可透過 TBDB 建立起一個與國際相關研究的平臺，除了和 CBDB 對話，並與其他已建置類似勘考、分析工具的國家的系統互相比較，吸取更多經驗，也使 TBDB 更完善。

關鍵字：數位人文、方志、人物志、傳記、文本探勘、臺灣歷史人物傳記資料庫、社會網絡分析

CrossAsia: Services and Resources for Digital Humanities in Asian Studies

Miriam Seeger*

Abstract

CrossAsia, operated by Berlin State Library, is designed as the central access point for the Asian studies academic community in Germany and for all researchers located at German research institutions who work on Asia with Asian language materials. The paper will present three main services of CrossAsia to invite discussion on library services of the future such as advanced licence agreements, including digital rights management, management of digital objects and metadata, services for digital humanities and social sciences in Asian studies and the role we envision for infrastructural institutions such as CrossAsia to establish cooperation and collaboration among research institutions, libraries, copy right holders, and data providers.

The three main pillars of the CrossAsia platform are: providing access to electronic and printed resources for humanities and social sciences in Asian studies, a metasearch interface called CrossAsia Search, and an Integrated Text Repository (ITR) with the aim of storing the huge amount of data and digital objects licensed in the context of CrossAsia as well as offering these resources to research and education through authorised protocols.

Several decades ago CrossAsia initiated a special interlibrary loan system to make printed materials in East Asian languages available to all Asian studies institutions in Germany and Europe under the umbrella of one-to-one special interlibrary loan contracts with Berlin State Library. This considerably levelled the field in Asian studies in Germany and beyond. Beginning in the early 2000s CrossAsia began to build up a service parallel to the special interlibrary loan system that provides access to electronic resources. Currently the CrossAsia platform subscribes to more than 130 electronic databases including scholarly information published in academic journals and eBooks, huge collections of classical texts from the different regions and time periods, governmental and official archives, up-to-date newspapers, and statistical data. We are able to grant access to these databases for all scholars working on Asia, as long as they are registered with CrossAsia and are affiliated to a German academic institution that has signed the named special interlibrary loan contract. For our license agreements with

* Berlin State Library East Asia Department, Email: miriam.seeger@sbb.spk-berlin.de

copyright holders we strongly focus on nationwide access for the Asian studies research community as well as on hosting and full text indexing rights. As far as possible we also acquire bibliographical metadata for all digital objects on the most granular level. These activities build the basis for the two other services.

The CrossAsia Search provides access to more than 100 million bibliographical data of all kinds of electronic and printed materials with a focus on Asian studies. It is the access point to the State library's East Asia collection, which is the largest collection of East Asian literature in Europe. All scholars, students and teachers of German and European research institutes focussing on Asian studies and having signed the mentioned special interlibrary loan service contract with Berlin State Library can order printed materials directly from the metasearch interface or – after log in – can access licensed electronic resources. The contents of several important electronic databases licensed by CrossAsia are searchable and accessible on the book and article title level as well as via metadata and abstracts.

Bibliographic metadata from a selection of publishing houses and national libraries are integrated into the CrossAsia Search to serve as orientation for our users and as basis for PDA (Patron Driven Acquisition) requests. In addition the extended search area of CrossAsia Search allows searching in bibliographic data of other relevant sources for Asian Studies.

Another aim of the CrossAsia Search is to bridge the divide between the visibility of electronic and printed materials. Since electronic materials became more and more visible and more easily accessible in online catalogues and discovery systems in comparison to printed materials, we are preparing a “full text search” in our holdings of printed books. Agreements with database providers allow us a full text search in their eBook databases for titles we have in print in our collection. We will show users snippets of 2-3 lines in their search results and a link by which they can order the printed books via our special interlibrary loan service. We plan to also build-up a full text index from our own new acquisitions of printed material by scanning and OCR.

The third service, we are currently developing, is an Integrated Text Repository (ITR) with a twofold aim: first, to organise and host the huge amount of data and digital objects that has been licensed in the context of CrossAsia as well as data that is in the public domain, and second, to offer these data for research and education. By integrating a digital rights management, for license-protected materials we guarantee to database providers that the copyright will be protected.

For the pilot project a pool of data with searchable full texts, digital images and descriptive metadata is extracted from their original database contexts with partly strictly predefined access possibilities. A first set of data will include full texts and metadata of Local gazetteers (2,000 titles), Renmin Ribao articles (1946-2012), titles

of the collections Xuxiu siku quanshu and Siku quanshu (ca. 10,000 titles), Taiwanese eBook metadata (7,000 titles), and English language databases of Adam Matthews. The metadata will be converted into a standardised format and full texts are stored ideally on the page level and in a meaningful hierarchical structure. Registered users will be able to perform guided searches within full texts via the CrossAsia Search and will be able to work with these texts with digital tools via virtual interfaces, of course taking into account the copyright conditions. The enhanced or processed data will be stored according to the copyright restrictions in the virtual research environment of CrossAsia, called CrossAsia Campus, which is protected by authentication.

As the central library and access point for Asian studies in Germany it is essential to prepare for the needs of future users. CrossAsia aims at providing relevant resources for the research community in the most comprehensive and integrated way to strengthen textual and digital research in humanities and social sciences, to extract information for their research, and to allow scholars to use their own tools and methods to analyse these materials.

Keywords: Asian Studies, Integrated Text Repository, License Agreements, Metasearch, Electronic Full Text Databases

A Critical Approach to the DH:

The Digital Influence on the Subjective Sphere

Paolo Casani *

Abstract

1 Introduction

In the midst of a significant expansion within academia (Nyhan & Flinn, 2016), the DH has rapidly evolved in significant part to keep pace with innovations in digital technology and corresponding shifts in culture and society (Jones, 2016). By embracing new methods and practices, such as in the visual rendering of large data sets, it has gone beyond its traditional focus on digital archives that in the past characterized it as a field (Klein & Gold, 2016). From being defined by the metaphor of the big tent, so to emphasized its shifting but all-inclusive boundaries (Svensson, 2011) (Terras, 2011) (Klein & Gold, 2016), the DH has more recently been theorized as an expanded field that engages in reciprocity with other areas of research. This takes place both at a shared field-specific level of practices as well as over theory by “confronting habits of minds” (Lauren & Gold, 2016).

Along with praise for the inferred growth in collaborative work (Nyhan & Duke-Williams, 2017), there has also been a growing but healthy inner criticism of DH from its practitioners, questioning such motives as its lack of self-reflexivity, and the candid fashion in which it has tackled political and social issues. Part IV of the 2016 edition of the *Debates on Digital Humanities* (Edited by Klein & Gold) which is devoted to its critics, has topics that range from the DH utopian ambitions (Greenspan, 2016), the neoliberal positioning it has taken within academia (Chung et al, 2016), and its assumed high-brow Anglo-American monocultural view (Fiormonte, 2016). Rather than going on the defense, the DH has embrace new possibilities for a more activist role in what some practitioners have defined as a critical turn (Berry, 2013) (Berry & Fagerjord, 2017). Overall, these critics seem to agree on the need of a cultural critique in the DH taking up, for instance, some of the themes such as gender, race, ideology, politics etc. that were familiar to cultural studies. Yet, in asserting and promoting techniques and methods of the digital, what is the place of DH as a critic of digital technology itself?

* Centre for Digital Humanities, Department of Information Studies, University College London, UK,
Email: paolo.casani.14@ucl.ac.uk

This short paper upholds another stance in regards to critique in the DH: as a humanities and humanist enterprise that questions and challenges the subjective experience of the digital. In doing so, it outlines the theoretical framework of the author's work-in-progress PhD thesis.

2 Levels of Critique

At comfort sitting between their digital technology know-how and socio-cultural understanding, digital humanists as practitioners of DH are ideally placed to examine and analyse the ways in which technology affects society and culture. But that same vantage point also enables them also to uniquely question and think critically about digital technology itself. Digital humanists have theorized and proposed some of the possible tenets of a critical DH, with the target of criticism varying in scope between ideology, gender, race and culture at large. Yet not so much has factually been factually realized (translated) in research studies that empirically tackle with these issues.

Many commentators (Lauren & Gold, 2016) (Berry, 2017) have attributed to Liu the opening of the debate proper about criticism within the DH. In his paper *Where is Cultural Criticism in the Digital Humanities* (2012), Liu points out that the DH rarely extend its critical reach to tackle social issues in such areas as economics, politics or culture. This timid stance sets it apart from the other humanities fields, as it is the case for new media studies, for instance, that confront social issues head on (Liu, 2012). In order to belong to the traditional humanities, Liu maintains as a provocation, the DH must at once embrace culture (2012). In this way, the DH can help to rethink the idea of instrumentality by thinking critically about digital technologies, thus taking into account the forces behind power, finance and government (Liu, 2012). Liu makes two key recommendations: the first is for the DH to come into full dialogue with new media studies and media archaeology, where issues close to DH can be given cultural meaning; the second is for DH to also come into an exchange with science-technology studies, and absorb such ideas as objectivity and actor-network-theory.

But why should the DH be concerned with critique? The fact that its neighbouring areas of research such as media studies are it is not in itself a sufficient reason. Although the overall argument in Liu's 2012 paper is that the DH must do so in order to be recognize as a full partner with the humanities, his outline of a historical, cultural but most of all methodological contextual entitlement to critique is more pressing. DH must also engage in criticism, Liu states, because of its proximity to the debate between *close* and *distant* reading in the literary studies (2012); in particular, because of the fact that by using representational digital methods it implements the practice of distant reading. In his groundbreaking book *Graphs, Maps, and Trees*, Moretti describes distant reading, in the context of the study of literature, as a specific form of knowledge that, by using alternative

representations to text such as graphs and maps, gives an immediate and comprehensive sense of the interconnections between its parts (2005). These representations not only allow to give voice to what remains unread and thus unheard of, (only a limited amount of texts is studied and analysed by critics) but to reveal the forces that come into play in shaping the forms of these cultural objects (Moretti, 2005). On account of the unmasking of latent social and cultural (and economic) forces by the application of distant reading, critique appears to be intrinsic to the DH. This gives it what Liu overall calls a 'special potential' (2012), which thus only needs to be unleashed and fully realized.

Liu appears to maintain that the humanities, at least since the cultural upheaval of 1968, have a mission to communicate and advocate their vision of humanity and critical values, ultimately for the betterment of society. Yet, that leaves a more radical question unanswered: is critique in our times still needed at all? Latour much cited *Why Has Critique Run Out of Stream?* (2004), starting from the observation that the internet has favoured an overindulgence of critique in such aberrations as conspiracy theories or instant revisionism, revisits the foundations of the same critique of scientific knowledge that he helped to create and promote. Rather than a misuse and abuse of the relativist approach to knowledge by unsophisticated minds, for Latour that is evidence of a flaw inherent to the critical spirit (2004). This calls for a renewal of the critical mind, where critique does merely subtract from empirical facts by dissecting their underpinning social construction; instead, by getting closer to facts it allows to generate new ideas, in this way helping to cultivate a 'stubborn realism' (Latour, 2004). The philosophical route that Latour sketches out is for critique to return to what he names as matters of concern, via the understanding that the very subject matter of enquiry, the thing, is in itself both instantiated and negotiated (2004). Although there are valuable lessons to be drawn from Latour's theoretical reevaluation of critique (2004), it does not seem to address the pragmatic problem from which he sets off: the gap between on one side the knowledge of experts, and on the other more populist (as raising from the bottom up) form of knowledge enabled by Web 2.0 open platforms. It is misguided to assume that criticism in the latter only results in conclusions that are farther away from fact. Furthermore, there are fallacies also in the type of knowledge that academic experts produce when it comes to the domain of human experience which can result, for instance, their highly institutionalized roles, and the very abstracts and long-term perspective that they take on events. Conversely, Liu acknowledges the breakdown of the humanities in their role to communicate to the public, but frames it in a change of the dynamics of knowledge with advancement of forms of networked public knowledge practices (2012). In this picture, with their instrumental approach and use of new technologies, the DH are ideally placed to help to bridge this gap.

3 Dh as a Humanist Enterprise

For the little agreement discussed above about what the DH is, and the uncertainty about its future directions, the DH has ultimately succeeded in bringing the humanities and the sciences, or at least computing, closer. Assessing the balance and the role of its two composing parts, the digital on the one hand, and the humanities on the other, can help to re-examine the possibilities of DH.

The balance between the digital and the humanities in the DH has already drastically shifted with historians of DH delimiting the turning point at around 2006, at a time when it was redefined changing its name what was the humanities computing to digital humanities (Nyhan & Flinn, 2016). Since its inception in the work of father Roberto Busa in 1949, humanities computing had mostly been concerned with applying digital techniques in the study of text (Nyhan & Flinn, 2016). In spite of the fact that its practitioners required a training in classics, its purpose was the application of the newly developed digital techniques to what can broadly be included within cultural heritage. From 2006 onwards, its new name DH, which was already in use (Svennson, 2016), was chosen to reflect not only the broadening of applications of the digital, but the fact that the digital technologies had become widespread to society. Yet, question still remain to this day about what the DH exactly is.

Yet, even though the DH have a very tightly knitted research community, and that there exist many DH associations, it's questionable whether DH can be defined as a distinctive discipline. Svennson argues that rather than a tent uniting together it can be viewed as a trading zone (2011). In tracing the journey from humanities computing to DH, McCartney suggest to interpret the DH as 'an interdiscipline' (McCartney, 2016, quoted in Nyhan & Flinn, 2016), a discipline therefore that brings together and bridges other disciplines. Yet, at a pragmatic level of analysis, it can be said that the DH is a discipline because it functions as one (Nyhan & Flinn, 2016). In the brief history of DH traced above, the direction in the relationship between its two constituent parts have predominantly been from the digital to the humanities. This can be accounted for in no small part to the dominant pro-technology tenets in society. Of course, advancements in technology, in particular with the relatively recent computer revolution, have been a radical force for change that has in no small measure ameliorated our lives. This general consideration in itself would suffice to justify the stronghold of the digital in DH. In our networked tecno-society, however, computer technology shapes the culture we live in to such an extent that an examination of technology is essential to understand how it finds meaning in cultural practices. Its reach affects the personal subjective dimension as well, thus to the way in which we come to think and we envisage the digital. As Berry acutely points out, the influence of the digital can go as far as a "treating the computer as a truth machine" (Berry 2011, 2014) (Berry & Fagerjord, 2017). Taking this into account, it can be argued that the

time is ripe for the DH to look more acutely at how the relationship between humanities and the digital can also further flourish at the reverse: from the humanities to the digital. Drucker rightly remarks that the essence of computational environments, which includes encoded protocols [of operating systems], machines languages, compilers and programming are fundamentally resistant to qualitative, thus humanistic approaches (2012). Since the very nature of these environment excludes a half way humanist compromise, she proposes for the humanities to assert their cultural authority over the digital by demonstrating that “the theory and methods of the humanities have a critical purchase on the design of platforms that embody humanistic values” (2012). Such role for a humanistic study need not only be that of spectators or critics of the contemporary persuasive digitalization or onto the effects of technology. Instead, Drucker goes as far as proposing DH as “a humanistic informed theory of the making of technology” which can be implemented at the level of design, modelling or information architecture, interface and protocols (2012).

On similar lines, in a talk given at Yale university to commemorate the three years from the birth of the Centre for Digital Humanities at University College London (2013), Claire Warwick spells out an interpretation of what the DH should be about. Although the DH must indeed keep engaging with cultural heritage and text, Warwick maintains, they should also be a force by which humanities tools can engage in questioning the digital (2013). By way of acknowledging the importance of the use of such digital technologies as social media are used by museums, Warwick states that social media should be examined as a phenomenon in itself (2013). Moreover, she argues that the DH should take on board the fundamental question that humanists have been asking forever: how do we express ourselves and our own identity? What is like to be human? (Warwick, 2013). Together with what are considered the main areas on concern of DH, Warwick posits that questions about what is like to be human are just as relevant (2013).

Giving that humanities theorizations and models of thinking and understanding the world we live in, together with established methods and instrumentality in the DH, can help to question and criticize the digital, the following question arises: how can the DH best approach, understand, and critique the digital?

4 The Subjective Perspective

Apart from delineating the new computational areas of humanities research, the digital in DH encompasses all the technologies of digital media. We live in what is named a digital age and we communicate via digital networks. In this sense, the digital does not need to be defined as it is all around us.

Even if one defines and understands what the digital is, there is still the problem about how to approach it. Berry & Fagerjord maintain that one challenge "is to bring the digital,

which includes software and computation, back into visibility” in order to do research and critique “both a material and as well as an ideology” (2017). There is a sense in which the digital is immaterial, inscribed not only in the technology that we use every day, but also in common practices. The objective therefore would be to reveal its materiality.

In the keynote speech at the Digital Humanities conference in 2014, Bruno Latour spells out his thinking about what the digital is, by defining in the first instance what it is not. The main target of Latour is to highlight some of the fallacies that we hold when we think about digital technologies, thus the digital (2014). In the cloud effect fallacy, the more people talk about the digital, the more material it is; in the cognitive fallacy, the more people talk about cognition, the more they actually describe a socio-technical environment; in the digital/analogue fallacy people think of the digital as a native function, when in fact it results from a redundancy in the computer institution; in the closed book fallacy, the more people talk about the difference between a book and the web, the more they are much of the same when the whole network is taken into account, necessarily being the ultimate end of analysis in itself, rather just the first step of enquiry. With the increased magnitude in which intelligent digital artifacts influence the conceptualizations that we make of our personal and relational spheres, comes the necessity to critically monitor, question and analyze what are the forces that are going to shape this future. This brings humanities enquiry and the forefront of an analysis of the digital.

In its embracing of close and distant reading, the DH can alternatively or jointly use quantitative and qualitative methods to shed light on the interconnections between how the digital influences personal experience, and in turn how personal experience shapes the digital space and culture at large. This puts the digital in relation to our more intimate and inner sphere of our self. With distant reading practices of visualization, the DH can also reveal some of the forces of shape described by Moretti (2005).

The DH as an interdisciplinary can tackle the challenges posed by the digital from a myriad of angles by combining different methods and understandings of the world. Given the complexity of the world that is prescribed by the digital, mixed methods research applied by the DH can help to more deeply understand the reality that is experienced, expressed and exponentially created for consumption by economic and political forces. My current PhD research, for instance, engages qualitatively with personal testimonies and reflections about the experience through and about digital platforms, and combines it quantitatively with the richness of the affordances that digital technologies allow in order to reveal the general expression by looking at the digital traces that we leave behind on the medium. This approach to the DH gives a vantage point on which to examine, question and challenge the influence that the digital has on our nature and reflexively, on society and culture.

5 Conclusion

This short paper has outlined an alternative critique for the digital. As it further expands, the cultural critique in the DH can target the digital by shifting the balance towards thinking and methods of the humanities. Rather than looking at the digital as a physical reality, it can engage as a starting point of analysis with the subjective experience where the digital impacts the nature of who we are.

6 References

- Berry, D. M., & Fagerjord, A. (2017). *Digital humanities: knowledge and critique in a digital age*. John Wiley & Sons.
- Berry, D. M. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, 12.
- Drucker, J. (2012). Humanistic Theory and Digital Scholarship. In Gold M. (Ed.), *Debates in the Digital Humanities* (pp. 85-95). University of Minnesota Press.
- Fiormonte, D. (2012). Towards a cultural critique of the digital humanities. *Historical Social Research/Historische Sozialforschung*, 59-76.
- Kirschenbaum, M. (2012). What is digital humanities and what's it doing in English departments? *Debates in the digital humanities*, 3.
- Floridi, L. (2008). The Method of Levels of Abstraction. *Minds & Machines*, 18: 303.
- Gibson, W. (2007). *Spook Country*.
- Greenspan, B. (2016) Are Digital Humanists Utopian? In *Debates in Digital Humanities*, ed. M. K. Gold. Minneapolis: University of Minnesota Press.
- Golumbia, D. (2009). *The cultural logic of computation*. Harvard University Press.
- Jones, S. E. (2013). *The emergence of the digital humanities*. Routledge.
- Latour, B. (2010a) The migration of the aura or how to explore the original through its fac similes, in Bartscherer, T. (ed.) *Switching Codes*, University of Chicago Press.
- Latour, B. (2010b) Coming out as a philosopher, *Social Studies of Science*, 40(4) 599–608.
- Latour, B (2013) *An inquiry into modes of existence: an anthropology of the moderns*, Harvard University Press.
- Latour, B. (2014) Opening Plenary, *Digital Humanities 2014 (DH2014)*, available from <http://dh2014.org/videos/opening-night-bruno-latour/>
- Liu, A. (2012). Where is cultural criticism in the digital humanities? *Debates in the digital humanities*, 490-509.
- Liu, A. (2016). *Drafts for Against the Cultural Singularity*.
- McCarty, W. (2014). *Humanities computing*.
- Nyhan, J., & Flinn, A. (2016). *Computation and the Humanities: Towards an Oral History of Digital Humanities*. Springer Publishing Company, Incorporated.

- Ramsay, S. (2016). Humane Computation. *Debates in the Digital Humanities*, 527-29.
- Svensson, P. (2012). Beyond the Big Tent. In Gold M. (Ed.), *Debates in the Digital Humanities* (pp. 159-86). University of Minnesota Press.
- Terras, M., Nyhan, Julianne, & Vanhoutte, Edward. (2013). *Defining digital humanities: A reader*. Farnham: Ashgate.
- Warwick, C (2013) *Making Ourselves Indispensable: The UCL Centre for Digital Humanities at Three Years Old*.

Humanity , *Quo Vadis?*

Graydon Wetzler*

Abstract

I bring together three articles each taking a position on the current stakes for the future of digital humanities, and that as a collection account for the DADH 2017's theme. For a more-or-less shallow sampling of the previous now of digital humanities, I discuss Eve Kosofsky Sedgwick and Adam Frank's 1995 Critical Inquiry article, "Shame in the Cybernetic Fold: Reading Silvan Tomkins." Interestingly, to sample the now-now of digital humanities, I discuss Alexander Galloway's 2014 text, "The Cybernetic Hypothesis," in which the author looks much deeper into the past conjoint of the digital and humanities. As a future referent that will in fact frame the previous two as an impetus is the editor's introduction to the most recent publication of PMLA (March 2017). In her compact text, Yale Professor Wai Chee Dimock suggests a provocative proposal to conceive 'experimental humanities' as a protocol born through the synthesis of what Rudolf Carnap distinguished as the tool-building spirit of active (versus passive verification) science and what I will call the humanities' preference for 'one-to-many' driven inquiry. With these three texts, or positions with respect to future digital and humanities convolutions, I introduce a series of currently ongoing research projects that I will aim to convince provide exemplars for conceiving future DH projects. After summarizing these canonical projects, I draw some possible ways to operationalize both their general as well as certain of their specific research frequencies (a conceptual schema that I introduce at the same moment).

0.1 experimental humanities

In her Editor's Column of the most recent issue of PMLA (2017), Distinguished Yale Professor of English and American Studies, Wai Chee Dimock, deploys experimental as a signifier for testing the feasibility of an isomorphism mapping scientific practice literary study as a field of knowledge. Borrowing Rudolf Carnap's distinction between active and passive as protocols constituting different fields of knowledge, Dimock offers NASA as a contemporary case of science "overflowing with human input, featuring practitioners working with far more than telescopes" by transitioning from onlookers into "builders and makers" (242). Intrinsicly tied to this transition, Dimock further points out, is also to make failing, as "the here and now" of

* New York University

performing experiments. A key fulcrum for my presentation is Dimock's important link from fallibility to engineering practice where alongside the privileging of tool-building, "error is both constraint and catalyst for an experimental science: one way by which as revised trajectory is built on a prior one — even an erroneous one — a cumulative effort extending steadfastly from the past into the future" (243). I suggest Dimock's register of 'experimental' as a productive displacement of positivism. I also suggest this register as a fulcrum for considering two previous texts taking up similar positions in the conjoint of digital and humanities.

0.1.1 Cybernated possibilities and dissolutions

Dimock's test persuades me to consider, in conjunction, Eve Kosofsky Sedgwick and Adam Frank's "Shame in the Cybernetic Fold: Reading Silvan Tomkins," (1995) along side Alexander Galloway's article "The Cybernetic Hypothesis" (2014). These two readings offer productive presuppositions, deeply resolute as well as in tension with Dimock's challenge for humanities to go experimental. Both of these texts share multi-modality as a *modus operandi*, as well as, preferring one-to-many, affective strategies as was forward through computational humanities.

0.2 æffecacy

In a recent Social Research article, NYU Professor of Media and Culture Stephen Duncombe proposed the term 'æffecacy' as a criterion for answering whether a work of activist art 'works' by assessing its causal link between discernible effect and expressive affect ¹. The 'artist' and EDT/b.a.n.g. lab member, Ricardo Dominguez, has offered a variation on Duncombe's metric by characterizing the lab's artistic output as aiming to be "concurrently effective and affective" by, for example, writing effective code that "works" while simultaneously coding poetic affect. Dominguez has referred to this concurrent aim in several places. Among these are Leila Nadir's Hyperallergic interview, "Poetry, Immigration and the FBI: The Transborder Immigrant Tool." How might these and other dynamic couplings of effect and affect situate the computational with aesthetics and politics, as well as, make humanities research practices more affective (embodied) and effective (in deploying theory)?

I introduce the idea of 'performative æffect' as prepending Duncombe's ligatured measure in order to shift from a critical to practical and solution based method without yielding to positivist strictures. With performativity as a basis, research draws on settings of post-object aesthetics, post-humanist perceptive models and surveying speculative approaches to how computational methods can co-situate discernible effect

¹ "Does it Work?: The Æffect of Activist Art", Social Research: An International Quarterly, Volume 83, Number 1, Spring 2016, pp. 115-134

with affective surplus. More poignantly, how might these and other dynamic couplings of effect and affect situate the hope of computational approaches towards aesthetics and politics in practice, as well as, liberal arts research more generally? Towards this I discuss three research settings that I participated in each giving a productive variation on the situated question, and which I will thematize under the notion of:

0.2.1 antidisciplinarity

First is a creative repurposing that the City University of New York Professor of Computer Science Zhigang Zhu achieved from his lab's research in multimodal sentient surveillance engineering to serve as a material basis for conceiving a design strategy for a wearable "multimodal sensing, display and learning platform enabling visually impaired users to see with their hands, bodies or faces. Next, is where current stand-off surveillant media become a mobile wayfinding interface in the former, a locative media (Global Positioning System) is transformed into a dislocative media (global poetic system) in the second case. Extending this double-design twist is also possible by considering how work in the neuroscience of wayfinding finds new purpose in the design of neuromorphic architecture here.²

In closing, I unpack how these projects can provide canonical models for what Professor Dimock alludes to in 'experimental humanities' as rich isomorphisms across quantitative and qualitative fields of knowledge discovery.

I look forward to unpacking this vibrant troika as inroads towards an experimental humanities.

² Michael A. Arbib, "Brains, machines and buildings: Towards a neuromorphic architecture", in *Intelligent Buildings International* 4(3):1-22, July 2012.

Embodiment, Becoming-animal, Nomadic Subject on Subject Issues of Rosi Braidotti's Situated Posthuman Ethics

Yan-Ping Gao*

Abstract

There are three keywords in Rosi Braidotti's situated posthuman ethics, embodiment, becoming-animal, nomadic subject, respectively. Braidotti treats embodiment as a starting point for the discussion of posthuman ethics from the standpoint of feminism, and she points out during the process of analysis of sexual differences that embodiment is not only about the body, the cognition, but also the becoming-subjects on the basis of monistic philosophy. It is necessary to include animals as reference frame to discuss subject issues in the context of posthuman. Braidotti also views zoe-centred vitality as a transverse power to connect people and animals, while the nature of the creatures and the whole planet as a reference dimension of post-human subject is also important. Braidotti advocates the participation of the nomadic spirit in specific social practice and establishment of a positive politics which bases on the current and a politics of hope which bases on the future, so as to defend the legitimacy of posthuman subjects in a theoretical level and show an enlightenment for ethics in the posthuman era.

Keywords: Embodiment, Becoming-animal, Nomadic Subject, Situationality, Posthuman Ethics

* Department of Chinese Language and Literature of East China Normal University, Email: 741446222@qq.com

台灣清治時期地景與事件的文本探勘

林淑慧*

摘要

空間書寫蘊含描述人類存在的現象，主要功能在於重新建構人類的各種體驗過程，這些體驗包含人與自然或環境的關係等。文學與地理學皆是表意作用（signification）過程，即在社會媒介中賦予空間意義的過程。從清治時期的方志史料觀察，一些文本中的地景，亦是歷史事件發生的地點。台灣檔案文獻中的地景與歷史事件究竟有何關聯？從清治時期諸多來台的官員所撰八景詩，所選的地景雖不盡相同，然皆顯現遊宦文人的敘事位置。如八景之一的「水沙連」，從台灣歷史數位圖書館（THDL）查詢出現次數最多為 1787（乾隆五十二）年及 1788（乾隆五十三年）年，此與發生於 1787 年台灣清治時期三大民變之一林爽文事件密切相關。至於地名的變遷，如 1847（道光二十七年）年的文本中，已出現「日月潭」的稱名。

地景兼具人為與自然特性，所以研究地景不可避免必須研究人類的活動，無論是經濟上、社會上或文化上的。此類富涵意義的文化地景，則與社會現象密切關聯。就現今社會救助的概念，是以照顧中低收入戶及救助遭受急難或災害者，並協助其自立等為主要的範圍。回溯台灣清治時期弱勢族群的社會救濟，又呈現哪些時代的現象？從《臺灣文獻叢刊》中有關養濟院、普濟堂、育嬰堂（育嬰院、保嬰局）、留養局（孤老院）等地景觀念叢，可蒐集到諸多相關的史料與文本。如彰化養濟院位於八卦山下，以收養麻瘋殘疾的患者，遺址於今彰化市南屯路靜和醫院。彰化〈善養所碑記〉為 1852（咸豐二年）年生員吳桂芳所撰，文末銘記地方仕紳倡建募捐的善行，此地景遺址位於彰化市民族路關帝廟旁。Tim Cresswell 於《地方：記憶、想像與認同》指出，地方是眾人將意義依附於空間而產生的。台灣在地文人如何書寫風景，則反映自我觀看的心境，及作品流露的地方感。這些或官方出資倡導、或民間資金援助，與現代社會福利或社會連帶責任不同，性質較似慈善事業。文本所建構的行善空間，伴隨著人物特殊的經驗、歷史性脈絡的累積，且意識到人對於臺灣情感上的依附，因而表現地方感的意義。研究慈善事業的地景與事件有助於理解歷史厚度，或重新賦予地景現代的意義。

台灣清治時期育嬰堂的資料，不僅與溺女事件相關，更呈現地方人士參與社會救濟的情形。舉例而言，施瓊芳於 1854 年（咸豐四年）所寫的〈育嬰堂給示呈詞〉，強化協助籌措長期經營育嬰堂的費用，蘊含深厚的生命感受，是由當時

* 國立台灣師範大學台灣語文學系教授，Email: limsiokhui@gmail.com

溺女的社會現象所激發的人道關懷。1866（同治五年）淡水廳枋橋街（擺接堡）的富紳林維源，自捐五千銀設立育嬰堂，1870年（同治九年）再由淡水廳官民義捐，於竹塹城內及艋舺街設立兩所。從國家圖書館《臺灣記憶》碑碣拓片資料庫查詢，如〈艋舺新建育嬰堂碑記〉提到陳培桂 1870年（同治九年）鑑於艋舺到淡水未有照顧孤兒及棄嬰的機構，於是倡建「育嬰堂」，以作為育幼撫孤之用，此育嬰堂碑現存於台北艋舺仁濟醫院門口。曾將此地景作為校外教學的參訪地點，從想像到對照文本與親臨現場參觀古蹟，以引導學生建構歷史記憶。

本文擬透過蒐集資料庫中的檔案、台灣文獻叢刊等豐盈的文本，應用文本探勘的研究方法，重探台灣清治時期地景與事件的關聯。DocuSky 建構主題文字庫，從內文比對及詞頻統計等技術，進行概念擷取、文本分類，再加以脈絡化的詮釋。並結合擴增實境(Augmented Reality, AR)技術，將定位後的文本建置為 Wikitude 的 World 檔案。使用者可在自己的手機或平板等行動裝置上，觀察周邊出現哪些地景文學，再選擇瀏覽文學作品或摘文；甚至利用導航功能，按圖索驥探訪這些地景。期望有助於理解文本與外緣背景脈絡意義，亦提供人文之旅及文化創意產業的參考，進而提昇聯繫土地的情感，具體發揮數位人文學術研究及應用的影響力。

關鍵字：文本探勘、地景、事件、台灣文獻、地理資訊系統

應用語料庫分析取徑探討臺灣女性研究主題： 以臺灣學術機構典藏為例

張映涵*

摘要

世界核心知識體系的建構，普遍以男性主導的世界觀為準繩，女人常位於邊緣地位，女性主義興起，開始挑戰以男性為中心的單一思維體系。幾十年來，臺灣受到西方婦女運動、女性主義思潮以及本土婦運的影響，性別議題逐漸熱門，成為一個新的研究焦點。隨著語料庫分析法（corpora analysis）技術的不斷進步與發展，語言與性別研究（language and gender studies）也是近期才興起的一個研究領域，回顧相關透過語料庫分析法探討性別相關議題者，Koller 與 Baker(2010) 以 feminist（包含名詞、形容詞）、feminists、feminism 為關鍵字，蒐集英國、美國的學術發表、雜誌報導及演講文本，透過搭配詞分析等語料庫分析法，比較英美兩國在 1990-2010 時期的女性主義發展情形。而後 Baker（2014）更進一步發表《利用語料庫分析性別（Using Corpora to Analyze Gender）》一書，專門探究語料庫分析法與性別研究之間的應用。綜觀臺灣女性研究多以理論哲學角度或書目計量法進行文獻回顧與分析，較少透過大量文本及語料分析法來汲取核心概念，並觀察女性研究的發展脈絡。本文依前述相關研究基礎，以臺灣學術機構典藏（Taiwan Academic Institutional Repository, TAIR）為研究範疇，擷取研究題目出現「女」字之資料作為分析文本，並透過詞頻統計（frequency list）、搭配詞分析（collocation analysis）等語料庫分析取徑，探索臺灣學術領域中女性研究詞彙運用、研究發展情形及主題樣態。

本研究透過臺灣學術機構典藏查詢工具，於進階搜尋的「題名」欄位中，選擇標題出現「女」字之所有檢索結果為分析文本，蒐集範圍僅限臺灣學術機構典藏所收錄之 104 所大專／技職院校與研究機構的文章，年代限定自有紀錄以來至 2016 年止，共計檢索樣本數 13,741 筆、99 個單位，並匯出題名、年代與機構等三個欄位作為後續分析依據。資料清理部分，以人工逐筆檢視扣除完全重複（機構、年代、篇名均相同者）、關鍵字錯誤組合（例如非指女性之專有名詞、女子學校的工作報告）、無年代（年代紀錄為 0，無法判讀者）、非研究性質（例如新聞消息、課綱、期刊目錄、名人校友專訪）之內容。另由於資料擷取過程中，標題文字有所遺漏及亂碼，為使後續分析更為精確，逐筆查詢原始文章標題後由

* 國立臺灣大學圖書資訊系暨研究所博士生，Email: d04126002@ntu.edu.tw

研究者自行補足。本研究總計分析文本數共 10,837 筆、97 個機構，總計分析詞彙共 140,511 字。研究分析採用語料庫為本 (corpus-based) 取向，觀察臺灣學術機構典藏女性研究的題名詞彙運用情況，再以聯合國千禧年高峰會議通過之「性別主流化行動」為分界點，進行前後期女性研究主題分析。本研究先利用 Excel 進行描述性統計，包含整體發表數量分析及發表機構分析，以瞭解國內對於女性研究的發表數量及其變化；在研究主題、詞彙運用情形及研究主題趨勢部分，則採用庫博中文語料庫分析工具 (Corpro) Pre-alpha 64 位元版本進行分析，分析項目包含總體詞彙及特定詞彙詞頻、搭配詞及關鍵詞、顯著詞分析等，交叉比對呈現臺灣女性研究自 1950 年至 2016 年來之重要研究發展過程。

審視研究資料分析過程發現，性別主流化的推動確實影響了臺灣女性研究產出，而女性主體隨著時代推移也有所轉變。本研究主要結果有以下四點：(1) 千禧年後女性研究數量大增：本研究統計，臺灣女性研究發表數量在 2000 年後大幅成長，其中 2000 至 2009 年間的發表篇數更達到 5,788 篇 (53.12%)。性別主流化政策使得各領域學門之女性相關研究成果豐富而多元，本研究統計結果正好與該浪潮相呼應；(2) 「女」的主體隨時代演進而有所不同：雖本研究以「女」字做為資料擷取依據，但觀察詞頻分析結果，發現用來表示女性的代名詞非常多元，從中亦可推論相關的研究主體與身分。若以 2000 年為分界點來看，前期 (1950-1999 年) 女性研究以「婦女」為主要的性別代名詞，可推論身份主體為已婚女子；後期 (2000-2016 年) 則以「女性」為主要代名詞，身份擴大為未婚、已婚女子皆有可能。顯示出女性傳統固有的角色逐漸被釋放，不再拘泥於婚姻、家庭、生育教養之身份束縛；(3) 女性健康與醫療技術的發展歷程：女性健康是女性研究中非常重要的研究主題，亦為全球共同關注的焦點。本研究樣本分析臺灣女性研究在 70 年代即有對更年期及生育問題等醫療面向之探究；90 年代開始，第一次提及尿失禁症狀，而生育方面也擴展到墮胎的探討，顯示在敏感的女性議題上有所突破；2000 年後，飲食習慣的改變影響著疾病地演進，開始出現了乳癌、肺腺癌、肺癌等病症；而 2010 年後則出現子宮頸、骨盆、骨質疏鬆等婦女疾病。由此可知，透過不同年代的關鍵詞與搭配詞分析，可初步了解不同時期對於女性疾病關注之重點及層面，還有醫學技術隨著時間發展而日趨進步的軌跡；(4) 女性研究主題圍繞女性生活世界：根據詞頻分析結果，臺灣在女性研究主題上圍繞著女性生活中的個人經驗、家庭、健康、社會互動等議題。觀察前後期主要顯著詞彙可發現，前期主要研究主題為女性的就業及職場；後期則是關懷弱勢的多元族群 (如新移民、外籍)。但不論前後期，都包含了教育議題的探討 (如讀、國小、學習)，可推論無論時代如何演變，教育都是一項作為探討社會價值觀對於性別角色形塑切入點的重要指標。

本研究透過語料庫分析，歸納臺灣學術領域中的女性研究詞彙運用與主題樣態，並初步理解臺灣女性研究趨勢。儘管上述結論是基於臺灣學術機構典藏中的資料文本所產生之推論，然而本研究藉由客觀的分析技術，初步探索出臺灣學術領域的女性研究議題圖譜，可供從事相關專家學者參考。未來還有許多研究課題

需要學者專家共同努力，具體包含以下二點：(1) 擴大研究範疇：藉由涵蓋多元的資料類型及不同的地理區域，發展相關比較研究；(2) 深化理論基礎：透過技術協助，引入各女性主義理論派典，進行內容判斷與識別，必更能以理論架構的觀點與脈絡來探討女性研究的趨向與主流，使分析與詮釋上更有深度。展望未來，期盼本研究的初步結果，能以不同觀點與方法，發掘出臺灣女性相關研究潛在的議題與脈絡，後續更可憑藉不斷進步的研究技術，對相關領域有所貢獻。

關鍵字：語料庫分析、女性研究、婦女研究、臺灣學術機構典藏

以語料庫分析取徑探究

社會對《一例一休》新法案之態度：

以 2015-2017 年台灣四大報紙電子新聞為例

吳孟家*、簡翊淇**、葉克芸***、陳品臻****

2016 年臺灣中央政府正式通過勞動基準法修正案，一例一休改革政策的推動引發社會各界關注與討論，不同類型與層級的組織機構，包括勞方與資方、行政機關與產業機構、工會組織與民間團體，各自表述立場。本研究企圖透過報紙媒體的大量新聞報導，探討新法案上路後，各界正面與反面的態度。研究者以四大報紙在 2015 年 5 月至 2017 年 5 月報導關注「一例一休」議題之新聞全文為研究對象，最終獲得之文本數量合計 3,314 篇，總字數 2,629,077 字；以臺灣大學關河嘉教授與陳光華教授共同開發之語料庫工具—「庫博 (CorPro) 中文語料庫分析工具」進行大量文本語料分析，並輔以社會網絡分析工具「Gephi」做後續詞彙共現詞的分析。

本研究首先找出該議題範疇最常提及的重要關鍵詞彙，配合社會現象進一步解讀文本，並針對四大報紙個別進行顯著詞分析，將顯著詞依照頻率排序，統一擷取詞彙頻率前 100 名之顯著詞，探討不同報紙對於「一例一休」側重的主題，以及哪些組織及團體較常對該議題發聲。透過關鍵詞脈絡分析新聞文本中正面情緒和負面情緒詞彙的使用情形，以及正反面情緒詞彙於四大報的分布差異。最後以社會網絡分析探討「一例一休」新聞常用辭彙在四大報紙之間的使用差異，研究者採用 SNA 分析工具 Gephi 呈現詞彙之間的關係，將各報出現的詞彙分別建立從屬關係，並以詞彙出現頻率做為權重，觀察不同辭彙之間的關連與意義。

研究結果顯示一例一休新聞的社會關注度以 2017 年 1 月熱度最高，且新聞報導大多使用負面情緒詞彙。勞資雙方對於一例一休各自著重的議題點有極大差異。勞方關注自身權益的保障，對於薪資、工時更動、休假等議題尤為敏感；資方則是對於一例一休影響人事成本與組織團體的利益等議題較為關心。四大報紙對於一例一休重視的議題有明顯差異。聯合報較著重在新法案對製造業帶來之衝擊；中國時報較常提及政黨與新法案之間的關連與影響；自由時報則著重關注政府與勞方的關係；蘋果日報除了報導「一例一休」影響服務業、倉儲業之外，尚提及役男議題。

* 國立臺灣大學圖書資訊學系研究所碩士生，Email: r04126011@ntu.edu.tw

** 國立臺灣大學圖書資訊學系研究所碩士生，Email: r04126020@ntu.edu.tw

*** 國立臺灣大學圖書資訊學系學生，Email: b02106054@ntu.edu.tw

**** 國立臺灣大學圖書資訊學系學生，Email: b02106037@ntu.edu.tw

關鍵字：語料庫分析、一例一休、新聞語料、社會網絡分析

威權到民主：歷屆總統就職演說的文字探勘

鄧志松*、周嘉辰**、李宥霆***

摘要

1947 到 2016，近七十年間，中華民國歷經五位、十四屆次的總統，而台灣的政治發展，也逐漸從威權走向民主。本文運用文字探勘(Text Mining)的技巧，比較歷任總統就職演說的差異，追溯威權到民主，價值觀遞嬗的過程。文字探勘的方法，最早出現 20 世紀 80 年代中期，但早期的研究偏向純人工的文字比對，在過去的二十年裡，文字探勘與資訊科技結合，這個領域乃獲得快速的進展。分析的步驟首先將文本作自然語言處理(NLP)文字轉化為數據，以便進行複雜的計算。

研究發現反共抗俄與法統存續是兩位蔣總統就職演說的重點，從李登輝開始，這方面的論述就不復存在了。與此相反的是台灣認同論述，陳水扁時期大幅躍昇。政策議題，除了經濟政策外，著墨不多。蔡英文是個例外，似乎她有意降低意識型態的糾葛，著力於政策議題。比較十四篇就職演說可以發現有兩個集群，兩蔣是個集群，陳水扁以後的諸位總統是個集群，李登輝如同是兩個集群間的橋樑。作者深知僅從這十四篇就職演說，很難掌握台灣政治變遷的全貌，但是經由新方法的導入、跨學科的合作，或許可以相互補充，作為政治學、歷史學分析之助。

* 國立臺灣大學國家發展研究所副教授，Email: terry@ntu.edu.tw

** 國立臺灣大學國家發展研究所助理教授，Email: chelseachou@ntu.edu.tw

*** 國立臺灣大學國家發展研究所助理教授，Email: ytandylee@ntu.edu.tw

Panel: Eurasian Past meets the Future: Digitalizing the Mongol Empire

Chair: Professor Michal Biran

In the past few years the study of the Mongol Empire (1206-1368) has entered a new phase as Digital humanities tools expanded the understanding of the Mongol Imperial enterprise and its impact on world history.

Nearly five year ago Prof. Michal Biran established at the Hebrew University of Jerusalem the project “Mobility, Empire and Cross Cultural Contacts in Mongol Eurasia” which seeks to explain why, how, when and to where people, ideas and artifacts moved in Mongol Eurasia, and what were the outcomes of these huge movements. Studying the Mongol Empire in its full Eurasian context, the project combines a world history perspective with close reading in a huge array of primary sources in various languages (mainly Persian, Arabic and Chinese) and different historiographical traditions, and classifies the acquired information into a sophisticated prosopographical database, which records the individuals acting under Mongol rule in the 13th and 14th centuries. (www.mongol.huji.ac.il)

This panel presents papers of four project members who use a variety of digital methodologies: The Jerusalem database, social network analysis, biographical databases and much more to demonstrate how the novel technologies can shed new light on a crucial period in world history, known as "The Mongol Moment" (1206-1368).

In the first paper Mr. Matanya Gill reconstructs the commercial networks of the Mongol successor state in Iran through a prosopographical analysis of data obtained from 255 biographies of merchants who lived and worked under the Ilkhanate, thereby manifesting the global character of Mongol trade.

Next, Ms. Vered Shurany uses online prosopographical databases to reexamine the 1328 succession struggle in Yuan China on the basis of biographies of known and lesser-known Yuan commanders who participated in the civil war.

In the third paper Ms. Qiao Yang utilizes digital databases to reconstruct the biography of the Uighur astronomer Aruhun Sali, demonstrating how the databases can help in combining pieces of information for reconstructing and analyzing biographic information.

Lastly, Mr. Ishayahu Landa will present the advantages and disadvantages of various digital approaches in the context of historical genealogical research that spread across Mongol Eurasia.

Reconstruction the Commercial Networks of the Ilkhanate with Digital Database

Matanya Gill*

Abstract

Many scholars who have studied the commercial networks of the Ilkhanate (the Mongol state in Iran, Iraq, Azerbaijan 1260-1335) focused mainly on a few famous merchants and their involvement in specific regional trade networks namely the trade networks of the *??b?* family and the Chinese merchants via the Persian Gulf and the Indian Ocean; the commercial activities of the Italian merchants in the Black Sea; and the Mediterranean trade networks via Tabriz and Armenian-controlled Cilicia. Furthermore, the commercial networks that operated between the Ilkhanate and Western as well as Central Asia did not receive sufficient attention in the research literature.

This paper aims to reconstruct the commercial networks of the Ilkhanate, both within and beyond its frontiers, through a prosopographical analysis of data, obtained from biographies of 255 merchants who lived and worked under the Ilkhanate, that were uploaded to the digital database of the ERC research project “Mobility, Empire and Cross Cultural Contacts in Mongol Eurasia”.

The aforementioned digital database enables the usage of sources from various languages (e.g., Arabic, Persian, Chinese, Syriac, Armenian and Latin) in order to locate relevant data on the individual merchant as well as on the merchants as a group. The accumulated information on these merchants, obtained from both primary sources and secondary literature, will be put together, indexed, analyzed and visualized through the use of this digital database, thereby enabling me to reconstruct the commercial networks of the Ilkhanate.

Reconstruction of the above commercial networks will provides a better understanding of the relations among the Ilkhanate and the Mamluk Sultanate (1250-1517) of Syria and Egypt, the Chaghadaid Khanate (1260-1678) of Central Asia, the Golden Horde (1260-1480) of Eastern Europe, the Yuan dynasty (1271 -1368) of China, as well as South East Asia, India , Africa, Yemen, Anatolia, Cilicia, Western Europe and the Italian city-states of Genoa and Venice, thereby manifesting the full scope of Mongol trade.

* The Hebrew University of Jerusalem.

The War between the two Capitals Revisited

A Digital Humanities Supported Prosopographical Perspective

Vered Shurany*

Abstract

Mongol and non-Mongol generals played a leading role in the conquest of the Southern Song Dynasty (南宋朝 1127-1279), the establishment of the Yuan dynasty (元朝 1271-1368) and the unification of China after more than three hundred years of disunion. Less than a century later, the military commanders became once more important but this time as players in the court's political intrigues. The changes in the role of the Yuan commanders became especially clear in the civil war that broke on August 1328 known as the "War of Restoration". After the death of Yesun Temur, (Taiding 泰定, r. 1323-1328) a succession struggle began between the supporters of Yesun Temur's son in Shangdu (上都) and the supporters of Qaishan's (Wuzong 武宗, r. 1307 - 1311) son Tugh Temur (the future Wenzong 文宗 r. 1328-1332) in Dadu (大都). This war turned out to be the most deadly and destructive civil war in Yuan history.

My presentation seeks to reexamine this turning point in the Yuan military system through a prosopographic lens by examining biographies of known and lesser-known Yuan commanders and soldiers who participated in the civil war. These biographies will be read, translated, indexed, compiled and analyzed through the use of online prosopographical databases such as the Jerusalem-based ERC project "Mobility, Empire and Cross Cultural Contacts in Mongol Eurasia", and the Harvard University's "China Biographical Database Project" (CBDB). The Jerusalem Database also enables the researchers to locate information on these Yuan generals from non-Chinese, mainly Persian, sources.

In this paper, I will map these commanders' and soldiers' movements and visualize their social networks, in order to demonstrate how the usage of prosopographic analysis provides a better understanding of the structure, function and administration of the military in the most important Mid-Yuan civil war.

* The Hebrew University of Jerusalem, Email: vered.shurany@mail.huji.ac.il

Experts, Institutions, and Networks: An Uighur Astronomer in the Yuan Court

Qiao Yang*

Abstract

Calendars as well as astronomical and astrological predictions had important political and ideological implications on the Yuan dynasty. The Mongols in China combined this Chinese perception with their own political tradition of astronomy. On the one hand, the Yuan dynasty adopted a Chinese astronomical system, which the Chinese astronomers significantly improved in the 1270s and 1280s. On the other hand, the Mongols brought non-Chinese astronomers to Yuan China in order to help them rule a multi-ethnic and multi-cultural population in China as well as balance the political influence of Chinese scholar-official astronomers. These foreign astronomers, like other foreign experts serving the Yuan court, are usually underrepresented or misrepresented in Chinese sources. This forms the biggest challenge for studying the role the non-Chinese played in the Yuan and for understanding the pluralistic nature of the Yuan administration, politics and culture.

This paper attempts to overcome the limitation and bias of sources with the database “Mobility, Empire and Cross Cultural Contacts in Mongol Eurasia”. I will reconstruct the biography of the Uighur astronomer Aruhun Sali (阿魯渾薩理 A-lu-hun sa-li), whose significance in Yuan astronomical institutions, politics and culture has been underestimated because of the scattered information on him and his misrepresentation as Sinicized in his biography in the Official History of Yuan (元史 Yuanshi). I will also analyze the social networks of Aruhun Sali and highlight his relationship with the Mongol rulers and his Chinese and non-Chinese counterparts. Aruhun Sali will serve as a case study to demonstrate how the database serves as an efficient tool for combining pieces of information to reconstruct the biographies of non-Chinese experts, and more importantly, for in-depth reading and analysis of the biographic information.

* The Hebrew University of Jerusalem, Max Planck Institute for the History of Science.

**Genealogy between History and Text:
The Digital Humanities' Quest in Researching
Multi-Generational Chinggisid intermarriages in
Mongol Eurasia**

Ishayahu Landa*

Abstract

The paper discusses the advantages and disadvantages of the digital approaches in the context of historical genealogical research. The case study is the reconstruction and analysis of the multi-generational power networks of the different tribal and ethnic elites and their relations with different Chinggisid clans in Mongol Eurasia (13 – 14 centuries). The paper will discuss three major challenges that genealogical research has to cope with: a) the multiple cross-Eurasian human migrations during this period, due to which even small groups (tribes, clans, separate lineages) dispersed across thousands of kilometers in a short period of time; b) the complex multi-layered, levirate- and sororate based networks of the marriages between the Chinggisid rulers and the military nobility, and c) the biased and partial nature of the available literary and non-literary sources from the various cultural and linguistic contexts. Could the usage of the digital humanities' approaches and technologies themselves solve these problems, and help scholars in reconstructing the missing links in the genealogical research? The presentation will show, that only through the mutual usage of three digital approaches – “big data” analysis, usage of the TEI or similar textual markup and network visual reconstruction (SNA) – can the set of problems mentioned above be solved at least partly. Simultaneously, the presentation will also argue that all of the above mentioned approaches are useless for the historical research's purposes without a) the establishment of the text-based (and not record-based) relational database and b) without any efficient cross-checking mechanism of the biographical data made by a human researcher.

* The Hebrew University of Jerusalem.

基於頭戴式裝置與行動裝置平台之虛實互動式導覽：

以花蓮牛犁社區為例

周子凱*、詹進發**

摘要

虛擬實境(Virtual Reality, VR)及擴增實境(Augmented Reality, AR)嘗試讓體驗者如同在真實世界進行感知，不同於以往純粹視覺性呈現，進一步延伸至多元化的人機互動(Human-Computer Interaction, HCI)方式。本研究目的為以價格合理及普及性高之行動裝置搭配 Gear VR 頭戴式顯示器(Head-mounted Display, HMD)並結合手勢感測器 Leap Motion，嘗試增進使用者體驗設計穿梭虛實的感受及多種虛擬探索方式，建置大稻埕歷史街區的虛實互動式導覽平台。

本研究以跨平台遊戲開發引擎 Unity 設計虛實穿梭及虛擬探索的體驗。虛實穿梭部分，由於 Gear VR HMD 搭載行動裝置平台，因此透過開關行動裝置之相機鏡頭並搭配 Vuforia 擴增實境軟體開發套件(Software Develop Kit, SDK)達到虛實穿梭的體驗。虛擬探索部分，由於 Leap Motion 無法與行動裝置連接，因此以 Raspberry Pi 3 為伺服器端(Server)透過 Socket 傳輸手部資訊至 Gear VR HMD 搭載行動裝置之用戶端(Client)並依照大稻埕歷史發展沿革分為四大虛擬探索體驗：(1)發源 稻香四溢、(2)械鬥 頂下郊拚、(3)開港 徜徉河畔、(4)舊址重現臨床講義。

關鍵字：虛擬實境、擴增實境、人機互動、手勢感測器、遊戲開發引擎

* 國立政治大學地政學系碩士生，Email: 105257029@nccu.edu.tw

** 國立政治大學地政學系教授，Email: jfjan@nccu.edu.tw

輔助社群媒體打卡研究之分析工具研發

李蔡彥*、梁芷瑄**、吳筱玫***

摘 要

打卡(Check-in)是 Facebook 平台上使用者經常使用的功能之一。過去關於打卡的研究大多採用質化的方法，而質性研究者在訪談使用者之前，為了解使用者的資料，往往需要手動查看使用者的 Facebook 塗鴉牆，收集、整理資料，費時費力。另外，使用者在 Facebook 上的打卡方式具多樣變化，許多的打卡不具有即時性與適地性。為了了解使用者的打卡動機，我們尚需取得使用者在手機上的操作與位置等資訊，方能還原打卡時的情境。為了了解打卡研究者的研究歷程與需求，我們讓質化研究先行，以進行需求分析，再依其開發一個協助研究者的分析工具，收集整理來自 Facebook 打卡資料與手機 Log 的資料，透過視覺化與列表的方式呈現，提供研究者能快速、深入分析、探索使用者在 Facebook 上打卡的行為與動機。本實驗邀請 5 位受試者扮演打卡研究分析者的角色，透過系統教學讓受試者學習使用系統，最後讓受試者自由探索使用者的資料集，並記錄下探索歷程與發現。實驗的評分結果使用 5 分量表，有用性向度的平均分數為 4.6，受試者認為本系統能協助他們分析使用者的打卡行為與進行後續研究；易用性向度的平均分數為 4.1，系統的操作方式對有的受試者需要時間來學習，但大部份受試者仍對本系統的易用性表示同意，證明本系統兼具有用性與易用性。另外，我們也發現受試者在探索過程與探索結果中展現了對此系統的創用性，是在設計者的預期之外，可見本系統的工具本質。

關鍵字：打卡、社群媒體分析、資料視覺化、臉書、手機紀錄

* 國立政治大學資訊科學系教授，Email: li@nccu.edu.tw

** 國立政治大學智慧型媒體實驗室，Email: yugixyuki@gmail.com

*** 國立政治大學傳播學院教授，Email: smwu@nccu.edu.tw

自然語言處理技術在數位人文研究中的應用

梁繼文*、葉文豪**、王東波***

摘 要

數位人文由人文計算發展演變而來，是一個將電腦學科與人文學科相結合的新興交叉領域。本文在對數位人文和自然語言處理進行概述的基礎上，指出自然語言處理技術是數字人文中重要的研究手段，並回顧了自然語言處理技術在近年來國外開展的數位人文研究中較有代表性的幾類應用與實踐，對其研究物件和方法進行詳細梳理，以期為我國人文學科發展提供啟發及參考。

關鍵字：數字人文、自然語言處理、人文研究

* 南京農業大學碩士研究生，Email: 2016114008@njau.edu.cn

** 南京農業大學碩士研究生，Email: 2015114008@njau.edu.cn

*** 南京農業大學副教授，Email: db.wang@njau.edu.cn

中文斷詞器的後設與理論意涵：語用學與社群的觀點

劉慧雯*

摘要

一、前言

近十年來，社群媒體發展、使用的狂潮席捲了人類社會生活的每個面向，從行銷公關，到政治參與，再到自我形象構塑或人際互動，社群媒體可說無役不與。這種現象也造就了學術研究上對社群媒體材料的高度興趣。由於社群媒體上的素材由使用者製造 (user-generated)，且不論是哪一種社群媒體，皆以數位編碼為其基本物理型態，因此，以計算機進行社群材料的分析，就成為近三年的顯學。

根據項潔、涂豐恩 (2011) 的說明，對數位化資料的研究工作，是以「對既有資料挖掘新的意義」，由此，在數位化檔案中進行查詢、交互比對、關鍵字追索，乃至於借重視覺化對大量資料的簡潔、直覺化呈現，就成為數位人文領域中最核心的研究模式。對社群媒體的研究者來說，由於由使用者生產的文件、檔案、意見數量豐厚驚人，且參與者眾多，若要理解社群媒體上的人際網絡或文本意義，就不得不借助計算機強大快速的計算能力。其中，對社群媒體上的文本進行分析，是本文特別關注的對象。

二、研究問題：中文斷詞器功能設定中的後設意涵

針對社群媒體材料的研究主題，相當程度集中在關鍵詞與語詞頻分析 (frequency analysis)、文字叢集關係，以及「情緒分析」(sentiment analysis) 等工作。關鍵字/詞分析工具能夠找出文本中常出現的詞彙；為了呈現重要性排序，關鍵詞通常會以「詞頻」呈現 (即，字詞出現的頻率排序)。由於語言使用是成串的線性模式，因此，與特定字詞一起出現的其他字詞，就成為界定該字詞意義的重要管道；在資訊工具上，「共現分析」(co-occurrence analysis) 可找出詞彙與詞彙間的關係。「情緒分析」可指出人們談論特定議題時之態度；由於態度具有方向性，因此也常被稱為「輿情分析」。為了進行這些分析工作，對中文材料的研究者來說，進行文本分析的第一步驟就是以合理合適的方式將文本進行「斷詞」。

在英文的語境中，「單字」(word) 所帶有的意涵，與中文所稱「詞語」較為接近；而中文的「字」(character)，多指一個最小的表義單位。一個中文表義單位到底應該與在他之前的字理解在一起，還是後一個 (或者前後多個) 是語言學上的繁浩工程。單就我們對使用者生產內容的好奇來說，若不能成功的「斷詞」，就有可能蒐集、計算錯誤的對象。例如「全台大停電」這短短五個字，便至少有「全，台大，停電」以及「全台，大，停電」兩種斷法。從斷詞所能形成的意義

* 國立政治大學傳播學院副教授，Email: moskito.liu@gmail.com

落差來看，斷詞確實是中文意義探究的關鍵第一步。

斷詞工作在資訊科技領域的概念是「斷詞/分詞器」。針對簡體與繁體中文的分詞器，至少有十數種工具。目前台灣研究者最常用的中文斷詞工具是中研院資訊所與語言所合作，歷時多年開發出「CKIP 中文斷詞系統」（參見：<http://ckipsvr.iis.sinica.edu.tw/>），以及初始由百度開發者撰寫，後因使用開放原始碼而由台灣工程師加入繁體中文字典的分詞器「結巴」（Jieba，參見：<https://speakerdeck.com/fukuball/jieba-jie-ba-zhong-wen-duan-ci>）。CKIP 因為有中研院語言所的語言學家長期投入，（號稱）分詞正確率可達 95%-96%（參見：<http://ckipsvr.iis.sinica.edu.tw/apply.htm>）。而「結巴」則因具有自建字典的功能，特別受到特殊語料研究者的關注。

如果考察中文斷詞器的元件可以發現，理論上，以詞庫為斷詞基礎的系統，都會受到詞庫本身的豐富性影響。過去，像是 CKIP 這樣的系統多以制式的內容，如：典籍、課本、新聞等，作為建立詞庫、機器學習的藍本。這是因為這類內容的詞語使用規則清楚可循，對於講究規則的機器學系體系來說，是較容易訓練的模式。考慮到詞庫本身的侷限性，CKIP 設計了「未知詞」欄位，將分析範例中出現頻次高，卻無法自詞庫中比對出來的詞彙羅列於未知詞欄位。理論上，只要詞庫夠豐富，未知詞的數量應該會很少。

相對於 CKIP 的未知詞欄位一方面保留了詞頻高卻未能符合詞庫的字，一方面同時維護了詞庫的神聖不可侵犯性，坐實了詞庫的「字典」地位。相對的，Jeiba 在協作前提下，發展出「自定義詞庫」。Jeiba 繁體中文版的主要撰寫人林志傑便指出，自定義詞庫使得個別使用者可以以小規模材料為範疇，建立分析核心詞彙；例如分析歌詞後，建立情境歌單、為創作者設計自動填詞功能、為聽眾提供相似歌詞推薦等功能。以上這些應用的關鍵可說是客製化，也可說是「每人一本自己的字典」：各人字典。

由以上針對 CKIP 與 Jaiba 這兩個斷詞器對「詞庫」的設定來看，斷詞器本身涉及了對語詞字典合法判定標準。不僅如此，隨著數位典藏、數位人文乃至於社群媒體研究的發軔勃興，內建於各個套裝軟體、分析器中的斷詞器，也與相關的其他功能，如：視覺化、關鍵字文本分布等，產生互動。

本文將由 CKIP 與 Jaiba 為基底，討論目前較常見的中文數位文本分析器/軟體在斷詞上的介面與功能設計，探索其中可能的後設與理論意涵。本文並試圖從語用學（pragmatism）的角度出發，討論這些分析器的後設與理論意涵對社群媒體文本分析研究可能產生的偏向或後果。

三、研究架構與預期成果

本文將至少分析 CKIP, Jaiba 以及包括「庫博」（Corpro）或 Gephi 在內的數個中文斷詞器或文本分析套裝軟體。我們將分析區分為兩個層次。第一層次是「斷詞器」的功能設置；在這個層次中，主要討論以斷詞為主要，甚至是唯一功能的斷詞器，如何建立、運作斷詞邏輯。我們同時從社群媒體研究者的角度出發，

以一組長久以來仰賴人工分類的語料為測試標準，討論不同斷詞器對該語料的斷詞結果。這組語料包括合理的多重分類結果（如：「全台大停電」），以及必須考量語言使用習慣與脈絡條件（如：「阿不就好棒棒」）等材料。我們的目的不在毀壞斷詞器的分類準確度，而是試圖從語用學的角度說明斷詞工作中可能遭遇的困難。

第二個分析層次將針對分析套裝軟體進行與斷詞器相應相連的其他文本分析功能（如：共現、情緒分析、視覺化等）。在這個層次的分析中，我們將側重「斷詞之後能從事的研究行為」，以說明斷詞在中文數位文本中的關鍵地位。同時，我們亦將以前述測試語料為標準，試圖描繪這些帶有高度模糊性的測試語料需要經過哪些條件、調整，才能在文本分析套裝軟體中指示出與人工分析類似的結果。

在以上的工作中，我們希望討論中文斷詞與文本分析結果的理論意涵。

關鍵字：中文斷詞器、文本分析、語用學

History of Governmental Funding in Shenzhen: Analyzing the R&D Structure and Effects through the Spatial Analysis of Metadata of Research Literature

Yawen Zou*

Abstract

With the digital turn bringing big data and computational tools to analyze those data, humanities are becoming more like science or social sciences in several regards. First, although the subject that humanists study may be different, for instance, poetry, classics, images, or sounds, most digital humanities projects produce papers that follow a methods-results format, like a scientific paper. Second, humanists would find it useful to learn digital tools, code by themselves, or work with computational scientists. Third, statistics or mathematical modeling is not unusual in a digital humanities project. In the deluge of data created every day, mastering these approaches seem inevitable, rather than ancillary.

This paper uses the case study of history of governmental funding in Shenzhen, China to demonstrate how a history project can benefit from the digital. Shenzhen is at the forefront of Reform and opening-up policy and has witnessed great success in the past four decades. The city is in the Pearl River Delta, one of the most prosperous areas in China. Universities are important in the innovation system. Compared to Beijing and Shanghai, due to historical reasons, the number of universities in Shenzhen are less and the contributions are not as significant. However, the is changing in recent years. The Municipal Government of Shenzhen, and other firms have dedicated tremendous R&D funds to support the research centers based in the city, which include high-tech zones, incubators, universities, research institutions, and companies. In 2005, the R&D expenditure of Shenzhen city is 73.238 billion RMB, and the governmental expenditure is 3.556 billion RMB. Although firms, like Huawei Technologies and Zhongxing Telecommunications (ZTE), Tencent, have invested billions in R&D, this study focuses on the funding from the Shenzhen government, because the transparency of governmental funding is higher than that of entrepreneurial funding.

* Chinese University of Hong Kong, Shenzhen, Email: yzou20@asu.edu

As the R&D funding is increasing every year, it is greatly important to assess the effects of governmental funding to justify the spending for taxpayers. This study addresses key questions, which include: what funding agencies belong to the Shenzhen city and who, from where, is receiving the money from those agencies? How many projects are exclusively funded by Shenzhen? How well do those research projects perform, as reflected by the journal quality and citation count? Has the regional inequality in different districts been intensified or ameliorated over the years?

This study addresses these questions by analyzing the metadata of research articles funded by agencies of Shenzhen and deposited in the Web of Science database. Previous studies have shown great advantages of using GIS technology in humanities research and social sciences, such as in archaeology, history, music, media studies, and literary studies. GIS can offer great tools for analyzing the distribution of funding, because each fund recipient has a physical location that can be mapped by GIS. Many studies about funding often have two limitations. First, they cannot offer a visual and straightforward representation about the places of researchers who receive the funding. Second, they often offer a general, rather than nuanced, description of the overall pattern; they either offer a qualitative interpretation or contain too many numbers. Such analyses may not be suitable for policy makers or researchers on policy to gain insightful knowledge. The GIS research can overcome these limitations. GIS offers new representations utilizing big data and has complex information underlying each map. A single map can show the situation of many regions, instead of one region.

A starting point for this study is to collect the metadata of articles deposited in the Web of Science Core Collection, with the word “Shenzhen” in their “Funding Agency”. This means the articles have at least one funding agency from Shenzhen. First, we catalogued the funding agencies associated with Shenzhen. Second, we retrieved information from the metadata and created a database that shows the corresponding author’s geographical location, city, district, funding agencies, and publishing year. Third, we geocoded their geographical location and mapped the publications on maps using different colors and shapes, based on funding agency, publishing year, district, respectively. Next, we analyze the statistics longitudinally and interpret the results. Meanwhile, traditional close reading of literature is not considered obsolete but is still used to facilitate the interpretation of the data.

Our study shows various funding agencies, projects, and plans, including Shenzhen Peacock Plan, Shenzhen Basic Research Project, Shenzhen Science and Technology Innovation Committee, in addition to national funds, such as National Natural Science Foundation of China, and provincial funds, such as Natural Science Foundation of Guangdong Province, have greatly boosted the research output of Shenzhen over the years. This is reflected in several aspects. First, the number of

publications funded by Shenzhen is increasing tremendously. Second, more institutions in Shenzhen receive the funding over the years. Third, there is a tremendous regional inequality in terms of research funds that are received by each district in Shenzhen. Nanshan District enjoys a great advantage, partly because that's where Shenzhen University, Southern University of Science and Technology, and University Town of Shenzhen, where THU, PKU, Harbin Institute of Technology, with other universities, have set up campuses. However, Longgang district, has been catching up very quickly in recent years.

Although publications are a major output of research funding, the author does not intend to claim this research captures all the research output, which include other output, such as patents or the setup of research facilities and infrastructure. Also, this research assessed research output biased toward higher-end publications, because the publications in the Web of Science database are from reputable English journals.

This study not only has implications for the regional planning of Shenzhen city to transition into a knowledge economy but can also shed light on the funding structure of other cities in China. This study offers a case study of scholarship that highlights the importance of bibliographical and geographical information in the humanities and social sciences, especially historical studies. The methods outlined in this contribution can be relevant for other humanists, who want to write the history of other subjects of their interests.

Keywords: GIS, Funding, R&D, Metadata

Walking into a Brave New World:

Visualizing Character Paths through Shakespeare’s Plays

Brian Kokensparger*

Abstract

There are 38 plays attributed to William Shakespeare. Digital Humanities allows researchers to question traditionally-held theories based on “close” readings by extending research to “far” reading. In this paper, the researcher uses the Ubiq+Ity tool provided by the Visualizing English Print (VEP) Website to analyze the spoken text of all major characters within all 38 of Shakespeare’s plays. The Ubiq+Ity tool provides several Language Action Types (LATs), which are then classified as relational versus selfish, and active versus reflective. These classifications produce a visualization of the paths the characters “walk” through the plays, leaving their trails behind that can then be visually analyzed. Upon initial analysis, using the play *As You Like It*, minor characters such as Phoebe and Silvius revealed dramatic arcs in their speeches. Rosalind’s two personas (as Ganymede and as herself) started differently but grew to similar-shaped paths. The three sets of couples who are married at the end of the play demonstrated different dynamics giving further insights into their relationships, and perhaps into what Shakespeare wanted to say dramatically. Some areas for further analysis and tool development are suggested.

Keywords: Corpus Linguistics, Textual Analysis, Data Visualization, Early Modern English Drama

* Department of Journalism, Media & Computing of Creighton University, USA. Email: bkoken@creighton.edu.

Revelation of Prehistoric Architecture through Augmented Reality: Representation of the Sanctuary at Tas-Silġ

Shih-Chueh Kao*, Vince Briffa**

Abstract

Many prehistoric sites and architectural ruins have not survived the ravages of time and have lost their visual appearance, hence also losing what we today consider important historical data. Time has severed their original connection with their surroundings, causing an altered context with their intended surroundings.

From a representation and an experiential perspective, it is meaningless for visitors to look at such a site without prior knowledge of its original aesthetic and topographic context. It is common practice for the representation of such important heritage sites to commonly apply photographs, videos, drawings and physical models to emulate their original look and feel. However, such visuals usually demonstrate limited and fragmented information, and may assume that the viewer is already informed of the architectural language of the time and is able to interpret the vicissitudes undergone by the site.

Augmented Reality (AR) allows developers to add virtual objects onto real world objects via viewing through mobile or head-wearing devices, which provides a great potential in digital heritage representation. There have been studies carried out on ancient heritage and Virtual Reality (VR) applications and they are categorised as: i) works running on mobile devices to perceive the virtual models blended with physical models. For example, Gongli Li et al. (2013) developed a mobile application to reconstruct Yuanmingyuan in viewers mobile device by AR technology. ii) Mixed Reality (MR) works applying VR interface to create different perceptions in the physical environment. Such as the work by Martin White et al. in 2007, the virtual interface of the museum applied various multimodal interaction techniques and allowed the user to better explore heritage information. iii) works presenting an interactive information representation. Derya Güleç Özer et al. (2016) employed Multirama, a software of developing virtual information in AR for scale architecture model, to

* Department of Digital Arts, University of Malta, Malta. Email: shih-chueh.kao@um.edu.mt

** Department of Digital Arts, University of Malta, Malta. vince.briffa@um.edu.mt

represent Parion Theatre with AR architectural information. The approach utilised in this paper can be considered as falling within the first group. A mobile application will be built by employing Vuforia, an AR software development kit for mobile devices, to represent a virtual prehistoric architecture when viewing through the camera of a mobile device. Vuforia applies Computer Vision technology to recognise and track target object (2D images or 3D objects). In this work, the viewer will look at a 3D printed physical model of Tas-Silġ through a mobile device's camera to trigger the virtual sanctuary that was built on the site between 200-100 BC. The virtual object then tracks the position and orientation of the physical model in real-time so that the viewer's perspective on the object corresponds with their perspective on the target object. With the freedom of moving the mobile device around the physical model, the viewer can decide which part and how close to look at the details of the virtual heritage object. The implication of this work is to demonstrate a more immersive and interactive viewing experience of digital representation by using AR technologies. The easier perception will give the viewer a more clear idea of the look and feel of the original site.

Tas-Silġ is Malta's only prehistoric site to be used by the Phoenicians for their own religious needs. In the last two decades, archaeological research on the sanctuary at Tas-Silġ has documented an interesting context for this long forgotten influential religious centre. Built on the top of a low hill near Marsaxxlokk bay in the south of Malta, the sanctuary was located on an imposing structure dating back to the Tarxien phase of the Maltese Late Neolithic period (2500-3000 BC). The part that still remains of the main temple building is a large lobe with a curved front and inner apses. For general visitors, these fragmental structures invigorate archaeological significance but connect nothing to the original architectural look. According to the report published by The Missione Archeologica Italiana in Malta in 2011, the architecture of the sanctuary between the end of the 2nd and the beginning of the 1st century BC was restituted with the peristyle-court by illustration. The figure outlines that the central area of the sanctuary was monumentalised by the building of a peristyle-court enclosing the sacrificial space and the temple itself. In the light of this information, this study will create a digital 3D model based on the current structure of the site and the 2D visualisation of the architecture to reveal the original look of the ancient sanctuary.

This study will carry out technical works in Vuforia and site prospection simultaneously to commit the scale and the proportion of the digital model in accordance with the site. With 3D object tracking, Vuforia combines the virtual object, the sanctuary, and the physical object, 3D-printed site structure, through augmented reality. The viewer looks at the scale model through the camera of a mobile device and the virtual sanctuary will be placed and combined with the scale model. For the scale model, the measurements of the site structure are referenced in Autodesk Maya, a 3D

modelling and animation application, to model the structure accordingly. The finished high quality polygon model will then be converted into .OBJ file format for 3D printing with Polylactic Acid (PLA) material, and the resulting output will be polished and painted in necessary colour to represent the scale site structure. The sanctuary will be built on the site structure model that was created previously in Autodesk Maya. The top view and the side view of the sanctuary will be developed based on the illustrated visualisation in the previous work to support the modelling of the architecture. The resulting sanctuary model will be imported in Unity, a game engine application, as an asset to be launched by Vuforia when the camera detects the target object (scale model). Subsequently, a mobile application 'Tassilg_AR.apk' will be built by using Unity and installed in either Android and Mac mobile devices. While viewing in an exhibition, the viewer will hold the smart device that has installed the application 'Tassilg_AR.apk' to look at the physical site structure model, the application will then trigger the virtual sanctuary object in the mobile display. The viewer then have the freedom to look around the architecture and decide which part he or she likes to focus on for details.

Practices in this work are not only to validate that the heritage representation by AR has its great potential to be adapted in exhibitions, museums or libraries, but also to implicate that a better understanding of the original look of the heritage site will support archaeologists, historians and restorers to have more precise analyses and interpretations in relative studies.

Keywords: Digital heritage, Augmented reality, Heritage representation, Tas-Silg

Research on the Teaching and Learning of Using 3D Motion Capture Analysis on Chinese Opera Performance

Tai-Jui Wang*

Abstract

The difference of performance movements are a frequent reality for training Chinese opera performers. Many aspects of the acting training make performers particularly susceptible to being different with each other's. Thus, the process of 3D motion capture could investigate the data of movements in between the instructor and students.

The purpose of this study focuses on a comparison of Chinese opera performance movements' difference for the professionals and students in training place. For the training issues, the "Eight Attributes" are the foundation of performing skills in Chinese opera. It concerns the manner of hands, eyes, head, feet, legs, and body which present the movements of a Chinese opera performer's physical capabilities. Therefore, the aim of this paper is directly towards one of three types of disciplines, which can be describe as "Basic Exercises" (Ji-ben-gong). It is the most important foundation in the training for traditional Chinese Opera curriculum.

For the methods, the researcher use PERCEPTION NEURON, a motion capture equipment, for capturing the movements of instructor and participants. A Neuron houses an Inertial Measurement Unit also known as an IMU, with a gyroscope, accelerometer, and magnetometer. The experimental processes consists of two phases. The first phase of experiment are using pre-experimental design in one-shot case study. The second phase of experiment are using true-experimental design in randomized control-group pretest-posttest design. We used motion capture to track the movements of an instructor and nine voluntary Chinese opera students as they performed a series of basic exercise movements. After a self-directed warm-up, subjects were given brief descriptions of each movement, and then asked to perform the movements. Thirty-one (31) IMU (Inertial Measurement Unit) markers on the whole body were used to track the movements of the subjects.

* Department of Mass Communication, Chinese Culture University, Taiwan. Email: wtd@pccu.edu.tw

The subjects and conditions randomly and voluntarily participated. Furthermore, an electronic data analysis program are used to facilitate analysis. The main program used are SPSS 18 because of its versatility in statistical analysis. For the calculate methods, the bivariate correlation, paired sample T-test, ANOVA, IPA (Importance-Performance Analysis), and multiple regression are used at a two-tailed significance level of .05 for comparison of the movements of instructor and participants variables such as right hand, left hand, hips, right foot, left foot, and head.

Furthermore, The analysis here is just to look all the data from the captured movement and to display all three values for each of the joints against each other of the three (x,y,z). A total correlation of right hand, left hand, hips, right foot, and head of velocity, acceleration, and angular velocity in 2.5 seconds. The qualitative grades for nine subjects has given from instructor and put in the table for comparison with quantitative results. The most important specific moment that Chinese opera performance called “liàng-xiàng” (posing) showed in figure as X axis and Y axis. The graphics indicated that the differentials of instructor and nine subjects. The shortly timing caparisoned minor changes of velocity, acceleration, and angular velocity in between 0.3 to 0.7 seconds. The spatial patterns showed comparing the same movements, done by both instructor and nine subjects. As the scatter showed, the patterns of right hand from each subject were different with instructor. These dot-to-dot path tracking could easily see how different that the movements of instructor and nine subjects.

Therefore, 3D motion capture can use for the future application of identifying detailed factors through a live demonstration involving visual blind spots and a rotating 3D doll perspective, to enhance the accuracy of movement observation. Analyzed data from 3D motion capture can benefit training. A database analysis could applied to quantify and qualify cross-examination; individualized guidance can enable professionals and students to mitigate the problems of differences. Hopefully, the database of Chinese opera performance movements will be established and not only supported teaching and learning in the real training field but also applied for VR (virtual reality), AR (augmented reality), and MR (mixed reality) in the near future.

網絡分析方法在音韻學教學中的應用： 以《廣韻》反切系聯為例

胡佳佳*

摘 要

音韻學是中國傳統“小學”的一個分支，與文字學（形），訓詁學（義）相依而並列。它的研究材料是古代的韻文，中古的韻書，以及古代文獻典籍中存留的語音現象和訓釋材料。作為古代漢語專業研究生必修科目之一的音韻學一直是最令老師和學生望而生畏的一門課程。其中一個根本原因就在於它所面對的是漢字記錄的書面語，並用漢字作為工具來分析語音。這也造就了傳統音韻學研究的一大特點，那就是音類關係的建立要比微觀音值的構擬更為重要。而進行音類研究主要的方法之一就是“系聯”法。

關鍵字：網絡分析，音韻學，反切系聯

* 北京師範大學文學院，Email: hjj81@126.com

《一切經音義》全文檢索資料庫的構建： 以日本所藏十種古寫本為例

劉冠偉*、李乃琦**

摘 要

公元 645 年玄奘大師設譯經院，翻譯其從印度帶回的大量佛經。譯經院中的字學大師玄應將譯經中的難解字詞另行闡釋並整理成書。此書正是成書於 661 年的《一切經音義》，也是現存最古的佛教辭典。全書共 25 卷，40 餘萬字。其中收錄的 450 餘部經目，為研究唐代所通行的佛經內容和類型提供了真實的依據。另外其就 9,000 餘字詞進行的闡釋，成為研究唐代長安音的寶貴語料庫。

隨著佛教的傳播，《一切經音義》作為佛教工具書也被廣泛傳抄，目前有刻本和寫本（抄本）兩種形式。刻本印於宋代之後，大部分藏於中國。寫本包括日本所藏古寫本以及英法德俄所藏敦煌·吐魯番斷片群。《一切經音義》的刻本和寫本在書寫形式和內容上有較大差異，根據成書年代推測，寫本更接近《一切經音義》的原本型態。但日本十種古寫本至少分為兩個系統，且各系統間有較大差異。由於文本內容數量巨大，這些差異的具體情況和產生原因無法簡單歸納。因此，將數位人文引入本課題是勢在必行。

近年，隨著 Open Data（開放資料）與 Open Science（開放科學）的興起，典籍的數位化已成為人文研究的重要方法，其中佛學和語言學等研究領域內亦有卓越的成果。本研究以日本所藏十種《一切經音義》古寫本為研究對象，通過其全文檢索資料庫的構建，對各個寫本間相異內容進行梳理，以期窺探《一切經音義》成書時原貌，以及傳入日本後的受容和變容，為佛教史和中國語言文字史的研究提供切實的依據。

* 日本北海道大學，Email: toyjack@gmail.com

** 日本北海道大學，Email: linennki0527@yahoo.co.jp

爭取無票者的民意：香港特首候選人的社群印象管理戰術

江建璿*、黃珮綦**、傅威***

摘要

香港的特首選舉在《基本法》的管轄之下，因其複雜且不透明的制度而被稱為「小圈子選舉」。儘管 2017 年特首選舉未落實民主普選，但選前的戰場規格卻透過網路拉抬至整個民間社會，「無票者的聲音」開始浮上檯面，甚至積極幫助候選人「輔選」。三位候選人（曾俊華、林鄭月娥、胡國興）的 Facebook 競選專頁皆有數萬以至數十萬的追蹤者，顯見大多數香港人手中雖然不握有選票，仍關注候選人在社群媒體上的言說甚至頻繁地投入社群選戰，也期待透過民意影響選舉結果。在《基本法》與香港人期望民主的矛盾間，三位背景與立場各自不同的候選人在社群媒體上的活躍行動，也提供了香港市民參與選舉過程的平台與機會。

本研究關切旨趣為，第五屆香港特首選舉的候選人透過社群媒體的言說建立起何種形象，在印象管理戰術（*tactics of impression management*）過程中如何回應政治背景與現況，以及香港人渴望民主或政改的情緒，進一步爭取「沒有實質投票權的民意」。透過人工撈選與語料庫，本研究旨在分析(*corpus analysis*) 三位特首候選人於 Facebook 粉絲頁上的大量貼文內容，檢視候選人如何透過其所偏好的詞句建構形象，最後以批判言說分析（*critical discourse analysis*）取徑，從主題、策略及語義層面探究三位候選人的形象，發揮印象管理戰術（*tactics of impression management*）如何對應到香港政治氛圍下的衝突與爭議。

本研究的分析對象為香港 2017 年的特首選舉候選人：曾俊華、林鄭月娥與胡國興的臉書粉絲專頁貼文作為分析語料。研究者蒐集從 2017 年 2 月 14 日至 2017 年 3 月 25 日，即選舉提名日至投票日前的競選期間，總計貼文為 6 萬 4 千 749 個字建構語料庫。本文以 Wordsmith 6.0 軟體輔助語料庫分析後，進行批判言說分析。首先，研究者引用中研院發展的斷詞軟體系統進行中文斷詞（其詳細簡介見可參考其網頁：<http://ckipsvr.iis.sinica.edu.tw/>）；接著運用軟體系統，從詞頻基礎進一步分析詞頻統計分析（*frequency list*）、併置詞（*collocation*）及共詞句（*concordance line*）分析（本文研究架構詳見圖一）。其中併置詞（*collocation*）部分，以 t-檢定（*t-score*）探究某一樣本分佈平均數和整體分佈在統計上的顯著意義，研究者預設 t-檢定顯著水準為 99.99%（ $p < .001$ ），依樣本達到顯著程度

* 政治大學傳播碩士學位學程，Email: kkchung87@gmail.com

** 政治大學傳播碩士學位學程，Email: pi0310@gmail.com

*** 政治大學傳播碩士學位學程，Email: weifu2005@gmail.com

的字彙亦列入本研究討論。綜合上述，本文透過語料庫分析與批判言說分析取徑，透過字詞頻率、共詞組合及本文的脈絡，探討本研究關切的三位香港特首候選人在競選期間如何透過「語義偏好」與特定詞彙，管理與建構其形象爭取民意，以回應本文研究問題。

關鍵字：香港選舉、特首、印象管理、語料庫分析、批判言說分析

經典與注疏之數位化系統設計與建置： 以「《成唯識論》及其注疏編撰」為例

洪振洲*、蔡伯郎**

摘 要

在佛學研究領域中，多文本間的對讀以及單一文本之於諸多注釋書的解讀，是一件基本、必要而且相當重要的工作，佛教經典在長時間的傳佈、弘揚中，同一部經、論常有多種不同的轉譯本，包含漢譯與藏譯，而這些不同的譯本間，可能是因為譯者翻譯時的理解與用語不同，或是其所依譯的原本不同，常常呈顯出大小不同的思想上的差異，因此在掌握某一時期或某一學派一部經論的思想時，對於其同本異譯間的比較對讀，便成為一個基本的研究工作。另外，由於一部經論常常是內蘊的思想豐富，而文字卻極為精簡，因此在解讀上若無補述的注疏本，則對於後人來說，研讀起來是極為吃力的，以此注釋書對於經論文本的理解常扮演著解惑的關鍵性角色。

以本論文所嘗試建立的「《成唯識論》及其注疏編撰」的文獻數位化專案為例，我們的目的是要將《成唯識論》依《新導成唯識論》的科判做分科，並將窺基的《成唯識論述記》、《成唯識論樞要》、惠沼的《成唯識論了義燈》以及智周《成唯識論演秘》（《樞要》、《了義燈》、《演秘》被稱為唯識三疏），依《成唯識論》論文分段標注，並列於一地，使初學者可以省去大部分的機械化整理作業，而能方便閱讀、理解《成唯識論》的文意。

利用傳統以紙張為載體的經論，要做文本上的對讀，或是了解諸多注疏對某一文本的詮釋，在紙本上的翻找是極不方便、也極為耗時的，常常在此一部分便已耗費學者相當多的精力。而 CBETA 電子佛典集成的出現，無疑對此提供了很大的助力，快速解決了對等文獻查找的麻煩。但是即便如此，若要將兩部或多部經論譯著對比閱讀，仍是相當不便。因此，在許多佛學經論的研究者，在其研究或教學上，常會以 Microsoft Word 或其他便於編輯的文字編輯器，自行切割文獻以製作便於對照閱讀的內容。但這些坊間所見之「所見即所得」的文字編輯器，在編輯格式上常有許多的限制，特別是在更動、編輯其內容或做圖表呈現時，更是常常受到其格式的限制。然大部分的佛學研究者，因已慣用這些編輯軟體，熟悉其編輯規範與操作模式，因此大都只能接受其限制而採用此種文字檔來做經論間的排比對讀，而這也是個人在多年的教學中經常採用的方式。

就做經論的某一段、或某一部份的對比閱讀來說，這些「所見即所得」的文

* 法鼓文理學院，Email: jenjou.hung@gmail.com

** 法鼓文理學院，Email: bolang@dila.edu.tw

字編輯器確實容易操作，並可以快速達到對讀的目的，但是當超過二個或三個以上的文本在排對、閱讀上就並不很適當；此外，若是長篇幅的文本整體的對比，更會讓一個文件檔過大，而不易掌握。以此，個人每當在做這種文獻對讀、比較性的研究與教學時，常希望能有一種快速任意轉換對讀文本、或顯示出詮釋該經論之注疏書的模式，同時也希望在某種條件下，可以讓使用者簡易並方便地隨意增補自己之所見與心得。

另外，以資料長久保存的角度來說，將資料儲存為一個封閉的商用格式並不是一個長久之計。這些採用商用封閉格式的資料，非常倚賴於商用軟體的相容性。由於現今軟體功能發展快速，軟體廠商因發展新功能的需要而捨棄相容性舊資料格式的決定，時有所聞。因此，在製作相關數位內容時，我們便希望採用較為獨立、開放且製作人員能掌握所有細節的資料格式。

綜合上述的需求，在經過多次的討論後，我們選擇了 TEI 標記格式，作為經典與注疏之數位化系統資料的主要處理型。但一直以來，TEI 格式最為人文學者詬病之處，便是其基礎標記模型的複雜性與缺乏一個可以快速轉換為其他所需格式的友善界面。而要解決這樣的問題，就必須為經典與注疏的數位化工作，訂定一個較為容易使用的 TEI 相容格式，並考量經典與注疏的數位化系統需求，設計數位文獻資料庫的操作界面與功能與打造容易使用的轉換格式，讓所有使用者皆能運用這套模型或工具來製作自己所需的數位注疏資料。

本團隊於去年起，開始執行「《成唯識論》及其注疏編撰」的文獻數位化專案，除將《成唯識論》與《述記》、唯識三疏之內容彙整之外，並在《成唯識論》本文中以〔 〕、（ ）符號，插入的夾字、夾句，在特殊的佛學名詞，加以標記簡要之注解，以省去讀者查找字典的麻煩，同時在重要段落將論文做成圖表，於該段論文之科判處或文末標注，以使讀者能將其展開、閱讀，快速掌握該段要義。此外，在目前已有的諸多文獻數位化成果中，通常多限於網站上呈現與使用，其主要功能為查找及對讀，對於使用者來說，只能單純地遵循網站上所呈現的功能、模式進行閱讀，但卻無法進行個人化的編輯與運用，在本計畫中，我開發了將後端 TEI 文件資料，轉換為 Microsoft Word 相容性的格式的供能。讓使用者可以於慣用的文件檔模式中，添加個人的見解與心得，編撰成屬於自己個人化的講義資料。

經過兩年的努力，我們終於將《成唯識論》資料庫的內容與相關功能初步製作完成。並且我們已經為數位注疏系統的需求，定義了一個良好的 TEI 模型，也根據這個 TEI 模型製作了一套完整的數位資料庫界面轉換工具。利用此套標記規範與轉換工具，其他學者將可以簡易進行類似的數位注疏系統之製作。但於數位資料庫的製作過程中，我們多半仍由人工進行資料的爬梳與輸入，在借用人工智慧或文本挖掘的功能來協助製作資料庫的方面，仍待後續積極努力。

在多年的佛典經論的研究與教學中，深刻體認到文本對讀的重要性與需求，同時也了解到大多數的人文學者或學生，對於文獻數位標記的不熟悉，因此若要使經典文獻研究者自己能將文獻加以標記數位化，達到文本比較對讀之效果，則

便必須開發出一的極簡易的操作方式，使人文學者不會因為對於資訊語言的隔閡而導致杆格不入，要能使人文學者只要能簡單了解、學會 TEI 的標記，便能透過所開發完成的程式，執行出想要的結果，未來若再以使用者需求的角度來考量，修訂 xml 的呈現模式與功能，將有可能把佛典乃至儒、道經典的數位化，變得簡易化、普及化，進而變成每個研究者都可輕易上手的個人電子書的出版。以此只須再建立一個分類的統合交流平台，則將有機會形成一個具有個人化，同時又有如 Wiki 百科可編輯性的資料庫平台。

在本論文中，我們預計深入的討論與說明注疏對於文獻理解的重要性，並以佛教經典與相關注疏之常見體例為出發，進行注疏文獻特性的討論，並找出數位注疏系統的整體功能考量。之後輔以本專案所完成「《成唯識論》及其注疏編撰」的文獻數位化專案為例，說明《成唯識論》的思想背景、重要性、特色、文獻處理對象，以及系統文獻標記方式、系統架構、程式功能。最後我們將以一些研究範例的運行方式，進行系統的評估與找出未來系統發展的建議。

關鍵字：數位注疏系統、文獻標記、成唯識論、玄奘、窺基

Institutional Infrastructure for Fast Project Development, Research Tools Deployment and Research Outputs Integration of Digital Research Lifecycle

Hou Jeong Ho^{*}, Florian Kräutli^{**}, Esther Chen^{***}

Abstract

In 2016, Max Planck Institute for The History of Science (Max-Planck-Institut für Wissenschaftsgeschichte - MPIWG) started planning a new institutional infrastructure for integrating digital research outputs to the library and for the increasing digital research activities. The new infrastructure should be able to unify digital projects (including a decade legacy projects and the future ones) into library's sustainable, reusable research portal. Also the new infrastructure should help the IT researchers to effectively respond to the increasing digital research needs, such as digital tools deployment and digital project development (digitization, system development, service deployment and project transfer) in the digital humanities research lifecycle. This paper illustrates the motivation of initiating such an infrastructure as well as describing the requirements, challenges, and the current design of this infrastructure at MPIWG. We believe such an infrastructure is more and more important as digital research methodologies become more and more popular.

Keywords: Institute Infrastructure, Research Lifecycle, CIDOC/CRM, RDF, Docker, Repository, S3 storage

^{*} Max Planck Institute for The History of Science, Germany. Email: bho@mpiwg-berlin.mpg.de

^{**} Max Planck Institute for The History of Science, Germany. Email: fkraeutli@mpiwg-berlin.mpg.de

^{***} Max Planck Institute for The History of Science, Germany. Email: echen@mpiwg-berlin.mpg.de

Aceh Tsunami Recurrence from Historical Records and Local Interviews

Nurjanah Jane^{*}, Inoue Hiroki^{**}, Hidenori Watanave^{***}

Abstract

The Indian Ocean tsunami of 2004 is one of the biggest catastrophe of the last 100 years and has resulted in damaged infrastructure, loss of individual properties and environmental destruction in many coastal areas around the Indian Ocean region. And the death toll from the tsunami is estimated at more than 200,000 person with countless casualties. Moreover, scholar believe that the 2004 tsunami was simply one of many to have swept through in Aceh`s history. As such, there exists a body of local knowledge regarding these past tsunamis and earthquakes. The preservation and communication of this knowledge between generations has not, unfortunately, been prioritized. As such, there is a major information gap in most of the Aceh region resulting from the lack of continuity of this local historical knowledge. Had it been more widespread, this knowledge could have played a crucial role the 2004 tsunami, and it could be equally important for mitigating the risk in future disasters. Today, there are several ways to uncover the past tsunamis record of Aceh. One of them is paleotsunami, which combines a historical and scientific approach to create a record of the tsunami-history of a place. With paleotsunami, scientists can access the tsunami history of Aceh over thousand of years, i.e. by using carbon dating on tsunami sediment deposits, scientists can record a history of tsunami since the prehistoric period around 5000 years ago. When the tsunami causes sea water to flow inland, it creates horizontal sediment deposits, typically comprised of offshore and beach sand. These tsunami deposit act as geological evidence, which can be analysed using carbon dating to determine when the historical tsunami events occurred. Coastal geomorphology records can also detect past tsunami events by examining the evolution of the beach as an indication of past disturbances potentially caused by tsunamis. A third way to reconstruct the tsunami history of Aceh is through a historical approach. Ancient literature and personal prose manuscripts can provide written evidence of tsunami occurrences since 1000 years ago. By integrating paleotsunami with historical and scientific approaches, scientists can ensure more complete and precise records of tsunami occurrences in Aceh. In the future,

^{*} Tokyo Metropolitan University, Japan. Email: nurjanahtmu@gmail.com

^{**} Tokyo Metropolitan University, Japan. Email: rot1024@gmail.com

^{***} Tokyo Metropolitan University, Japan. Email: hwtntv@tmu.ac.jp

develop Aceh Paleotsunami Digital Archive in the open data source platform is one of the ways to augment information regarding past disasters for the sake of higher quality global natural disaster information.

Keywords: Paleotsunami, Gap Information, DRR, Global Information, Digital Archive

The Library as a Medium: Strategies for Tackling Challenges of Digital Humanities

Feng-En Tu*, Sharon Yang**

Abstract

Nowadays we see increasing numbers of online databases and digital tools available for scholars in East Asian humanities and social sciences. These new resources has brought profound impact on the ways in which people conduct research, but they also brought challenges to the academic libraries. In the past decade, libraries around the world made a tremendous contribution to this development by digitizing many of their rare and unique collections. However, digitization is but one of many things that the library can, and should, do to improve its service. In addition to digitization, the library also has to help faculty members and students navigate through the massive transformation of the research environment brought about by digital technologies. In this paper we introduce a few endeavors that the Harvard-Yenching Library made in the past few years in order to play a more proactive role in developing digital scholarship of East Asian studies, including creating a platform for researchers interested in digital resources, tools, and methodologies to exchange ideas, share experience, showcase ongoing projects, and address their needs. As we will argue in the paper, the library should see itself not as a space where the books are stored but as a medium through which scholars and knowledge can connect.

Keywords: Library, New Media, Digital Humanities

* Harvard University, USA. Email: fengen.tw@gmail.com

** Harvard-Yenching Library, USA. Email: yang8@fas.harvard.edu

Public Practice of Anthropological Collections (Archives) and Its Digitalization In Museums of Taiwan

Hui-Tuan Chang*

Abstract

Under the spirit of positivism, practice of collecting evidences has been carried out for scientific study. Research always has to be based on collection and accumulation of relevant data or materials. The use of such collection could further go beyond individual's life to be inherited as public resources shared for community benefits if being incorporated into permanent institutions such as museum or documentary center. And public practice of such collection can be expected to build solid and stable foundation in the long term for common knowledge development of human beings, as can be demonstrated with the well-known anthropologist Claude Levi-Strauss's case, whose important contributions of mythology and structuralism would not be possible without the opportunity to read and re-analyze records and materials donated by France Boas - founder of American anthropology, whose systematic investigation among American Indians of the north-western coast allowed rich materials left to be conserved by American Anthropological Archives, National Museum of Natural History in the United States.

Following the trend of technological development, museums first attempted to digitalize their collection management in Taiwan since 1990s. Promoted by National Science Foundation (Ministry of Science and Technology), almost all research institutions with collections have joined the national project for digitalization of collections and archives during 2000s. The paper intends to examine both effects and deficiencies in execution of the project, taking digitalization of anthropological collections and archives into account. It aims at institutional aspect of museum collections and archives in Taiwan. Whether the imagination of a freer and open communication with more diverse and pervasive conservation of heritage brought about by digitalization would come true? And what institutional conditions are needed for the implementation? It intends to examine collections and archives policy and management, types and amount of collections and archives achieved, and their use frequencies.

* Department of Ethnology of National Cheng-chi University. Email:huituan@nccu.edu.tw

Relevant institutional web-pages concerning collection and archives access policy and guides will be reviewed. It will focus on whether public access is improved to increase the number of beneficiaries, different archives and collections have built cooperative relation based on complementary works so as to improve cost benefit? etc.

Digitalization of archives as an advance thought unavoidable by research institutions in Taiwan, if not being able to recognize limitation imposed by its own cultural and social conditions, efforts could be vain and lost in the currents. It is necessary to study the practice of some advance case oversea for comparison so as to identify its own problems and possible solutions.

For instance, the environment and institutions of knowledge production and communication in Taiwan are different from that in the United States. American museums are regarded as public due to the duty in service of social development, and museums always adjust themselves responding to changes and needs of the society. Curators of museums have interactions and communications with university scholars of different disciplines, but they are regarded as a profession with special charge and function. However, public museums in Taiwan are founded and supported by the government, and are considered as agency of the government. Stable budget and personnel provide conditions for public museums to take the job of leading museum development of the country. Being financially supported by the government, public support is not very important or even irrelevant to them as there is no issue of accountability to the society for museums in Taiwan.

It has been pointed out that the cognition of professionalism in the society of Taiwan, especially regarding its relation to the public or social responsibility, is somewhat inadequate. Effective collaborative governance as well as constructive and positive forces therefore cannot be formed in the civil society, and professional self-regulation cannot be developed in the public interest groups and organizations.

Under the improper governance of non-government organizations, professionalism among museums in Taiwan has actually not been formed. One crucial problem is that there is no clear division of labor with knowledge production between public (or national) museums and other academic institutions. And the exclusive business of museum such as collection management, exhibitions policy-making and its evaluation has not been instituted with consensus. Only quantitative indicators such as the number of published papers in specific academic journals and of projects supported by the Ministry of Science and Technology are valued, taken as accumulated points for formal assessments of performance. Meetings or conferences are held now and then but mainly as platforms for presentation of academic research results as academic publications are among the most important criterion for one's promotion. Professions in the field of public service, academics and culture are not able to establish justified

mechanism for assessment and evaluation of themselves so as to take social responsibility. They mainly rely on formal evaluation of performance under endorsement of the government such as diploma or degree conferred by universities or published academic papers to manage themselves.

As cultural industries are given more weight in government policy in recent years, successful museum exhibitions are valued as management indicator. By contrast, museums and the public hardly pay attention to the off-stage collection management. Except for the academics, value and significance of museum collection and archives are scarcely recognized by the government and the society. There appears no consensus about the public interest of museum collections and archives between academics and the public, which is perhaps due to the leading role used to be played by the government and the academia, and civilian participation is excluded in the work of museum collection and archives. This could make the public aloof towards and unfamiliar with collection and archives as memory assets. As a result, the idea of heritage preservation would not be able to implement and take roots in the civil society. No passion, involvement and contribution would then be aroused from the civil society.

Financial assistance or grants are not provided and assessed in a long term perspective and to be accorded with the public practice and performance quality provided by institutions or groups. Neither museums, collections and archives have their own standards of evaluation, nor are they taken into account as one's personal achievements. These are the problems to be pointed out for solution and improvement.

Keywords: Anthropological Archives, Digital Archives and Collection Management, Museum's Public Practice

體育情報學發展歷程、現狀及發展前景展望

張鵬飛*

摘 要

回顧體育情報的發展歷程和與情報學的淵源。分析體育情報發展現狀。在詳細的闡述體育情報基本內容的基礎上，總結體育情報的功能和重要性，並展望體育情報今後發展趨勢。體育情報會促進體育產業發展、促進體育競技水準提升、推廣體育運動等。

關鍵字：體育情報、競技體育、高科技化、體育產業

* 南開大學，Email: 363428708@qq.com

淺談次文化的跨文化：

從郝毅博藉《後宮甄嬛傳》打入台灣鄉民圈來看

黃南穎*

摘要

本研究採用肯尼斯柏克（Kenneth Burke）所提出的「戲劇五因分析方法（Dramatistic Pentad）」，進行內、外在分析，嘗試理解英籍人士郝毅博（Ben Hedges）評論《後宮甄嬛傳》，其背後所隱含華歐溝通之隔閡，期許於跨文化溝通（Intercultural Communication）之途略盡絲薄。

有別於西方人有話直說的溝通方式，華人顯得較為隱諱保守，這樣的情形在跨文化溝通上容易形成隔閡，郝氏對《後宮甄嬛傳》的評論影片便闡述了此種文化交流上會遇到的難題，且該影片一出，隨即在網路上被大量轉載、討論，引起各界關注。網路可以是次文化的傳播管道，而郝毅博身為外國素人，卻能突破跨文化溝通的隔閡，更成功打入網路這個次文化圈，本研究盼能探索其原因。

戲劇五因用來研究言者的動機。人生就是戲，人做任何事都有想要達成的目標，因此藉著塑造出一個戲劇場景、運用每個戲劇元素來讓自己的行為合理化，進而透過這種編、導、演的方式讓這齣戲的劇情進展成自己想要的結局。戲劇五因後來又演變為「內在分析」與「外在分析」。

本研究於內在分析方面，乃針對文本內容來探索演說者之評論動機；於外在分析方面，則蒐集文本以外之相關資料或背景來予以剖析，從而提供前者詮釋之基礎。結果發現，評論者與其《老外看中國》之工作團隊乃以目的（Purpose）因素來主導行動（Act）、行動者（Agent）、場景（Scene），以及手段（Agency）等因素。於內在分析上，他們除了藉此表達自身對華夏文化的讚賞外，郝氏亦意圖尋找其童年居住香江之回憶，並為其本人奠定在外籍中文圈與臺灣鄉民圈之地位。至若外在分析的時空背景部分，其一，全球社會商業貿易往來頻繁，儼然成為文化大熔爐的地球村（Global Village）；其二，當華人社會播出收視率高升的《後宮甄嬛傳》時，適逢薄妻谷開來涉及英商海伍德（Neil Heywood）命案，致使薄熙來遭懲處而終結政治生涯。斯時，郝氏評論既出，無疑的為「甄嬛」耀錦增添艷妝，是以網際間網友熱烈轉發分享「郝評」，益令其本人與團隊一時之間聞達於國人。

周敦頤《通書·文辭》：「文所以載道也。」文化者，修治文德以施行教化；綜括之則涵蓋：道德、宗教、語言、文字、藝術、科學等人類在各方面活動、過程所發展之總成果。然而，由於地域、種族、血統、國家、思想模式等差異，容

* 世新大學新聞傳播學院口語傳播學系碩士班，Email: linsen790609@yahoo.com.tw

或衍為各種不同文化。而今之社會時空限縮，文化交流之場景不斷上演，東道西說，可能造成衝突；然而南腔北調，或可共譜和諧動聽的樂章；是絲是麻，亦能併續金采耀眼的鏐艷。

不論不同文化相遇時會如何發酵，若各方都能夠秉持互相尊重之心態以降低對立之勢，則易於共同締建諧和之溝通平臺。尤以自任文化溝通相關領域工作之志士，能以荷擔橋梁為職司，自我要求日益精進，則人類社會必當獲致「苟日新，日日新，又日新」之境。

巨量資料與公民社會：從數位隱私談起

鐘孝宇*

摘要

巨量資料與演算法技術的應用，以統計相關性的知識論，形成不同的洞察與價值，其以資料驅動技術為主所辨識出的現象模式，結果似乎比傳統科學方法論更為準確，建立其客觀性的優勢。因此，巨量資料科技的應用，在萬物聯網的數位環境之中，全面滲透人類日常生活的各面細節，賦予我們改變生活基本面向的能力，包含我們的對話溝通、學習、工作、購物、娛樂，甚至我們參與政治與社會生活的方式。事實上，在巨量資料脈絡下的數位技術物（artifacts），已不單純僅是日常生活的輔助工具，而是積極介入、給予指示、引導行為，甚至定義我們的身分（identity）、調整（modify）、調控（modulate）我們行為的知識生產與治理模式。本文主張，巨量資料科技的知識生產與治理模式，對於公私部門而言，雖然具有提升公共及商業利益的正面效益，然而，其知識生產與治理模式對於個人及群體所產生的權力作用，卻也伴隨極大的隱憂，其可能造成一種新型態的數位隱私侵害風險，進而可能影響健全的公民社會（civil society）發展。

本文認為，隱私是公民社會的構成元素，它保障個人在社會建構的形塑之下，保有日常行為實踐的能動性，得在自我自主經驗與社會建構的來回探索之間，生成個人的主體性。這個動態的主體性發展空間，使得我們得以開拓環境中的意外發現（serendipity），建立心智的批判獨立性，具備如此公民特質的社會，才有能力促進社會多元、創新、民主改革等實踐可能性，形塑健全的公民社會，這即是隱私的社會性與社會價值所在。然而，巨量資料監控的知識生產與治理模式，尤其是商業監控的「調控治理」，其權力作用機制可能造成嚴重的個人數位隱私風險。因為，商業組織的監控資本主義（surveillance capitalism）運作邏輯，乃是在利潤最大化的前提上，分派使用者至特定的「個人化」資訊環境，形塑資訊的過濾氣泡（filter bubbles），決定其所能看到的服務、廣告、內容，並線性的引導其持續的發展走向，它侵蝕了作為使用者的公民的個人主體性發展空間，削弱公民的心智批判獨立性。並且，在演算法的自我增強機制下，將阻斷個人日常生活實踐的意外發現可能，使得我們喪失接觸不同社群、資訊、概念、觀點、政治立場、意識形態的機會，使公民自始失去接觸「未知的未知」（unknown unknowns）的可能性。

數位科技是我們個人感官的延伸，它不僅單純作為資訊與訊息的傳遞通道，我們感知的世界，是透過網路資訊科技等技術物的媒介過程所傳達，技術物組織

* 政治大學，Email: zqperation@gmail.com

我們的世界，精巧的形塑我們理解世界的方式、影響我們與世界的關係，甚至深刻的影響我們的認知模式。因此，我們必須審慎評估巨量資料科技的政治性，並辨識其以統計相關性驅動的知識論對於特定群體或個人的分類、預測、調控等權力作用，可能會對公民社會所造成的傷害。本文認為數位環境中的巨量資料應用，特別基於其監控（surveillance）本質，可能對於個人形成全面的思想介入與編排，影響個人日常生活實踐的能動性，削弱個人主體性（subjectivity）的發展空間，進而影響自我決定、創新、人際交往互動等實踐，造成社會的負面外部性。

特別應注意的是，巨量資料作為一種知識生產與權力作用的治理模式，並不具有其宣稱的純然客觀性，偏見、歧視、誤解、惡意都可能鑲嵌於演算法的程式碼之中，而造成錯誤的分類、預測與調控。必須承認，調控治理對於個人的思想編排，其具體的傷害仍是隱微而難以主張的。其具體的傷害仍是隱微而難以主張的。但筆者認為，巨量資料對於使用者的身分認同、智識、意識形態、人際交往互動空間的全面介入，將嚴重削弱網路使用者的另一個身分，也就是公民的公民特質，而影響公民社會的發展。

然而，現今無論在學術或產業領域，巨量資料研究（big data research）作為一種新興的研究方法與治理模式，都已是不可逆轉的趨勢，而巨量資料研究的範圍、規模、與複雜性，早已超出傳統電腦或自然學科的研究議程，這使得傳統的規範和指導原則難以適用。因此，本文認為，無論公私部門，只要是以「人類主體」（human subjects）為研究對象的巨量資料研究，不管是社會、心理、經濟現象等研究主題，都應該建立並發展巨量資料的應用倫理，盡可能最小化潛在的損害，同時，亦應該重新思考現行個人資料保護機制面臨巨量資料監控造成的系統性數位隱私風險的侷限與修正方向。

聚焦於以上的問題，本文第一節首先將簡短介紹巨量資料的定義、各國政策發展與商業及公共治理的最新應用案例；第二節，本文將討論隱私與隱私權的社會價值，探討隱私權除為個人利益以外，亦為重要的社會利益，因為隱私權透過保障個人的主體性發展空間，可促成公民社會整體福祉的累積，在面對巨量資料的隱私侵害威脅時，可作為立法政策的理論資源；第三節，本文將試圖辨識巨量資料應用的監控性質，將其區分為「政府監控」與「商業監控」，並主張其各自代表不同的知識生產與權力治理模式，進而產生不同的隱私威脅。而筆者主張，在商業監控的脈絡之下，將會對於個人或群體造成更嚴重的隱私侵害風險；第四節，本文首先將指出現行各國的個人資料保護或隱私規範作為資訊社會的產物，無法因應數位環境中的巨量資料應用所造成的三個困境：個人資料性質的改變、告知與同意機制的失敗、資料汙染。接著論證三種可能的原則性因應策略：巨量資料的應用倫理、巨量資料利用的正當程序原則、社會責任與自律規範。最後，將視角拉回我國的個人資料保護規範，參考歐盟一般資料保護規則（GDPR）的相關立法，建議我國應盡速設置專責資料保護監管機構，並提供具體的個人資料保護法修正方向：資料管理機制、內部監督機制、創設使用者資料可攜權。

關鍵字：巨量資料、演算法、監控、調控、數位隱私、個人主體性發展空間、公民社會、個人資料保護機制

人形存有與數位尊嚴：數位人文要預示未來命運嗎？

蔡瑞霖*

摘要

數位人文學 (digital humanities) 是專技性學科，應具有自身所以成立的恆久命題：人性 (humanity) 或說以人之存在為中心概念。它有漫長演化過程，可以視為「人形混雜」 (hybrid beings with the humanoid) 發展史。即，我們可以通過歷史共業的混雜性之分析而描述其多重區分及混雜，即人與獸(蠻荒世代)區分、人與神(神話及其後)區分，還有人與機器(工業和資訊高科技時代)的區分，如今人形混雜已經進入到數位化階段而更加突出「人機拼裝」之異己狀態，在虛擬真實或實際上，甚至被設想成生化人或賽博格之模樣。

爰此，本文提出問題：首先，數位人文是建立在過去人形演化的共業結果上嗎？吾人應如何重新界定人性來看待當代人形存有 (humanoid being)？是否仍舊依據人性尊嚴來界定的？尊嚴或人性，都來自道德實踐與倫理力，一如康德所說的，「唯道德與人之本性才能有尊嚴之名」，也就是依自由意識而行動的人。人而如此，那麼混雜的人形存有呢？

其次，從人性論來看，如何確認所謂「數位尊嚴」 (digital dignity) 之權利，如同人權公約一般？其核心價值都歸屬於誰呢？在導向全球網絡、雲端運算與大數據的方向上，最終我們能預示當代人形存有以未來的集體命運嗎？畢竟大數據最後將服務於誰？藉由數位網絡化而不斷擴張的全球化怪獸，譬如媒體霸權(類似於訊息、資本與軍事的混合體)之透過雲端儲存與運算而集中的控制力量，是如何運作的？這些人形存有的哲學問題，都涉及到數位人文學確立與發展的可能意義。

專技操作的數位人文學，總是在人性恆在的前提下進行運算的。本文認為人形存有是歷史共業，而且數位人文學的核心價值應該是人性尊嚴。經過演化的人與獸、神以及機器的尊嚴可以和協混同起來，人之面對冰冷機器，卻能反饋於環境、浸溶到各種真實中形成「溫濕的生命世界」 (the moist world of life)，那麼數位人文學才有積極意義。雖然前所未有的現象，依然是全球人類還在不斷自我耗損和相互摧毀人類共有的尊嚴中。這是人形存有的災難，或是人類命運的真相？

* 考試院，Email: raylin.tsai@gmail.com

Panel: Spatial Humanities Mapping in Taiwan

David Blundell^{*}, Jih-Fa Jan^{**}, Ching-Chih Lin^{***},
Ann Heylen^{****}, Oliver Streiter^{*****}, James Morris^{*****}

For DADH 2017 we propose to revisit our book chapters for a Springer Handbook Big Data in Computational Social Science and Humanities edited by Shu-Heng Chen, to further discuss our current work proposed to MoST in Digital Humanities (DH) this year, and talk about our latest DH research efforts.

The purpose of this panel proposal is to bring studies that illustrate different approaches to bridge the historical distinction between humanities and social science with a digital solution. As crosswalks for information from multiple sources and in multiple formats spatial humanities – a sub-discipline of digital humanities based on geographic information systems (GIS) and timelines – create visual indexes for diverse cultural data and provide an effective integrating and contextualizing function for geo-cultural attributes. We are particularly interested by finding meaning and innovation in DH and enrich what scholarly efforts have already achieved.

In 2015, we initiated our Asia-Pacific Spatiotemporal Institute (ApSTi, <http://apsti.nccu.edu.tw>) on the campus of National Chengchi University, Taipei. Here we have created an environment for synergies to occur between researchers serving to facilitate studies as a home for innovative geographic information systems (GIS) based research and sharing advanced technologies in the digital humanities. Our institute offers a range of project services to facilitate new ways of configuring data based on geospatial tools. From the interfacing of spatiotemporal systems dynamic maps of unique information possibilities are generated. Researchers of various , disciplines contribute to dialogues about techniques, challenges, and results of digital humanities research. In short, we are facilitating capacity building and innovative ways of sharing information by digital methods for visualizing spatiotemporal aspects of human experience.

A far-reaching goal is to further standards in cartographic strategies through the

^{*} ECAI, ApSTi, National Chengchi University, Email: pacific@berkeley.edu

^{**} ApSTi, National Chengchi University, Email: jfjan@nccu.edu.tw

^{***} ApSTi, National Chengchi University, Email: cclin52@gmail.com

^{****} ApSTi, National Taiwan Normal University , Email: aheylen26@gmail.com

^{*****} ApSTi, National Kaohsiung University, Email: oliverstreiter@gmail.com

^{*****} IDAS, ApSTi, National Chengchi University, Email: jamesxmorris@gmail.com

utility of digitalization and animation of map content giving new possibilities in the hands of local and international collaborators. We provide examples for developing best practice standards applied to databases giving interactive multimedia utility aspects. This allows uniting the context of environmental landscapes with cultural data for making enhanced possibilities in spatial humanities with scales of data, large and small – with humanistic and scientific results.

For comprehensive developments in spatial humanities we consult Jo Guldi's introduction of the spatial turn for eight academic disciplines, "What is the Spatial Turn?" (2016) and Richard White's essay "What is Spatial History?" (2010). Digital mapping today gives resource affordability to researchers. Availability to digital resources allows novice or advanced researchers who are not cartographers, abilities to chart information.

Our research draws from state of the art studies in DH funded by MoST, putting Taiwan and its surrounding networks in the region on the map, and will be bundled in a comprehensive section of the current Springer Handbook Big Data in Computational Social Science and Humanities edited by Shu-Heng Chen.

Now historiography has fresh and innovative tools (Robertson 2012), and not about literary text mining. GIS provides history "the most exciting developments in both digital and spatial humanities" (Gregory and Geddes 2014) with advances in computing and information infrastructures offering researchers possibilities of reconsidering the entire strategy of analysis and dissemination of information. It features 'deep mapping' acknowledging multiple meanings in a place that "enables humanities scholars to discover relationships of memory, artifact, and experience that exist in a particular place and across time" (Bodenhamer et. al. 2010).

As the authors, we have experience in visualizing layers of ethno-archaeology from the perspective of sociocultural networks based on local self-represented environments, individual perspectives, and acculturation in space at a given time. We describe applications and methodological strategies for new ways to approach data from the grassroots. Our illustrations utilize GIS comprehending spatial happenings through time.

Our GIS research contributes to important DH academic discourse in many ways. Time maps are utilized to trace stories by the way people move through time. Other humanities mapping is data analysis of information quantitatively to understand spatial patterns from original sources. These visualized displays contribute to discovering knowledge, answering questions, and seeking other questions.

Spatial humanities produce a cycle of questions creating layers of maps portrayed in different ways. The projects introduced here further promote such innovative media of aesthetics and ethnography. Both analog and digital approaches are important to

explore the aesthetics and ontology of perspectives of history, social and physical sciences, and humanities. We ask what is heritage vis-a-vis present day society and what are the determinants of people in their context.

The aim is to recount human transformations from cartography, historical records, aesthetic determinants, and community research partnerships. That implicit conceptual underpinning of advanced hermeneutics research in our qualitative tradition is critical and able to potentially enrich and deepen perspectives based on elements seemingly unrelated, yet connected.

The interactivity of digital maps allows one to filter data to the desired scale and includes a multitude of sources that, through their abundance, allow for data gathering and trans-disciplinarily comparisons to be made. Sometimes, these techniques take confidence to master. Yet, we hope to provide case examples conducted by anthropologists and historians, not cartographers. We aim to open spatial humanities to anyone.

Each of the proposed papers will illustrate dimensions of this innovative digital infrastructure, taking the audience through the geography, history and landscape of Taiwan society, past and present. We are describing applications and methodological strategies for new ways to approach data from the grassroots. Our illustrations utilize GIS comprehending spatial happenings through time. Time maps are utilized to trace stories by the way people move through time, and this can be best illustrated with the technology and methodology of spatial humanities that produce a cycle of questions creating layers of maps portrayed in different ways. Our challenge is to imagine new methods for doing research and making those results available to broader user communities. Can we find meaning and innovation in the digital humanities beyond what scholarly efforts have already accomplished?

Spatial Humanities:

An Integrated Approach to Spatiotemporal Research

David Blundell, Ching-chih Lin, and James Morris propose to present on spatial humanities as an integrated approach to spatiotemporal research in Taiwan featured as a project on religious studies, part of the Asia-Pacific SpatioTemporal Institute (ApSTi, <http://apsti.nccu.edu.tw>), National Chengchi University. Spatial humanities are a sub-discipline of digital humanities based on geographic information systems (GIS) and timelines providing an effective integrating and contextualizing function for geo-cultural attributes. As information systems from multiple sources and in multiple formats they create visual indexes for diverse cultural data. Spatiotemporal interfaces provide new methods of integrating primary source materials into web-based interactive and 3D visualizations.

Our paper will be based on case studies from research related to (a.) Atlas of Maritime Buddhism and Austronesia Team (Electronic Cultural Atlas Initiative, ECAI <http://www.ecai.org>), (b.) GIS spatiotemporal analysis of Chinese woodblock prints (nianhua) and local religious networks in Taiwan, and (b.) documenting and mapping land god shrines to establish patterns of settlement, communal organization, and historical trade networks of communities in Taiwan, connecting to southern maritime China, Hong Kong, Macau, and outlying islands.

We examine GIS point locations tracing routes and networks imbued with historical meaning across the region linked to enriched attribute information. These are charted and visualized in maps and can be analyzed with network analysis, creating an innovative digital infrastructure for scholarly collaboration and creation of customizable visualizations. Our paper illustrates a range of ways to facilitate configuring social science data with geospatial tools featuring Taiwan research with GIS point locations, migration and historical trade routes, and religious sites of the region linked to enriched attribute spatial information (Blundell and Zerneke 2014; Lin 2016; Streiter and Morris 2015).

Citizen Science and Volunteered Geographic Information (VGI)

Jihn-Fa (Andy) Jan examines citizen science and volunteered geographic information (VGI). Owing to the rapid development of geospatial and mobile communication technologies in recent years, acquisition of high-quality spatiotemporal information has become much more efficient and cost-effective than before. As a result, many researchers deem that data collected by volunteers with little training can be used for scientific researches if carefully designed quality assurance process is performed. In this paper Jan explores the application of VGI in spatial humanities.

Geographic information is required by a wealth of scientific research for various disciplines. Due to much progress of geospatial technologies in recent years, acquisition of high-quality spatial and temporal information has become much more efficient and cost-effective than past few decades. Remote sensing provides massive high resolution imageries about Earth surface, which can be analyzed by image processing tools to automatically derive valuable information for various applications such as climate change, land resources inventory, environmental monitoring, and urban sprawl.

The advent of global positioning system (GPS) has revolutionized the process of surveying. GPS allows local users to directly measure coordinates of location very accurately and rapidly. Integrated with IMU (inertial measurement unit) and other sensors, GPS can also be used for mobile mapping when mounted on mobile platforms such as bicycles, cars, boats, aircrafts, and UAS (unmanned aerial system). Results of the research are visualized and posted through Web-based Arches (Jan and Mao 2016).

Expressing Dynamic Maps through 17th-Century Taiwan Dutch Manuscripts

Ann Heylen's presentation is the visualization of her Springer handbook chapter on expressing dynamic maps through 17th-Century Taiwan Dutch manuscripts that guides the reader through the process involved in applying digital software to 17th-century Dutch handwritten manuscripts in documenting the presence of the Dutch community in Taiwan. The key text for this exercise is the digitized version of the Church Minutes or Consistory Notes of the manuscript Kercboek, Brievenboek van Formosa, 23 januari 1643- 4 maart 1660 in Dutch and English translation, and applies geocultural space in mapping cultural practices and attitudes onto the geographic regions that encompass the historical and cultural events and narratives associated with that location. The narrative invokes the sense of place of Dutch VOC personnel between 1643 and 1649 (Heylen 2016). The visualization features different components of this digitized text integrated in a relational database framework Arches being developed by Jihn-Fa (Andy) Jan (see Jan and Mao 2016).

Simultaneous Invention or Propagation of Cultural Practices?

The Tanghao on Tombstones of Taiwan and Penghu

Oliver Streiter proposes to share his Taiwan regional tombs research uniting dimensions that contribute to the creation and transformation of cultural practices in one comprehensive spatial model expanding previous work on cemeteries describing interactions of material cultures, climate, religion and politics in a combination of loosely connected timelines, maps, correlation statistics and prose accounts.

In this paper Streiter illustrates the application of large data from tombs research to better understand the emergence, transformation, and interpretation of cultural practices. Having digitized tombstone inscriptions on Taiwan and Penghu, he begins with case studies of distinctive carving practices to trace their origin in time and in space. The analysis presented here uses the example of the tanghao (堂號), a type of place name that identifies origins of a clan. In one case he traced practices of carving a tanghao established on Xiyu, an island of the Penghu archipelago, where he finds it to be an expression of loyalty to the Qing dynasty (皇清) just after the occupation of Penghu by the Japanese in 1895.

In the research of such examples, Streiter provided representative data with a unified model of interaction networks to obtain through their visualization new insights on possible interactions and to evaluate mathematically the contribution of each entity in a network, substantiating previous mixed-schema (Streiter and Goudin 2013). These are part of his ThakBong project where research develops techniques and interfaces for the storage, annotation, transcription, distribution, analysis and visualization of data and media files for scientific and educational purposes.

These proposed papers trace the maturity of DH, crossing boundaries of geography and history to include a wider range of disciplines through our mutual trans-disciplinary synergies. We comprehensively search for spatiotemporal points to where networks existed and continue to do so.

Our spatiotemporal methodologies explore new ways of integrating primary source materials into crosswalks of interactive visualizations. Utilizing GIS we are able to chart the extent and dynamics of specific traits of cultural information to create layered maps. These elements are transmitted through time based on spatial points. The research outcome is a Web-based, interactive, and cultural platform acting as an atlas for a local community bulletin board designed for scholarly exchange.

References

- Blundell, D., and J. Zerneke. 2014. Early Austronesian historical voyaging in Monsoon Asia: Heritage and knowledge for museum displays utilizing texts, archaeology, digital interactive components, and GIS approaches. *International Journal of Humanities and Arts Computing*. 8: 237-252.
- Bodenhamer, D. J., J. Corrigan, and T. M. Harris. 2010. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington & Indianapolis: Indiana University Press.
- Gregory, I. N. and A. Geddes, Eds. 2014. *Toward Spatial Humanities: Historical GIS and Spatial History*. Bloomington: Indiana University Press.
- Guldi, J. What is the Spatial Turn? In *Spatial Humanities: A Project of the Institute for Enabling Geospatial Scholarship*. <http://spatial.scholarslab.org/spatial-turn/the-spatial-turn-in-history/index.html>.
- Jan, J. F., and W. H. Mao. 2016. Cultural heritage inventory using Arches: case study of a rural community in Taiwan. Paper presented at the Pacific Neighborhood Consortium (PNC) Annual Conference and Joint Meetings, Does Data Construct Reality? The Getty Center, Los Angeles, California, August 16-18.
- Heylen, A. 2016. Expressing power in dynamic maps through 17th century Taiwan Dutch manuscripts. European Association of Taiwan Studies (EATS) Conference, Prague, Czech Republic, March 30-April 1.
- Lin, C. C. 2016. GIS spatio-temporal analysis of Chinese woodblock prints (Nianhua) and religious networks in Taiwan. Paper presented at the Pacific Neighborhood Consortium (PNC) Annual Conference and Joint Meetings, Does Data Construct Reality? The Getty Center, Los Angeles, California, August 16-18.
- Robertson, S. 2012. Putting Harlem on the map. In *Writing History in the Digital Age*, Jack Dougherty and Kristen Nawrotzki, Eds. Ann Arbor: University of Michigan Press.
- Streiter, O. and Y. Goudin. 2013. The tanghao on Taiwan's tombstones. *Archivi Orientalni*, 81.3: 459-494.
- Streiter, O. and J. X. Morris. 2015. Researching Taiwan's gravesites with ThakBong and R. Paper presented in the ECAI workshop at the Pacific Neighborhood Consortium (PNC) Conference Annual Conference and Joint Meetings, Taking Data into the Public Domain, University of Macau, September 27-30.
- White, R. 2010. What is spatial history? *Spatial History Lab: Working Paper*, Stanford University. <https://web.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=29>.

歷史分期的數位人文研究： 以《人民日報》(1946-2016)中深圳主題報導為例

邱偉雲*、梁穎誼**

摘要

本文之研究目的，是以找到另一有別於過去從人文主觀角度出發進行歷史分期的方法，希望能夠改以客觀數據角度進行歷史分期的工作。歷史分期問題一直以來都是歷史研究者，尤其是長時段歷史研究必須進行的基礎重要工作。數位人文研究的優勢，即在於能鳥瞰的、遠讀的處理數據體量大，歷時時間長的變化議題，在這樣的脈絡下，當數位人文研究與歷史研究結合時，自然會召喚出法國年鑑學派布勞岱爾（Fernand Braudel，1902-1985）的「結構」（structure）和「長時段」（La longue duree）概念。布勞岱爾主張在長時段歷史中有一些穩定結構/集合/架構/現實長期存在，這些穩定結構遲滯/阻礙/支配著歷史流動速度，另有一些結構則是很快消失。在上述布勞岱爾的主張下，所謂的歷史分期，就是指長時段中的結構推移現象，這樣的結構推移是在前結構被時間逐漸磨損/緩慢推動下，直到一個分水嶺，前後結構已經小同大異之際，即可視為兩個歷史時期。然而過去都是經過人文學者定性主觀的去判別「小同大異」的時間分水嶺，通常是以歷史紀年、事件、意識型態去作為標準，但這些標準都無法客觀的證明「結構」確實出現「顯著相異」，如今本文正是想透統計觀點，找到一客觀計算出「結構顯著相異」之方法，亦即找到一結構轉變的時刻（moment），在此時刻前後確實具有詞頻/詞叢結構於數據上的顯著變化現象，為歷史分期研究帶來另一種視野，此即為本文之研究目的。

本文將以 1946-2016 年間《人民日報》中所有標題中包含「深圳」一詞的報導（以下稱為「涉深報導」）為研究範圍，共有 3,203 篇。《人民日報》從 1946 年創刊至今已長達 71 年，見證了新中國的發展史，在這 71 年中刊物未曾中斷，為中國第一大報，1992 年被聯合國教科文組織評為世界十大報紙之一，受到中外各國的關注，因此中外各國對深圳城市形象的接受，很大程度除了親自到深圳的感受外，其餘就是透過《人民日報》來進行想像與建構，故分析《人民日報》中的深圳形象，對於我們掌握國內外對深圳此一城市的理解具有指標性的意義。

本文在研究方法部分，將比較 1.人工主觀（歷史紀年、事件、意識型態）分期；2.數值高峰分期；3.關鍵詞叢異同分期等方法。其中在數值高峰分期部分，將以歷年涉深報導篇數為分子，以歷年報導總篇數為分母進行計算，最後得出歷年涉深報導的比例，透過此一數值的起伏進行分期工作；而在關鍵詞叢異同分期法部分，是基於不同時期的涉深報導當有不同的論述修辭結構，故不同時期中涉深報導之共現詞叢當會伴隨時間變化有所差

* 山東大學歷史文化學院歷史學系，Email: brianacwu@163.com

** 國立政治大學中國文學系，Email: htiosc85@hotmail.com

異，在此基礎上，本文將從關鍵詞叢歷時性的變化（新舊詞彙結構的新陳代謝）中進行歷史分期工作。

上述三種分期方法雖然都可達到歷史分期效果，但也都有些問題在，首先在人工分期部分過於主觀，故會有伴隨不同學者學術立場與見解的差異，透露出某種個人主觀偏好的現象；其次在數值高峰分期上，可能落入將數據的突變湧現，誤以為結構改變的陷阱；其三在關鍵詞叢異同分期部分，仍然缺乏客觀判斷詞叢新陳代謝之標準，故仍無法客觀指稱思想轉型的時間分水嶺。由上可見，三種分期方法雖然有其作用，然卻有它們的侷限性，其一為這些分期法並未有具體的數位計算流程，仍屬人工驅動，因此依然仍被視為主觀分期，不具有普遍性；其二是從《人民日報》文本資料結構去觀察，可發現詞彙的變化可能具有平均比例上的階段變化(Mean shift)，而前述方法並未考量到此種轉變型態。因此，本文將提出第四種歷史分期方法，以統計數據作為客觀分期標準，由人工驅動轉向全自動數據驅動歷史分期，為長時段歷史研究方法提供一新的視角，此為本文研究結果與價值所在。

關鍵字：深圳、人民日報、歷史分期

歷史文化多源(元)脈絡之動態呈現設計： 以澎湖石滬文化為例

吳如娟*、陳至柔**、王瑞琪***、傅詩涓****

摘要

歷史是早期當地居民環境觀察在地環境特性發展因地制宜的生活文化，為當地人們與環境相依存生活的表徵，現今成為人文研究者在推論當地早期生活與文化等層面的參考依據，隨著近年來資訊科技與內容數位化持續發展，帶來許多跨領域技術支援與應用，例如：在歷史人文研究領域所應用之數位典藏，運用資訊科技的手段來提升產業專業知識化、數位化之程度，提高文化在經濟地位上的角色，帶動了著重保存當地傳統文化的國家與人文研究學者採用數位典藏，進行知識層面的人文共享以及多元的資料加值應用，以宣揚當地的生活文化，讓這些珍貴的文化智產，透過數位典藏的形式永續經營，例如：全世界僅存澎湖早期石滬的文化保存與呈現。

經本研究觀察近年石滬相關文獻與史料發現目前世界各地的石滬相關資料保存現況與方法，隨著數位多媒體以及資通訊科技進步，仍有改善之處，例如：(1)石滬的相關史料並未完全數位化；(2)數位化資料過於分散且不完整；(3)石滬史料類別分割不清，難以重組歷史脈絡與面貌；(4)系統呈現大多採靜態內容形式，例如：近似於早期紙本典藏的呈現方式設計，缺乏時間軸的引導與事件的串聯等動態的呈現等。(5)數位典藏資源的應用與永續；且尚未完全歸納石滬活動相關組成因素，難以從中窺見石滬與當地生活的脈絡，以貼近研究者的需求，成為了目前人文研究者在石滬與時空環境(時間、經緯度)相關轉換與改變調查上的障礙，且石滬多元資源未能被有效整合與運用。

因此，本研究以澎湖石滬文化為例，從數位人文的觀點進行歷史文化多源(元)脈絡之動態呈現設計，先以活動理論歸納石滬活動組成元素，進而以資料倉儲概念針對石滬文化特性，建構石滬活動維度、介面、服務與流程，並運用超媒體概念進行介面元件的視覺化與串聯，以服務導向架構進行介面層、服務邏輯層、資料層的組裝與流程整合，還原原有石滬文化的歷史脈絡，提供了多重檢視角度歷史知識地圖，使用者透過多維度的參數條件篩選與組合，查勘不同時空下的當地人文與環境變化與歷史文化意涵，了解石滬對於早期澎湖社區的社會與經濟的關聯與意涵，透過數位化的形式永續經營石滬知識，能應用在其他歷史文化的數位典藏建置設計，以資通訊科技從事多維度的觀察與研究，諸如年代、地理環境、與人文社會變遷等關係，以有效協助其從事發掘歷史脈絡中的潛在意涵與價值。

關鍵字：石滬、數位內容、活動理論、資料倉儲、超媒體、服務導向架構

* 逢甲大學企業管理學(系)所，Email: katejcwu@mail.fcu.edu.tw

** 國立澎湖科技大學行銷暨物流管理系，Email: benson@gms.npu.edu.tw

*** 國立臺灣科技大學資訊管理所，Email: s9277702@gmail.com

**** 逢甲大學企業管理學系，Email: serena6044@gmail.com

唐宋詩歌活動空間的分佈與位移： 基於《唐宋文學編年地圖》的數據分析

王兆鵬*

摘 要

作為一種研究方法的數字人文，主要是運用大數據對研究對象進行統計分析，以發現新問題，提出新觀點，並對統計分析的結果予以可視化呈現。是故，數字人文研究，數據是基礎，統計分析是關鍵，發現問題是目的，可視化呈現是手段。筆者主持的國家社會科學基金重大專案《唐宋文學編年系地信息平臺建設》，即是數字人文研究的具體實踐。為完成此項目，我們組建了近百人的研究團隊，歷時五年，廣泛搜羅唐宋作家年譜、編年箋注之別集及相關生平作品考訂之論文，轉錄成唐宋文學編年數據，然後運用 GIS 技術，與地圖相結合，建設成“唐宋文學編年地圖”，本文的數據即來源於此。然則，如何用大數據來發現新問題，尋求新的學術增長點？本文即嘗試運用大數據來分析唐宋詩歌版圖的分佈與變化，旨在提供數字人文研究的具體實例，探索數字人文研究的可能路向。

本文為筆者所撰《唐宋詩歌版圖的空間分佈與位移》（載《中國人民大學學報》2016年第6期）的姊妹篇。《唐宋詩歌版圖的空間分佈與位移》主要依據唐宋詩人的占籍情況統計分析唐宋詩歌版圖的空間分佈與位移。統計數據顯示，唐宋詩人地域分佈的整體格局是，唐代詩壇的重心在北方，宋代詩壇的重心移到了南方。從流域分佈看，唐代詩人主要生長在黃河流域和環太湖流域，宋代詩人主要分佈在長江流域和環太湖流域。從發展變化看，唐宋時期詩壇重心的南移，始於晚唐五代，完成於北宋，而不是如前人所認為的始於安史之亂，成於靖康之亂。詩歌版圖的分佈與位移，隱藏著許多歷史的奧秘，比如，詩歌版圖重心的位移究竟跟哪些因素有關？是什麼力量、什麼因素推動著詩歌版圖格局的變化？詩歌版圖重心與政治文化中心究竟有什麼關係？一個地方的詩人輩出，詩學薪火相傳不絕，與當地的自然環境、人文環境、社會風尚有什麼樣的關係？這些問題，需要我們去思考和探索。

《唐宋詩歌活動空間的分佈與位移》則是以“唐宋文學編年地圖”的數據為基礎，依據作家的活動空間來對比分析唐宋詩歌版圖的分佈與位移，旨在瞭解唐宋詩人的占籍地理與活動地理有哪些差異與變化？唐宋時期詩歌活動空間的分佈是正態分佈還是偏態分佈？詩歌活動的中心是在什麼區域？是在政治中心的都城還是在遠離政治中心的地方州縣？是在文化經濟相對發達的地區還是在偏遠落後的地區？是在北方的黃河流域還是在南方的環太湖流域，抑或是在南方的長江流域？哪些地方的詩歌活動最為密集頻繁？是在城市還是山林鄉間？是在廟堂還是在江湖？唐代與五代、北宋與南宋詩歌活動空間的

* 中南民族大學，Email: wdwzp@yeah.net

流動趨向有什麼異同或變化？變化的趨勢是由北向南，還是由南向北？或是自東向西？變化的節點是在什麼時間？是什麼原因導致這種變化？詩壇領袖在詩歌活動空間的變化中是否起著引領的作用？詩歌活動空間的變化是否與詩壇領袖的流動變化有關？從唐宋詩歌活動版圖的分佈與變化，能夠發現哪些新的問題？數據統計的結果，將一一回答這些問題。

關鍵字：唐宋詩歌、活動空間、定量分析

DH Tools to Capture and Predict Literary Reading at City-Scale

John Shanahan*

Abstract

I am the creator, and a principal investigator, of the “Reading Chicago Reading” (RCR) project, begun in 2014 and presently supported by grants from the U.S. National Endowment for the Humanities and Microsoft Azure for Research. The project combines methods from the humanities, social sciences (especially sociology and geography), and computer science. While it originates in the study of literary reading in one large American city, it supplies lessons for scholars elsewhere who are interested in tools and methods for combining different types of cultural data for their own DH projects. I have presented some of this work at conferences and universities in the U.S., India, and the People’s Republic of China (see CV for a list), and I hope to have the opportunity to present this project at the 8th International Conference of Digital Archives and Digital Humanities in Taipei.

“Reading Chicago Reading” is a large-scale multi-disciplinary analysis of the popular and much-imitated “One Book, One Chicago” (OBOC) program of the Chicago Public Library (CPL). The Chicago Public Library serves many tens of thousands of citizens each year at its eighty city branches. Since Fall 2001, the CPL has chosen a book each year (originally every six months) around which to organize city-wide public events, book discussions, and other creative programming. The works chosen reflect the diversity of the city’s residents, their cultural heritages, interests, and concerns. Our project began from the observation that the OBOC program acts as a natural experiment -- each chosen work represents a time-stamped probe into library usage and, by extension, a window onto the reading behavior of the library patrons of a major American city. The program’s repetition provides a means of studying reading behavior comparatively over time.

The goal of our work has been predictive insights about literary culture for the use of librarians and DH scholars. Our guiding hypothesis is that text characteristics, library patron demographics, and promotional activities are variables that can be used to predict patron response to future OBOC titles. We know of no other project that combines these data sources to seek a fine-grained model of library patron behavior. A predictive model of this type could be used by librarians at CPL and elsewhere in order to understand the likely uptake of a prospective title by the reading public.

The project focuses on what Wayne Wiegand calls “library in the life of the user,” that is to say, putting the public at the center of discourse regarding the role and function of libraries and capturing libraries’ evolving role in the promotion of literary reading, especially among the

* Associate Dean and Associate Professor of English Literature, DePaul University, Chicago, IL USA. Email: jshanah1@depaul.edu

economically disadvantaged. Our project is exactly what a recent report of the U.S. Institute of Museum and Library Services recommended when it called for partnerships “with allied organizations beyond the library sector” that can “elevate the role of libraries and expand library services to new audiences.” While the “One Book One Chicago” program forms an excellent starting point for our project’s methodology, we expect that our models, tools, and techniques can be applied to library holdings more generally. In recent years, large-scale data mining has become a common practice, but research shows that libraries rarely make resource decisions based on data-driven considerations. Indeed, sociologists found that public libraries regularly choose texts based on “gut” instincts rather than empirical data. “Reading Chicago Reading” is working to create tools to close this knowledge gap, creating a path for the use of large-scale data mining to achieve the broad goal of understanding reading behavior of library patrons across a metropolitan region.

To date, we have focused on connecting our historical circulation data with branch-level demographics. The next phase of our modeling work incorporates content-oriented aspects of the books. This requires creating, for each book, a set of features that represents it for the purposes of modeling. These features should be ones that can be obtained for any book, so that the probable reception of any future library book choice can be predicted. We are now beginning to explore two sources for these features: text content and social media data.

An important new source of data for this project is HathiTrust, a large multi-institution digital library that contains almost 15 million volumes scanned and digitized, with thousands of new titles arriving each month. HathiTrust offers us the possibility of performing large-scale text analysis across many in-copyright texts, something that would not otherwise be possible. We have so far been working almost exclusively with the individual texts chosen as OBOC selections, but it is difficult to draw robust conclusions about them without comparative baselines. HathiTrust released the Extracted Features (EF) Dataset in late 2016, thereby offering page-level features from 13.7 million volumes including (for the first time) in-copyright works from the 20th century. We are exploring whether the features we have found of interest in full-text volumes can also be reliably extracted from the EF data.

We are trying a variety of techniques to quantify the most salient formal features of the books for modeling, including sentiment analysis, topic modeling, and type-token ratio analysis as well as the distribution of different types of verb classes across the texts. (Many of these text measures can be found on our project blog.) One text feature that we originally thought would be important was reading difficulty, a concept for which a wide variety of measures have been proposed by psychologists and linguists. However, these measures have been developed for very different purposes than ours (determining age-suitability for children, for example) and for different types of texts (homogeneous collections of short texts). As a result, to date we have found that these measures alone do not readily capture adult readers’ responses to the wide variety of texts that make up the OBOC selections, ranging from short story collections to novels to nonfiction.

Social media data is another important source of information about books and readers through which we aim to understand the impact of “One Book One Chicago” programming on a diverse readership. We build from studies like Ed Finn’s of literary networks made possible by Amazon.com’s book recommendation system; from Pinder on the social and technological networks that make up LibraryThing.com; and from Dimitrov et al. on the comparisons made

possible by mining Amazon.com and Goodreads.com. Microblogs such as Twitter and other social media sources form a rich and nuanced reflection of (at least some) readers' responses to an OBOC text, responses that supplement and contextualize the raw numerical circulation data we have obtained from the library system. They demonstrate, for example, how discussions around "One Book" texts may drive interest in other related books.

Recently, one of our student researchers extracted and parsed a corpus of just over 25,000 reviews of the seven most recent OBOC titles from the website Goodreads.com. This data includes tags, reviews, and other user-generated data, and represents readers' voluntary responses to a text including emotional reactions that are both positive and negative. From this data, we are exploring sentiment analysis and topic modeling to build representations from these data sources that will support comparison across the texts.

Another element of the project that I will explain is our process for archiving the many materials associated with the "One Book, One Chicago" program. These include CPL system promotional materials and other ephemera, planning documents and receipts, video and audio recordings, and digital records like the Twitter feeds we are already collecting. To date, we have concentrated on materials associated with the past six seasons of OBOC events, because these are ones for which circulation data is available. In the coming year, we will also explore archiving the born-digital materials including CPL's "One Book One Chicago" blog and patron comment areas. We are building an archive of value to scholars seeking to understand mass cultural initiatives of the early 21st century.

All of our code for data extraction and modeling is publicly available as a set of R workbooks from our project blog. We plan to move these files to the open-source code sharing site GitHub so that we can benefit from contributions by open-source developers. As we finish versions of other code elements, we will post those as well.

Quantitative Stylometry of Alice Bradley Sheldon: How Well Did Her Pseudonyms Hide Her True Identity?

Miki Kimura*

Abstract

This paper presents a quantitative authorship attribution analysis on the works of Alice Bradley Sheldon (1915–1987), an American writer of feminist science fiction who used two pseudonyms—James Tiptree, Jr. and Raccoona Sheldon—to disguise her true identity. Given that Alice Sheldon masqueraded as the male James Tiptree, Jr., for almost a decade as a commercial strategy, many critics have discussed the author’s identity and gender. Silverberg (1975) insisted, with reference to the style of Ernest Hemingway, that James Tiptree’s stories were written by a man. In this study, we performed a quantitative stylistic analysis of Sheldon’s work, comparing it with those of a group of male and female writers: Ernest Hemingway, Theodore Sturgeon, Arthur C. Clarke, Ursula K. Le Guin, and Octavia E. Butler. All except Hemingway were science fiction writers whose careers overlapped Sheldon’s. By comparing all of Alice Sheldon’s works with Hemingway’s works and those of her own contemporaries, we hoped to find clues about Sheldon’s allegedly masculine writing style.

The analysis used syntactic variables as discriminants, specifically the distribution of parts of speech in units of single POS, bigrams, and trigrams. One unsupervised methods (using principal component analysis) and two supervised learning algorithms (using support vector machines and random forests) were employed.

The study examined inter-author variations between the six designated authors. Both of the supervised learning methods successfully detected inter-author variations, although the unsupervised methods did not. Furthermore, based on the multi-dimensional scaling (MDS) plot of the random forests results, Alice Sheldon’s style is similar to the masculine style of Ernest Hemingway, just as many literary critics have asserted. However, none of the analyses support the related proposition that Sheldon’s two separate pseudonyms exhibited distinct styles.

Keywords: Authorship Attribution, Alice Bradley Sheldon, Random Forests

* Meiji University, Japan. Email: mk_ling@meiji.ac.jp

Digital Humanities Exploration: Text Mining on Modern Political Discourse and Chinese Classics

Hsaunlei Shao^{*}, Yu-Hsuan Tiffany Wang^{**}

Abstract

In recent years, Taiwan's academic circle has gained a deeper insight into Digital Humanities. In this essay, the authors apply digital concepts and techniques such as text exploration, subject analysis, and the latent Dirichlet allocation (LDA) model, to extract the main themes and build models of the vast breadth of modern political discourse. We attempt to find political messages and values that political leaders have put forth in their speeches. This kind of research is often lumped together with political communications in traditional political science, which has already had a lots of former researchers in this field. Thus, instead of continuing the traditional research on political discourse, this essay applies digital tools to review and examine the speech of modern political leaders to discern and visualize the target context behind the political discourse.

In the past, social scientists tended to use the political discourse of political leaders to observe and predict the behavior of their countries. Although scholars would also take in account conditions such as the international environment and the political and economic circumstances, the political discourse engaged in by leaders is still the most direct proof of the public policies of a country. This is especially true in studies of mainland China, which often struggle with a lack of objective information, and thus must rely on the political discourse of China's leaders for their analyses. But since the interpretation of research is often highly dependent on the experience of scholars, the variability of scholars' interpretative approaches and only a few comprehensive view are problematic for the long-term development of this field. Due to the introduction of more rigorous research methods to Taiwan's academic circle in recent years, these defects have been gradually improving. To that end, the authors believe that the introduction of digital technology will facilitate the continuing improvement of the field.

The authors collected the full corpus of speeches by President Xi Jinping of the People's Republic of China. Using the aforementioned digital methods, the authors ran qualitative and quantitative analyses of a type of modern political discourse. This study can also provide an entry point for the observation of Chinese politics, which is imperative to the academic circle in Taiwan.

Keywords: Text Mining, Mainland China Studies, Digital Humanities, Latent Dirichlet allocation (LDA), Political Discourse, Visualization

^{*} Norwegian University of Science and Technology, Email:hlshao2@gmail.com

^{**} Norwegian University of Science and Technology, Email:tiffanywang0815@gmail.com

Taiwan Biographical Ontology (TBIO): A Working Annotation Scheme for Biographical Information Extraction and its Application to Sociology of Literature

Alvin Cheng-Hsien Chen^{*}, Táňa Dluhošová^{**}

Abstract

Biographical information has been widely recognized for its usefulness in the sociological historical studies. Existing biographical databases in Taiwan often provide only descriptive texts, which may not be directly machine-readable for sophisticated digital methods and analyses. We propose developing a working scheme for a semi-automatic extraction of useful biographical information from existing texts. Taking the early post-war writers from the Taiwan Writers Catalogue (TWC) database (including 2600 biographies) as our data source, we propose a three-stage annotation scheme for data extraction to construct Taiwan Biographical Ontology (TBIO), with an ultimate goal to uncover structuring principles on literary field in Taiwan (1945-49). In the first stage, following the principle of Chinese Proposition Bank, we choose from the TWC a selection of core writers as our training set and manually identify “informative” clauses from their biographical narratives, each of which is semantically annotated with semantic roles and main predicates. The “informative” clause is defined as a grammatical unit that encodes social relations or networking of one agent to another socially-meaningful entity. Secondly, we proceed with the semantic role labeling for the other agents in the TWC in a semi-automatic way, utilizing the state-of-the-art dependency parser. In the final stage, each argument identified in semantic role labeling is further ontologically annotated, categorized and encoded in the Web Ontology Language (OWL). This primitive TBIO will be used to update our previous analysis of the literary scene of early post-war Taiwan, shedding light on the potentials and capacities of TBIO for future studies.

Keywords: Biographical Ontology, Prosopography, Annotation Scheme, Semantic Role Labeling

^{*} Department of English, National Taiwan Normal University. Email: alvinworks@gmail.com

^{**} Oriental Institute, Czech Academy of Sciences, Czech Republic. Email: tana.dluhosova@gmail.com

大數據環境下數字檔案管理趨向探析： 基於以內容管理為核心的六何分析法

郭先敏*

摘 要

檔案事業在大數據的工具和思維助力下有了長足發展，尤其體現在從實體管理像內容管理邁進。國內外對檔案內容管理的研究多着重技術方案和具體系統，鮮見基於全局的整體分析。本文利用分析與規劃中常見的六何分析法，結合當下中國大陸的數位檔案整體情況，以內容管理池（WHAT）和管理鏈（HOW）為核心，在管理驅動力（WHY）、場域（WHERE）、時間（WHEN）以及主體（WHO）的共同作用下，釐清六何元素邏輯關係，並構成完整的檔案內容管理模型。這有助於檔案內容管理的工作規劃和層次分拆，具有一定的理論與實際意義。

關鍵字：數位檔案、內容管理、六何分析法、大數據

* 中國中山大學資訊管理學院博士生，Email: 413866201@qq.com

目錄型文獻轉化為分析型數據的步驟與方法： 以《20世紀唐代文學研究論著目錄檢索系統》為例

邵大為*

摘要

“數字人文”是目前國際最具潛力的新興學科和前沿研究領域。在數位人文研究中，電腦的功能轉變引人注目。從最初幫助人文研究者海量、永久保存文獻的存儲功能，到強大的檢索、分析、統計功能，電腦不僅解決了許多人文研究者們在實際研究過程中可能遇到的具體問題，更啟發研究者深化研究方向、生發新的選題和思路，形成一種不同于傳統人文研究的新思維。從研究“工具”到研究“思維”，電腦扮演的角色發生了根本性轉變，使得“數字人文”不再意味著“數字”和“人文”的簡單相加，而掀開了跨學科之間真正相互融合、相互啟發的新紀元。

以筆者從事的唐宋文學研究領域為例，20世紀以來唐代文學研究雖然積累了豐富的研究成果，但因數量龐大，這些成果不可能被當代及今後研究者們一一閱覽，既遮蔽了許多研究盲區，也造成很多研究課題的重複。儘管有相關專業檢索目錄，但囿于傳統排版和紙質印刷，極大限制了檢索目錄的利用價值。筆者有感于此，欲借助電腦的強大功能，挖掘這些“目錄型文獻”背後的隱含資訊，使其變成能夠輔助研究者更細緻瞭解唐代文學研究現狀的“分析型數據”。在這一目標驅使下，筆者帶領團隊製作了“20世紀唐代文學研究論著目錄檢索系統”（以下簡稱《檢索系統》）。本文就以這段實際經歷為例，介紹將“目錄型文獻”轉化為“分析型數據”的具體步驟和方法，以期對其他相關數據庫建設者提供實際操作經驗。

文章首先介紹製作《檢索系統》的“工作步驟”，依次從功能設計、錄入校對、分類標注和系統設計四個環節展開。系統功能原本是要最後一個環節才能實現，但在操作層面，必須事先規劃、設計好，《檢索系統》至少應有檢索、統計、查重和匯出四大功能。錄入校對是項目中任務量最大也最需要細心和耐心的環節，原始資料有可能在錄入過程中出現異文和錯誤，這些都要經過統一化或去重處理。分類標注是和《檢索系統》今後的使用者關係最為密切的環節，數據標注分原始信息和衍生信息兩大類，具體標注哪些性質的信息直接影響今後使用者能在什麼方向和多大程度上利用《檢索目錄》。最後一個環節是系統設計，此處由程式設計人員充分瞭解設計者的意圖後，進行專業設計，然後再不斷調試優化。

其次介紹《檢索系統》的“使用方法”。依次從最初設計的檢索、統計、匯出、查重四大功能展開介紹。檢索功能分單一條件檢索和組合條件檢索，前者包括按時間檢索、按作者檢索、按作品檢索、按來源檢索、按對象檢索、按文體檢索、按單篇作品檢索、按學

* 中南民族大學，Email: shaodawei1985@126.com

者地域檢索和按成果形式檢索等九種形式，後者則是在單一檢索的基礎上，根據使用者需求進行邏輯疊加檢索。統計功能分全部統計和局部統計兩種，前者指不設限範圍，就全部成果目錄進行統計，局部統計是設置範圍，就特定範圍進行統計，如只統計某個年度、某個地區、某位元作者、某個對象的成果量等。匯出功能指使用者需要的統計結果可從系統中匯出成 Excel 表格，進行編輯，方便進行下一步研究和撰寫論文。查重功能是指檢索系統應能夠自動分析和提取那些被重複錄入或重複出現的作品，以供選擇和汰除，避免重複統計，影響統計結果。

本文是作者及團隊在進行數據庫建設里程中的一小段經歷，在製作《檢索系統》的過程中，作者既是設計者，又是假定的使用者，如何借力電腦本身的結構和程式師的技術，將設計者的理念和用戶的需求完美結合並呈現出來，是建設數據庫最有趣的地方，不經過實際操作難以有深刻感觸。本《檢索目錄》雖是個案，帶有本學科的具體特性，但在理念設計和一般操作上有普適性，希望在操作層面對數字人文研究有所助益。

關鍵字：目錄型文獻、分析型數據、數據庫建設、20 世紀唐代文學研究論著目錄檢索系統

中國近代「個人」觀念的形成與演變： 以非監督式學習法為主

鄭文惠*、梁穎誼**

摘要

傳統中國天下觀的社會運作機制依循著以倫常等級所建立的家國同構的道德共同體作為正當性理據與合法化基礎。然而，近代中國邁向現代性歷程中，知識分子在跨語際實踐過程中援引中國傳統及歐西與日本等各方思想資源，一方面關注於在由天下進入世界體系中，如何與帝國霸權的文化擴張進行政治頡頏、文化協商，及如何與全球化市場經濟無限制的發展和生產力超增長過程，聯結為一個總體而連動的關係；一方面進行價值改造，在標舉不同意識形態的國族認同的同時，注重個人權利的核心價值。其中個人權利作為一種新理據及新道德，在近代中國是一個全新的概念，而個人觀念的形成與變遷，又與新型國家的建制、國民的素質、權利觀念、主體意識等等息息相關。近代中國一切的變革，實質就源自於在劇烈的社會文化轉型與高度的跨文化流動下，如何透過跨語際實踐的過程，面對西方價值理性和工具理性二者合力下所產生的巨大衝擊，從而建構一套新的正當性理據或合法化基礎，才不致於導致原有社會全然的崩潰或解體。

傳統小我的價值是在君王政權結構及家國大我敘事框架中被範定，而具有本位意識的獨立個人——唯我式個人主義，是伴隨著近代中國王權解體、宗法制度鬆動，及議會民主制度失敗、第一次歐戰的殘暴本質與戰後談判權益受損等等因素而興起。中國知識分子從國家主義中覺醒，從而揚棄了大我敘事，促成 1915 年後個人的覺醒。近代中國在國家主義退潮的當下轉而注重個人的價值，實質上是一場場隨順著不同事件而緊密連動的話語行動。個人觀念在「天」、「帝」、「國家」、「群體」等概念解構後湧現，象徵著隨著國家主義的幻滅，大我敘事的消解，促成近代中國由「專制王朝」到「君主立憲」、由「臣民」到「國民」、由「家國大我」到「私情小我」等制度與觀念的巨大改變與轉向。這是面對中國傳統社會總體結構的巨大解體與文化秩序、心靈道德的多重性危機，及在帝國霸權夾縫中與世界視野下自我與他者異置而共在的自我生存與修復調節機制的轉向。然而，即使個人觀念崛起，又恆常是作為召喚國族巨靈的主要憑藉，個人往往又積極介入國家、民族話語的建構。近代中國在邁向現代性歷程中，個人與國族無不彼此牽動著、相互助力著，因而，探討近代中國個人觀念的形成與變遷，除了逼顯中國現代個人觀念與主體意識的建構，有其形而上的哲學意義外，在跨文化流動中透過文化斡旋的多重力的作用下，也可以廓清近代中國國族認同與國家、個人觀念形成與轉型的過程。

職是，本文主要研究目的有二，一是在數位技術上的推進，二是在人文議題上的發展。

* 政治大學中國文系教授，Email: whcheng123@gmail.com

** 政治大學中國文系博士後研究員，Email: htiosc85@gmail.com

在數位技術上的推進部分，主要針對過去運用 CUSUM 研究方法過程中所發現方法不足之處進行調整與推進。其一，過去只能進行四十個關鍵詞左右的分群研究，本研究希望能在技術上革新，進行至少一百個關鍵詞以上的分群研究；其二，過去 CUSUM 方法係利用關鍵詞歷時性線性增長時間相似性進行分群，而本研究則希望加入統計學中關聯性相似性進行分群研究，以協助人文研究者有更為客觀的分群標準，解決過去某些關鍵詞在視覺判決上未能有效分群的研究侷限。例如過去研究近代中國「思想」一詞的分群，發現它大致能視為與「國民」、「進步」、「人民」為一群的觀念，但由於時間曲線不明顯，導致視覺判決上具限制。因此若考慮 CUSUM 曲線之間的相似性，使用一些常用的指標如 Cosine、Jaccard 等相似指數加以分類，當能使 CUSUM 方法更具可重複操作性。

而在人文議題上的發展部分，主要以上述方法分析中國近代報刊、雜誌、專著中的觀念軌跡，如《新青年》、《少年中國》、《現代評論》、《新民叢報》、《新潮》等具有歷史代表意義之不同類型的語料，結合長時段中國近代個人觀念的發展，進行一宏觀且結構式的考察。本研究主要觀察從晚清到民國之際，「個人」觀念處在公/私、傳統/現代、理想/現實等三大兩歧性語境中的變化軌跡，並進一步探討個人觀念論述背後所牽涉到的話語、事件、行動與觀念系統的關係，除了推進中國近代「個人」觀念的研究外，也冀望從「個人」語料的歷時性考察對比中，勾勒出中國近代知識分子對「個人」的現代性想像，並探討中國近代大我敘事框架，如天、帝、國家、群體等概念如何解體？小我價值及個人權利如何湧現（emerge）？現代私情小我的唯我式個人主義如何取代傳統家國大我的發展軌跡？研究「個人」觀念的必要性與重要性，在於個人與個人主義的崛起乃是現代性的標誌，因此唯有掌握中國近代具有本位意識與主體建構的個人觀念的發展與變化，才能明白中國如何從傳統走向現代的發展歷程。

關鍵字：個人、概念、觀念史、非監督學習、相似度測量、集群方法

數位典藏系統初探：以國家圖書館為例

魏家惠*、洪偉翔**、鄭基田***

摘 要

數位科技越來越進步，漸漸的已經影響我們閱讀的習慣，從原本閱讀紙本書到現在閱讀電子書，而早期的紙本書保存不易，透過掃描為數位化的方式典藏，而數位典藏系統在建置時，面臨其中兩個問題，首先為了要呈現給讀者閱讀，必需要將原始的數位出版品檔案轉製成可以在電腦、智慧型手機、平板電腦可以閱讀的模式，然而數位出版品檔案格式多元，尚未出現標準統一格式，因此如何建立一個機制收藏電子書檔案格式。另一個問題，如何控制數位版權，出版單位出版電子書有各式各樣的目的，出版社或個人出版者出版電子書是為了販售，政府單位或財團法人出版電子書是為了宣傳或推廣讓大家知道，因此本文提供授權機制滿足各出版者的需求又可以免費提供閱讀，同時兼顧保護出版單位銷售的權益不提供閱讀，故能滿足各方的需求。

關鍵字：數位典藏、數位版權管理、數位出版品、電子書、數位學習

* 國家圖書館編輯，Email: chwei.phd@gmail.com

** 國家圖書館助理編輯，Email: weldon@ncl.edu.tw

*** 國家圖書館編輯，Email: robin0909@gmail.com

Near Synonymous Body-Part Terms in Chinese: A Corpus-Based Study

Hsiao-Ling Hsu^{*}, Huei-Ling Lai^{**}, Jyi-Shane Liu^{***}

Abstract

Forms and meanings have been the core in the field of linguistics, while few studies have looked into the word structures (forms) of metonymies or metaphors in Chinese. If the components of words are examined in terms of their form class identities, the systematic knowledge of understanding and using the words can be uncovered. Human body-parts are regarded as the most primary source people employ to understand abstract concepts. Indeed, body-part terms are pervasively found cross-linguistically to express concepts such as personality, emotion, and quality. The English expression - *pull/wear/put on a long face* and the Chinese expression-*bǎnzhe liǎn* 板著臉 ‘to look sad, glum or disapproving’ can illustrate. Many studies on metaphoric and metonymic expressions have suggested that a bodily and psychological basis is employed to create metaphors or metonymies (e.g., Yu 2002). However, previous studies are mostly based on linguistic data limited to dictionaries or underdeveloped corpora, leading to their findings being based more on theoretical reasoning rather than an empirically-based treatment (e.g. Tsai 1994; Yu 2008, 2011b; Wen & Wu 2007; Qin 2008; Zhao 2010; Xie 2011; Hung & Gong 2011). Thus, this study incorporates computational linguistics approaches, corpus-based approaches, and cognitive semantic theories with the aim to better capture the dynamicity and the generalized patterns exhibited by language use. Words containing three sets of body-part terms in Chinese--*face* (*liǎn* 臉 / *miàn* 面), *eye* (*yǎn* 眼 / *mù* 目), and *mouth* (*zuǐ* 嘴/ *kǒu* 口)—are analyzed in terms of their word component structures and meaning extensions. The three sets of body-part terms are peculiar since they are counterparts to each other, and are the most salient body-parts people utilize to understand and interact with the world. The investigation of their usages will be a good starting point to understand human cognition.

Our data are all extracted from the Academic Sinica Balanced Corpus of Modern Chinese 4.0. Instances containing either one of the six target body-part terms are extracted. First, we search the target body-term *liǎn/miàn*, *yǎn/mù*, or *zuǐ/kǒu* in the corpus. Then, the words containing the

* National Chengchi University, Graduate Institute of Linguistics, Email: heidimavis@hotmail.com

** National Chengchi University, Department of English, Email: hllai@nccu.edu.tw

*** National Chengchi University, Department of Computer Science, Email: liujsh@nccu.edu.tw

body-part terms are extracted according to the CKIP segmentation. Lastly, the data are exhibited in the excel templates with the important information: word, part of speech, and frequency. Based on our research aims and the information of our raw data, three steps are taken: Class Identification: The form classes of the six terms are identified as they are listed in authoritative *Chinese Dictionary* and *Chinese Wordnet*.

- (i) Structural Analysis: the form classes of the internal components of the disyllabic words are then analyzed based on the components' role and position within that word and the form classes as they are listed in the Chinese Dictionary and Chinese Wordnet. For example, a word like *diūliǎn* 丟臉 'lose face' is analyzed as a verb [V] composed of a verb and a noun elements ([VN]V); *yǎnjìng* 眼鏡 'glasses' as a noun [N] constructed by two noun elements ([NN]N); *zhèngmiàn* 正面 'front' as a noun [N] composed of an adjective and a noun ([AN]N).
- (ii) Meaning Identification: The various meanings are categorized based on their linguistic context and their usages in the real-world context. Words that have compositional meanings such as *zuǐchún* 嘴唇 'lips' is analyzed as words which denotes literal meanings. Those that have non-compositional meanings are analyzed as denoting non-literal meanings. In general, the disyllabic word is used as a compound and the meaning of that neighboring component helps embed certain meanings in the target body-part term. Combining the meaning of the target body-part term and its neighboring component, the disyllabic word is used to denote a certain metonymic or metaphorical meaning. For instance, in the word *shǎngliǎn* 賞臉 'to honor someone', it is *shǎng* 賞 that helps embed the metonymic meaning DIGNITY in the word.

The results show that the [NN]_N structure is the most dominant word component structure across the six terms as in (1). Except for words containing *miàn*, words that contain the other five body-part terms show the second highest percentage of occurrences in the [VN]_V structure, as in (2). In the data, only words that contain *miàn* occur in the [VV]_{w1} structure manifesting two syntactic categories: either [VV]_N (as in 3a) or [VV]_V (as in 3b).

(1) Instances in the [NN]_N structure

- a. *tóuliǎn* 頭臉 'person'
- b. *rénmiàn* 人面 'person'
- c. *yǎnméi* 眼眉 'spy'
- d. *tóumù* 頭目 'leader of a tribe'
- e. *míngzuǐ* 名嘴 'pundit'
- f. *hùkǒu* 戶口 'number of households'

(2) Instances in the [VN]V structure

- g. *fānliǎn* 翻臉 ‘suddenly get mad’
- h. *biànmian* 變面 ‘suddenly get mad’
- i. *yàoyǎn* 耀眼 ‘dazzling’
- j. *dèngmù* 瞪目 ‘stare in anger’
- k. *tiāozuǐ* 挑嘴 ‘be fussy about food’
- l. *jìkǒu* 忌口 ‘be on a diet’.

(3) Instances in the [VV]W1 structure

- a. *miànshì* 面世 ‘to be published’ [VV]_N
- b. *miànlín* 面臨 ‘confront’ [VV]_V

In terms of their meaning distribution, our results show that *miàn*, *mù*, and *kǒu* are used more frequently to denote non-literal meanings than their counterparts – *liǎn*, *yǎn*, and *zuǐ*. A sharing of labor in meaning is detected: Between the two candidates that refer to the same body-part, one is literal-meaning-oriented (e.g., *liǎn*, *yǎn*, and *zuǐ*) and the other is non-literal-meaning-oriented (e.g., *miàn*, *mù*, and *kǒu*).

In detail, five metonymic extensions are found for *liǎn*: PERSON, EMOTION, CHARACTER, DIGNITY, and APPEARANCE, as in examples (4a), (5a), (6a), (7a) and (8a); and seven are found for *miàn*: PERSON, EMOTION, CHARACTER, DIGNITY, APPEARANCE, CONFRONT, and QUANTITY, as in examples (4b), (5b), (6b), (7b), (8b), (9), and (10). In addition, five are found for both *yǎn* and *mù*: PERSON, EMOTION, ABILITY, INTELLECTION, and FOCUS, as in examples (11) – (15). Moreover, five are found for *zuǐ*: PERSON, EMOTION, FLAVOR, UTTERANCES, and CHARACTER, (as in examples (16a), (17a), (18a), (19a), and (20a)) and five are found for *kǒu*: PERSON, EMOTION, FLAVOR, UTTERANCES, and QUANTITY (as in examples (16b), (17b), (19b), (20b), and (21)).

(4) PERSON

- a. *tóuliǎn* 頭臉 ‘person’
- b. *rénmiàn* 人面 ‘person’

(5) EMOTION

- a. *fānliǎn* 翻臉 ‘suddenly get mad’
- b. *miànróng* 面容 ‘countenance’

(6) CHARACTER

- a. *hēiliǎn* 黑臉 ‘appear as the hatchet man’
- b. *jiǎmiàn* 假面 ‘masked’

(7) DIGNITY

- a. *diūliǎn* 丟臉 ‘lose face’
- b. *qíngmiàn* 情面 ‘face-saving’

(8) APPEARANCE

- a. *lòuliǎn* 露臉 ‘show up’
- b. *lòumiàn* 露面 ‘show up’

(9) CONFRONT (V.)

- miànshì* 面世 ‘to be published’

(10) QUANTITY

- yímiàn* 一面 as a classifier

(11) PERSON

- a. *yǎnméi* 眼眉 ‘spy’
- b. *tóumù* 頭目 ‘leader of a tribe’

(12) EMOTION

- a. *lèiyǎn* 淚眼 ‘tearful eyes’
- b. *dèngmù* 瞪目 ‘stare in anger’

(13) ABILITY

- a. *yīngyǎn* 鷹眼 ‘sharp eyes’
- b. *yīngmù* 鷹目 ‘sharp eyes’

(14) INTELLECTION

- a. *míngyǎn* 明眼 ‘discerning eye’
- b. *mùguāng* 目光 ‘vision’

(15) FOCUS

- a. *yàoyǎn* 耀眼 ‘dazzling’
- b. *duómù* 奪目 ‘to catch one’s eyes’

(16) PERSON

- a. *míngzuǐ* 名嘴 ‘pundit’
- b. *hùkǒu* 戶口 ‘number of households’

(17) EMOTION

- a. *dūzuǐ* 嘟嘴 ‘pout’
- b. *pīkǒu* 劈口 ‘yell at someone suddenly’

(18) CHARACTER

tiězuǐ 鐵嘴 ‘with a iron and bitter tongue’

(19) FLAVOR

a. *tiāozuǐ* 挑嘴 ‘be fussy about food’

b. *kǒugǎn* 口感 ‘texture’

(20) UTTERANCES

a. *dǐngzuǐ* 頂嘴 ‘talk back’

b. *kǒucái* 口才 ‘ways of talking’

(21) QUANTITY

yìkǒu 一口 ‘a bite’

Furthermore, no metaphorical meanings are found for *liǎn* while three are found for *miàn*: SURFACE, CONDITION and SIDE, as in examples (22)-(24). In addition, four metaphorical meanings are found for both *yǎn* and *mù*: NEAR SPACE, NEAR TIME, TINY HOLE, and CENTER, as in examples (25)-(28). But *mù* denotes two more: CONDITION and ENTRY, as in examples (29)-(30). Moreover, as illustrated in (31a), only one metaphorical is found for *zuǐ*: MOUTH-SHAPED, while *kǒu* denotes three: MOUTH-SHAPED, NEAR TIME, and GATEWAY, as in examples (31b)-(33).

(22) SURFACE

dìmiàn 地面 ‘ground’

(23) CONDITION

shìmiàn 市面 ‘market conditions’

(24) SIDE

běimiàn 北面 ‘the north side’

(25) NEAR SPACE

m. *yǎnqián* 眼前 ‘at present’;

n. *mǎnmù* 滿目 ‘meet one’s eyes on every side’

(26) NEAR TIME

a. *yìyǎn* 一眼 ‘see something at a glance’

b. *mùqián* 目前 ‘at this moment’

(27) TINY HOLE

a. *wǎngyǎn* 網眼 ‘tiny hole of a net’

b. *wǎngmù* 網目 ‘tiny hole of a net’

(28) CENTER

- a. *cíyǎn* 詞眼 ‘motif of lyrics’
- b. *gāngmù* 綱目 ‘outline’

(29) CONDITION

méimù 眉目 ‘take shape’

(30) ENTRY

shūmù 書目 ‘bibliography’

(31) MOUTH-SHAPED

- a. *húzuǐ* 壺嘴 ‘spout’
- b. *kǒudài* 口袋 ‘pocket’

(32) GATEWAY

jìnkǒu 進口 ‘import’

(33) NEAR TIME

dāngkǒu 當口 ‘at present’

The interactions of the word component structures and the meaning distribution can be explicated as follows. In general, the six body-part terms are all used to denote the metonymic meanings PERSON and EMOTION. It is found that when the six target body-part terms denote the metonymic meaning PERSON, the highest frequency of their word component structures occurs in [NN]_N. When *liǎn*, *miàn*, and *yǎn* denote the metonymic meaning EMOTION, the [NN]_N structure also shows the highest frequency of occurrences, while [VN]_V or [VN]_D shows the highest frequency of occurrences when *mù*, *zuǐ*, and *kǒu* denote EMOTION.

The analysis evidences that in each pair the target body-part terms are near synonymous to each other, exhibiting similarities and differences. For face-denoting, both *liǎn* and *miàn* share the same five metonymic meanings, but they contrast in their individual preference in denoting certain metonymic meaning: *liǎn* mostly used in a nominal construction for expressing EMOTION, while *miàn* mostly used as a predicate expressing CONFRONT (V.). For eye-denoting, both *yǎn* and *mù* are used to denote the same five metonymic meanings: PERSON, EMOTION, ABILITY, INTELLECTION, and FOCUS, they contrast in their individual preference in denoting certain metonymic meanings: *yǎn* mostly used to express EMOTION, while *mù* mostly used to express FOCUS. For mouth-denoting, both *zuǐ* and *kǒu* are used to denote the same four metonymic meanings: PERSON, EMOTION, FLAVOR and UTTERANCES, they have their own preference in denoting certain metonymic meanings: *kǒu* used more frequently than *zuǐ* in denoting PERSON. Only *zuǐ* is used to express CHARACTER and only *kǒu* is used to express QUANTITY.

As for the metaphorical meanings, liǎn denotes no metaphorical meanings, while three are found for miàn: SURFACE, CONDITION and SIDE, with SIDE occurring at the highest frequency. Yǎn and mù both denote NEAR SPACE, NEAR TIME, TINY HOLE, and CENTER. Yǎn used to denote NEAR SPACE the most frequently, while mù used to denote NEAR TIME the most frequently. However, some differences are still found: mù denotes two more CONDITION and ENTRY. Both zuǐ and kǒu are used to denote MOUTH-SHAPED, while kǒu has two extra GATEWAY and NEAR TIME, indicating that kǒu is more likely to be extended in use. Zuǐ is only used to denote MOUTH-SHAPED, while kǒu shows the highest percentage to denote GATEWAY.

The discussion of meaning extensions gives three important insights. First, the six body-part terms are all used to denote PERSON and EMOTION. Second, the metonymic meanings the six body-part terms denote may be strongly correlated with the physiological functions the three body-parts (face, eye, and mouth) possess. Third, some metaphorical meanings may be developed mostly based on the physiological shapes of the three body-parts. That PERSON and EMOTION are the most prevalent metonymic meanings may indicate that the three body-part (face, eye, and mouth) share two similar functions in social interactions: to represent individual identity and to show emotions. When denoting PERSON, most of them belong to PART FOR WHOLE

type of metonymy. Cases, such as tóumù 頭目 ‘leader of a tribe’, míngzuǐ 名嘴 ‘pundit’, and hùkǒu 戶口 ‘number of households’ can illustrate. In denoting EMOTION, most words are constructed are predicates, indicating emotions can be exhibited through the change of the state on the body-parts. Cases such as fānliǎn 翻臉 ‘suddenly get mad’ and dūzuǐ 嘟嘴 ‘pout’ can illustrate (face and mouth in these two cases). In addition, it appears that the metonymic meanings are not randomly created but are developed based on the physiological or social functions of each body-part. First, in addition to PERSON and EMOTION, the pair for face-denoting, liǎn and miàn, both denote CHARACTER, DIGNITY, and APPEARANCE. Our face can reveal our emotion and even our character, and it is where we can wear a mask to act like a certain character (CHARACTER). In addition to physiological face, a social face is related to our dignity (DIGNITY and APPEARANCE). Then, the pair for eye-denoting, yǎn and mù, both denote ABILITY, INTELLECTION, and FOCUS. The most important physiological function of eyes is our eyesight (ABILITY). And eyes are one of the most important organ for perception; thus, something important would first be noticed by our eyes (FOCUS). Because one of the ways to become intelligent is to read, we need our eyes to read to gain intellection (INTELLECTION). Moreover, the pair for mouth-denoting, zuǐ and kǒu, both denote FLAVOR and UTTERANCES, which coincides with the physiological functions of mouth: to taste (FLAVOR) and to speak

(UTTERANCES). Furthermore, some metaphorical meanings may be mostly based on the physiological shapes of the three body-parts. First, SURFACE that miàn denotes may be created based on the flat shape of faces. Second, TINY HOLE that yǎn and mù denote may be based on the shape of eyes. And MOUTH-SHAPED and GATEWAY are based on the shape of mouths and the physiological function of mouths, which is as an opening in the digestive system. But, it should be noted that although based on the physiological shapes and functions of the three body-parts, metonymic meanings are developed in the same domain (i.e. human), while metaphorical meanings are developed and mapped into different domains.

This current study adopts an innovative approach to investigate the relationship between the word component structures and the meaning extensions. The syntagmatics in meaning extensions have seldom been at the core of study in previous studies of metonymies or metaphors in Chinese. Most studies focus on the mappings between the source domain and the target domain and mostly focus on theoretical discussion. Even though some studies have investigated the syntactic structures of metonymies or metaphors, little research has looked into the structures inside a word, namely, the word component structure. This study decomposes the words into their components, analyzes these components based on their form class identities, and identifies their meaning extensions. This approach provides a systematic way to generalize the word component structures and their corresponding meaning extensions. In addition, this study provides evidence for the claim that the target body-part terms are counterparts and near-synonyms to each other in each pair, and distinguishes the similarities and differences in them. It is expected that the results may provide some insights for future study to refine procedures to help identify metaphors and metonymies in corpora and then to accelerate quantification and statistical evaluation.

References

- Cameron, L., and Deignan, A. (2006). The emergence of metaphor in discourse. *Applied linguistics*, 27(4), 671-690.
- Huang, C., Chen, K., Chang, L. and Hsu, H. (1995). An Introduction to Academia Sinica Balanced Corpus. [In Chinese]. *Proceedings of ROCLING VIII*. 81-99.
- Hung, Y. Y., and Gong, S. P. (2011). A Cross-Linguistic Approach to Body-part Metonymy in English and Chinese. *Effects of an On-line Syntactic Analysis Strategy Instruction on University Students' Reading Comprehension of English Science Texts*, 16, 146.
- Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.
- Qin, Xiugui. (2008). The Conceptual Metaphors of “eye” - A Comparative Study Based on the Corpus Between English and Chinese. *Journal of Foreign Languages*, 31(5), 37-43.
- Tsai, L. C. (1994). The Metaphor of Body-parts in Chinese. *Unpublished Master of Arts Thesis*, National Tsing-Hua University, Taiwan.
- Wen, X., and Wu, S. Q. (2007). A Study on the Cognitive Features of the Metaphoric Expressions Involving the English and Chinese FACE [J]. *Journal of Southwest University (Social Sciences Edition)*, 6, 028.
- Xie, H.Z. (2011). Comparative Study on Metaphorical Meaning of Vocabulary with the Denotation of “ face” between Chinese and English. *Unpublished Master of Arts Thesis*, Central China University, China.
- Yu, N. (2002). Body and emotion: Body-parts in Chinese expression of emotion. *Pragmatics & cognition*, 10(1), 341-367.
- Zhao, X.D. (2010). A Cognitive Study on Meaning Transference of Body-Part Terms. *Unpublished Doctoral Dissertation*, Fudan University, China.

Table 4.1 The distribution of metonymic meanings among the six body-part terms

<i>Metonymic meanings</i>	<i>Liǎn</i>	%	<i>Miàn</i>	%	<i>Yǎn</i>	%	<i>Mù</i>	%	<i>Zuǐ</i>	%	<i>Kǒu</i>	%	<i>Total</i>
PERSON	3	0.52	741	11.94	5	0.23	65	5.57	6	3.55	1345	33.48	2165
%	0.14		34.23		0.23		3.00		0.28		62.12		100
EMOTION	415	71.31	306	4.93	837	38.38	146	12.51	24	14.2	20	0.5	1748
%	23.74		17.51		47.88		8.35		1.37		1.14		100
CHARACTER	54	9.28	13	0.21					29	17.16			96
%	56.25		13.54						30.21				100
DIGNITY	78	13.4	245	3.95									323
%	24.15		75.85										100
APPEARANCE	32	5.5	1058	17.05									1090
%	2.94		97.06										100
CONFRONT(V.)			3835	61.8									3835
%			100										100
QUANTITY			8	0.13							74	1.84	82
%			9.76								90.24		100
ABILITY					288	13.2	16	1.37					304
%					94.74		5.26						100
INTELLECTION					667	30.58	411	35.22					1078
%					61.87		38.13						100
FOCUS					384	17.61	529	45.33					913
%					42.06		57.94						100
FLAVOR									26	15.38	657	16.36	683
%									3.81		96.19		100
UTTERANCES									84	49.7	1921	47.82	2005
%									4.19		95.81		100
Total	582	100	6206	100	2181	100	1167	100	169	100	4017	100	

Table 4.2 The distribution of metaphorical meanings in the six body-part terms

<i>Metaphorical meanings</i>	<i>Liǎn</i>	<i>Miàn</i>	%	<i>Yǎn</i>	%	<i>Mù</i>	%	<i>Zuǐ</i>	%	<i>Kǒu</i>	%	<i>Total</i>
<i>SURFACE</i>		4016	22.55									4016
%		100										100
<i>CONDITION</i>		3214	18.04			3	0.02					3217
%		99.91				0.09						100
<i>SIDE</i>		10582	59.41									10582
%		100										100
<i>NEAR SPACE</i>				895	76.89	296	1.69					1191
%				75.15		24.85						100
<i>NEAR TIME</i>				85	7.3	8909	50.86			26	0.71	9020
%				0.94		98.77				0.29		100
<i>TINY HOLE</i>				12	1.03	23	0.13					35
%				34.29		65.71						100
<i>CENTER</i>				172	14.78	4020	22.95					4192
%				4.10		95.9						100
<i>ENTRY</i>						4266	24.35					4266
%						100						100
<i>MOUTH-SHAPED</i>								35	100	438	11.90	473
%								7.40		92.6		100
<i>GATEWAY</i>										3216	87.39	3216
%										100		100
<i>Total</i>		17812	100	1164	100	17517	100	35	100	3680	100	

Baidu News Information Flow and Return Volatility: Evidence for the Sequential Information Arrival Hypothesis

Dehua Shen^{*}, Xiao Li^{**}, Wei Zhang^{***}

Abstract

This paper employs Baidu News as the proxy for information flow and investigates competing hypotheses on the relationships between information flow and return volatility in Chinese stock market. The empirical results show that: (1) trading volume and return volatility are not driven by the same variable, i.e., the information flow, and thus contradicts the predication of the Mixture of Distribution Hypothesis (MDH); (2) there exist significant lead-lag relationships between information flow and return volatility, which is in accordance with the Sequential Information Arrival Hypothesis (SIAH); (3) these findings are robust to alternative measurement of return volatility and subsample analysis. Generally speaking, these findings contradict the prediction of MDH and support the SIAH.

Keywords: Return volatility; Sequential Information Arrival Hypothesis; Mixture of Distribution Hypothesis; Information flow; Baidu New

* Tianjin University, China, Email: dhs@tju.edu.cn

** Nankai University, China, Email: xiaoli@nankai.edu.cn

*** Tianjin University, China, Email: weiz@tju.edu.cn

William Faulkner's Fictional Demographics: Mapping People and Places Using the Digital Yoknapatawpha Database

Johannes Burgers*

Abstract

Introduction

The Nobel-prize winning American author William Faulkner (1897-1962), spent the bulk of his literary career writing about his native Mississippi. In his fiction, he transformed his town, Oxford, and county, Lafayette, into the imaginary spaces of Jefferson and Yoknapatawpha. He populated his now-famous county with a teeming and robust cast of characters who reappear throughout the 14 novels and 54 different short stories specifically set there. The resulting literary works like *The Sound and the Fury*, *As I Lay Dying*, and *Absalom, Absalom!* are considered masterpieces of twentieth century American literature and are taught every semester in high school and college classrooms around the US and, indeed, around the world.

Operational since 2012, the Digital Yoknapatawpha Project (hereafter DY), aims to turn Faulkner's Yoknapatawpha fictions into searchable data, and then visualize that data through maps and other digital displays. Specifically, the project plots Faulkner's characters, locations, and events onto an atlas of deep maps based on the maps Faulkner himself drew in 1936 and 1946. Currently, the main visualization is still a work in progress and viewable here: <http://faulkner.iath.virginia.edu>. This presentation uses the character database to display demographic overviews of Faulkner's fiction with Quantum GIS. What is revealed in doing so is that Faulkner's fictional spaces are highly gendered, racialized, and class-based. Of particular interest is the way in which Faulkner's conception of these spaces changes over time. These changes, I suspect, not only track with literary developments throughout his career, but also with the social, ecological, and legal changes that took place in Lafayette county. This presentation

* Assistant Professor of Digital Humanities Ashoka University, New Delhi. Email: joostburgers@gmail.com

therefore resides at the unusual intersection of aesthetics and demographics.

Theory and Practice

The Digital Yoknapatwpha Project resides at the confluence of two different types of academic traditions: distance reading (Moretti) and thick mapping (Presner, Shepard and Kawano). From the former it builds on the idea of mapping the events in an entire corpus and from the latter it utilizes rich but fuzzy humanities data sources to layer a map. Yet, as combination of the two it also somewhat unique: it offers the breadth of the Moretti model and the depth of the Presner model. While there has been other work on the spatial analysis of literature (Piatti et al.; Westphal and Tally), much of the digital work tends to focus on biographical data rather than corpus analysis (Evans et al.; Jewell), or do literary-spatial analysis on a very limited scale (Goldsmith). Due to the size of the database the mapping visualization is not just a representation of the database, but also a means to navigate through it (Caquard, Vaughan and Cartwright 2).

Database Model

The core of the DY project is a relational database that structures Faulkner's fiction into three entities: locations, characters, and events. The standards were established through a working group meeting in 2013, but also relies on established knowledge within the field of Faulkner Studies (Railton). Part of the standard is to document every character who appears in Faulkner's fiction. While there are previous examples of guides and concordances that were quite comprehensive (Brooks; Brown; Connolly; Dasher; Fagnoli et al.; Kirk), the DY database includes approximately every character who appears. This includes unnamed characters who appear only once, characters who are mentioned in passing but never actually appear in the text, and characters created by other characters in the text. Yet, this entity is by no means complete or unproblematic. This is because along with identifying all of these characters, editors also enter as much demographic information as possible about the character, including but not limited to: race, class, gender, other stories the character appears in, aliases, birth date, etc. Needless to say, entering this data is not always simple, especially for peripheral characters, because more often than not, Faulkner, like any other author, does not explicitly indicate someone's class or race, and these usually have to be inferred from the text. These reservations aside, character data table with over 3,300 records and 25 different attributes, allows for a birds-eye demographic view of Faulkner's fictional county.

Data Visualization

Using QGIS it is possible to map out the character demographics using the X/Y data provided by the location table. Since this data is not georeferenced to any real-world map, it is essentially a scatter chart projected onto a digital version of Faulkner's

own maps. Through this plot, it is possible to know who a character is, but also, with varying degrees of certainty, where that character is on Faulkner's map. Beyond locations, editors also enter events. These are manually entered narrative units in which there is a unified space, time, and action. Characters, therefore, can be tracked across locations and across events.

Using this data, I have created a new data model that individuates each character in each event on the display and turns him or her – in Faulkner's world these are the only two options – into a point. The resulting visualization in QGIS gives insight into the distribution of characters across Faulkner's fictions (see Figure 1).

Character Distribution Weighted by Location Certainty

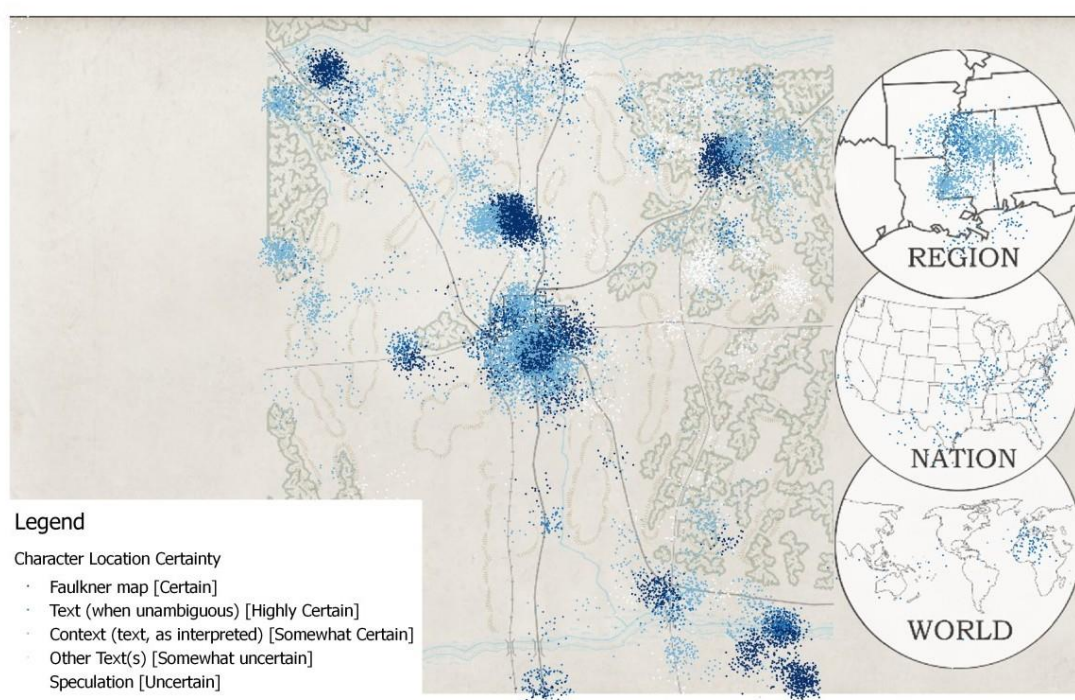


Figure 1: Character Distribution Faulkner's Fiction by Location Certainty

The above figure gives an impression of the distribution of all of the characters in all of the events in all of the texts currently in the database. As to be expected, much of the action and many of the characters appear square in the middle of the map in the center of town, the major hub of activity. Interestingly, when Faulkner actually imagines the US as a space he can only conceive of action happening in few places: the Southwest and West, the eastern seaboard, and most of all Appalachia. This is likely due to his familiarity with these spaces during his travels.

Some More Granular Examples

Beyond simple population distribution, it is also possible to pull forward certain demographic features. On that same map we can trace how Faulkner sees race playing out in his particular county (See figure 2).

Character Distribution by Race

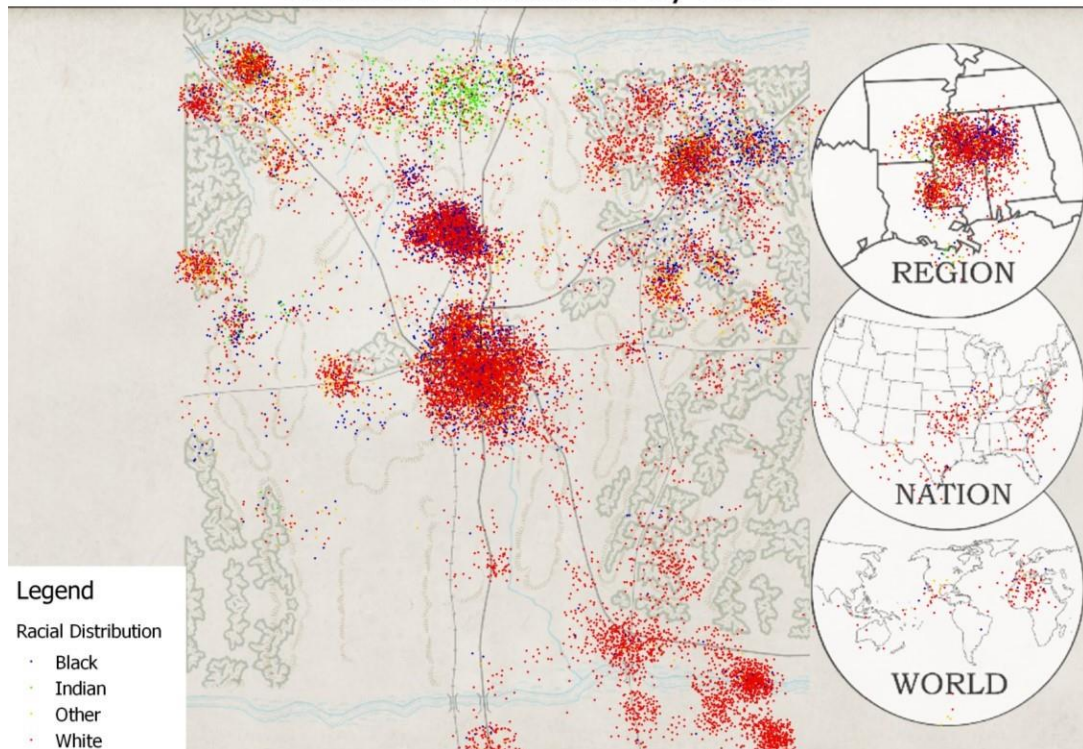


Figure 2: Racial Distribution Characters in Faulkner's Fiction

In the figure above a very clear racial picture of Faulkner's county emerges. First, Native Americans, indicated in green appear almost exclusively in the north of the county. This makes sense, since most of Faulkner's native American stories take place before Jefferson is ever settled. Interestingly, African-Americans can be found in two distinct locations: Jefferson proper and the big former plantation houses directly north, north-east and southeast of Jefferson. This is altogether not surprising to anyone familiar with Faulkner's fiction. Yet, what is remarkable is just how stark this contrast is. On only a few occasions in Faulkner's entire fiction do African-Americans have any type of agency outside of their relationships with the Anglo population. Their sense of space and range of movement is highly constricted by the racial laws and mores of the pre-Civil Rights era south. What is more, they are an obvious minority in Faulkner's fiction. Statistically, speaking they only appear in about 13% percent of his fiction. This despite the fact that Faulkner himself contended that African-Americans outnumbered whites two-to-one in Yoknapatawpha. There is therefore a stark contrast between Faulkner's imagined demographics and how those demographics play out across his fiction. Further afield, it is interesting to note that all "Other" racial minorities, those of mixed heritage or indeterminable, cluster in the center of town and the north-East, while being virtually invisible in the south, the site of Faulkner's working-class Snopes trilogy. Finally, though a somewhat obvious point, it is interesting to see just how white Faulkner's fiction is. There are virtually no

spaces on the map which in a certain sense are also not white spaces. Social relations in Faulkner's county are almost entirely determined through the presence or absence of white people. Drilling down into the data further we can also reveal that Faulkner's fiction does not just cut across racial lines, it cuts across gender lines as well (See Figure 3).

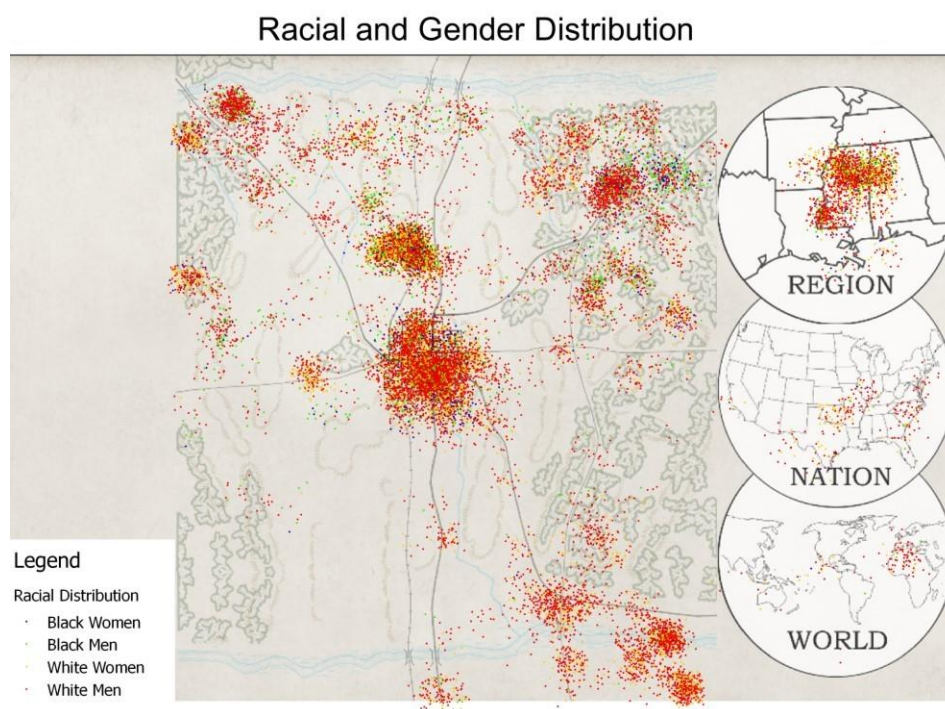


Figure 3: Race and Gender Distribution of Characters in Faulkner's Fiction

In the figure above the two main racial categories used for describing characters in Yoknapatawpha: African-American and white have been broken down by their race. What becomes clear is that once again the greatest space for gender and racial intersection is the center of town. This is, on the whole, to be expected. Yet, in like fashion the race and gender breakdown of Faulkner's fiction is quite stark. There are spaces for white men and white women, and there are spaces for white men and black men, and there are spaces for black men and black women, but there are virtually no spaces for white women and black men, or white men and black women. Black women can only be around white men if there is an implied presence of a black man in the background. This is a quite stunning revelation. After all, it speaks about Faulkner's anxiety about miscegenation. The races and the genders have to be paired off in a very particular fashion across the texts, and only certain combinations are possible.

Towards a Conclusion and Further Research

Currently, I am working on creating a dynamic visualization that creates the above plot for each individual text across Faulkner's career in chronological order. What I suspect it will reveal is that social spaces become more mixed across race and

gender lines as Faulkner's career progresses. The change would reveal Faulkner as an author who was both haunted by his region's racial past, but moved towards changing the representation of it through his fiction.

List of Works Cited

"Dislocating *Ulysses*: Re-Situating *Ulysses* and the Exhibit in Time and Place". 2017.
<<https://dislocatingulysses.wordpress.com/>>.

Brooks, Cleanth. *William Faulkner; the Yoknapatawpha Country*. New Haven,: Yale University Press, 1963. Print.

Brown, Calvin S. A Glossary of Faulkner's South. New Haven: Yale University Press, 1976. Print.

Caquard, Sébastien, Laurene Vaughan, and William Cartwright. "Mapping from above/Mapping from the Ground: Mapping Environmental Issues and the City." *Mapping Environmental Issues in the City: Arts and Cartography Cross Perspectives*. Eds. Caquard, Sébastien, Laurene Vaughan and William Cartwright. Berlin: Springer, 2011. 1-8. Print.

Connolly, Thomas Edmund. *Faulkner's World : A Directory of His People and Synopses of Actions in His Published Works*. Lanham: University Press of America, 1988. Print.

Dasher, Thomas E. *William Faulkner's Characters : An Index to the Published and Unpublished Fiction*. Garland Reference Library of the Humanities. New York: Garland, 1981. Print.

Evans, Martin, et al. "Authorial London: The City in the Lives and Works of Its Writers". Stanford University, 2017. Ed. Evans, Martin.
<<https://authorial.stanford.edu/>>.

Fargnoli, A. Nicholas, et al. *Critical Companion to William Faulkner : A Literary Reference to His Life and Work*. Facts on File Library of American Literature. New York: Facts On File, 2008. Print.

Goldsmith, Meredith. "Mapping Literary Visions". 2017.
<<http://www.meredithgoldsmith.com>>.

Jewell, Andrew. "Mapping a Writer's World - a Geographic Chronology of Willa Cather's Life". U of Nebraska-Lincoln, 2004-2017. Willa Cather Archive.
<<http://cather.unl.edu/geochron/>>.

Kirk, Robert Warner. *Faulkner's People, a Complete Guide and Index to Characters in the Fiction of William Faulkner*. Berkeley,: University of California Press, 1963. Print.

Moretti, Franco. *Graphs, Maps, Trees : Abstract Models for a Literary History*. London ; New York: Verso, 2005. Print.

- Piatti, Barbara , et al. "Mapping Literature: Towards a Geography of Fiction." *Cartography and Art*. Eds. Cartwright, William, Georg Gartner and Antje Lehn. Berlin: Springer, 2009. 177-192. Print.
- Presner, Todd Samuel, David Shepard, and Yoh Kawano. *Hypercities : Thick Mapping in the Digital Humanities*. Metalabprojects. Cambridge, Massachusetts: Harvard University Press, 2014. Print.
- Railton, Stephen. "Bibliography". University of Virginia, 2017. *Digital Yoknapatawpha*. <<http://faulkner.drupal.shanti.virginia.edu/content/digital-yoknapatawpha-bibliography>>.
- Westphal, Bertrand, and Robert T. Tally. *Geocriticism : Real and Fictional Spaces*. New York: Palgrave Macmillan, 2011. Print.

Collocation Analysis in News Corpora and its Implication to News Values

Huei-Ling Lai*, Chao-Lin Liu**

Abstract

News values determine what makes certain stories worthy of being news. From the discursive approach, news values are taken as being embedded in language (Cotter 2010) and are constructed through news discourse (Bednarek and Caple 2012, 2014). Corpus linguistics techniques are employed to analyze news texts and identify linguistic resources to represent each value, including negativity, proximity, prominence, consonance, and others. However, while the denotative meanings of linguistic devices can uncover textual values, newsworthiness can be underestimated or misinterpreted if linguistic resources are taken at its face value, as linguistically encoded meanings are often underdetermined. In addition to linguistic strategies, social-cultural contexts that provide communal common ground for interpretation are crucial for reinforcing a better understanding of newsworthiness (cf. Potts, et al. 2015). Hence, this study aims to carry out an in-depth examination of newsworthiness by investigating denotative, implicative and social-cultural meanings.

Among the linguistic devices for news values analysis, collocation analysis is considered useful as recurring patterns can identify specific indicators established around certain entities (cf. Baker et al. 2013; Potts et al. 2015). However, for news texts, significant collocates may not occur in the immediate co-text of the key word. A more rigorous computational procedure is needed to analyze collocation strength in a news text with collocates found beyond the immediate co-text. To this end, an ethnic term 客家 kèjiā ‘Hakka’ with underspecified referential meaning is measured regarding the collocation strength of kèjiā and its noun and verb collocates occurring nearby in news texts. The database is based on news corpora from four major newspapers in Taiwan—Knowledge Management Winner (KMW), Udndata (Udn), Liberty Times Net (LTN), and Apple Daily (AD)—during the years from 2005 to 2015. In total, 48726 new articles with 168116 tokens of kèjiā are found. Four dimensions are employed to measure collocation strength of the noun and verb

* National Chengchi University, Email: hllai.nccu@gmail.com

** National Chengchi University, Email: challinliu@gmail.com

collocates of kèjiā: (1) frequency, (2) mean and variance of the distance between kèjiā and its collocates, (3) Pearson’s chi-squared test and (4) pointwise mutual information (PMI). Moreover, verb collocates are examined regarding the news topics. The results of the analysis are given in Table 1 for noun collocates, Table 2 for verb collocates, and Figure 1 for the news topics, respectively.

The results show that some collocates have high frequency and strong semantic relations; some show long distance but strong semantic connections; others show high frequency yet long distance and weak semantic connections; still others show stronger semantic connections but low frequency and long distance. Such a precise computation gives rise to a finer-grained distinction of collocation strength regarding various collocates, enhancing the referential specificity of the term kèjiā. For instance, among noun collocates (See Table 1), xiāngqīn ‘people’, wénhuà ‘culture’, chuántǒng ‘tradition’, měishí ‘cuisine’, and tóngguā ‘Tong blossoms’ score relatively higher than other cases in terms of the four calculations, indicating that they carry stronger collocation strength with kèjiā. The collocate yìshù ‘arts’ shows high frequency, but scores quite low in the other three indicators. This case and other similar examples show clearly that some collocates do not have strong connections with kèjiā in meaning or distance, even though they show high frequent occurrence. A peculiar case is yìngjǐng ‘stiff neck; toughness’. It tops the other collocates in terms of PMI value, indicating the strongest meaning relation with kèjiā. However, this word does not show high frequency, nor does it show close distance with kèjiā. Furthermore, regarding verb collocates (See Table 2), jǔbàn ‘hold’, zhǎnxiàn ‘show’, chénglì ‘establish’, and tīyàn ‘experience’ score relatively higher than other cases in terms of the four calculations, indicating that they carry stronger collocation strength with kèjiā, and zhǎnshì ‘display’ tops the other collocates in terms of PMI value, indicating the strongest meaning relation with kèjiā. Moreover, some verbs show specific Hakka matters that are reported in the news. For instance, the verb guīhuà ‘to plan’ mostly is related to Hakka cultural parks, or Hakka earthen buildings, zhīchí ‘support’ is always related to presidential or other political candidates, and tuīguǎng ‘promote’ is usually related to Hakka culture, followed by Hakka language and music. These syntagmatic co-occurring collocates increase semantic specificity to the vague ethnic term kèjiā, hence giving more concrete specification of the semantic domain projected by the ethnic concept Hakka.

The results also reveal how news values are constructed in the news discourse. There are 168,116 tokens of kèjiā found in 48, 726 news articles in our database—meaning about 93 news articles per month related to kèjiā from four major newspapers in Taiwan in the recent eleven years. A strong correlation between kèjiā, its collocates and news topics is observed (See Figure 1). The most frequently reported topics are:

life and style (34%), art and culture (32%) and politics (11%). There are 14 verb collocates in the art and culture topic, and the verb *tīyàn* ‘experience’ exists in the art and culture topic most.

The results also show that, in addition to the linguistic context, the social-cultural context needs to be taken into consideration for a more in-depth examination of news values. Indicators of news values such as negativity, proximity, eliteness, impact, and consonance particularly are suggested to be re-examined with regard to the corresponding linguistic devices. First of all, the topic-associated collocates indicate that the most prevalently concerned Hakka news topics are straight news. Their attachment to Hakka nostalgic cultural traditions are foregrounded as being newsworthy. Such an emphasis perpetuates a stereotypical impression on what are significant for the Hakka ethnic groups. This shows the media’s efforts in raising the visibility of the Hakka ethnic groups by recognizing their deep anxiety in preserving their inheritance and assets, yet leading to the Hakka ethnic groups being the peripheral as opposed to the social mainstream. Specifically, when the top verb collocates are analyzed, they seem to carry positive meaning connotations. Nevertheless, within the news topics, they all refer to Hakka matters that need to be promoted or enhanced, implying their secondary lower status. The fact that the verb collocate *zhīchí* ‘support’ is always related to political candidates indicates that the Hakka ethnic groups are often politically manipulated as supporting certain political parties. The strongest semantic connection of a peculiar symbolic code *yìngjǐng* ‘stiff neck; toughness’ as shown by its highest PMI value reveals an ideological association that plays on ethnic stereotype—Hakka people are portrayed as being tough and indomitable by the news media.

In sum, this study investigates the use of an ethnic term in news discourse from linguistic, discursive, and social-cultural aspects. A more rigorous computational procedure than hitherto used is employed to precisely measure the collocational strength of noun and verb collocates of a whole news article in news corpora. The results indicate diversified distributions of the collocates with regard to their frequency, distance, and semantic connections. The findings enhance the specificity of the meaning of the ethnic term by revealing in more detail how this ethnic group is characterized, the trends in the choice of topics in which the term most commonly occurs, and the ideological representation of this ethnic group in a wider social-cultural context. The findings imply that an understanding of newsworthiness is deepened as the representations of the minority ethnicity in the news media are analyzed through three layers—the denotative in the linguistic, the referential in the discursive, and the implicit in the social-cultural context.

1900-1924 年西方「風景畫」的傳入與影響

彭卿*

摘 要

本文研究的是西方“風景畫”觀念的傳入與其對中國“山水畫”的影響。使用“全國報刊索引”“瀚堂近代報刊”兩個數據庫，以“風景畫”為關鍵詞切入考察，運用詞頻的數位人文研究法，探討在現代轉型背景下，從中西二分二元論（泛指 1900-1915 年）到五四新文化運動，中國人所接受的西方“風景畫”觀念變遷，及其對“山水畫”的影響。是一篇方法論介紹與應用研究相結合的文章。

關鍵字：風景畫、攝影、畫報、自然、印象

* 浙江師範大學，Email: 522131648@qq.com

圖錄與畫題的數位人文分析：佛利爾先生的中國畫想像

徐鶯*、邱偉雲**

摘要

晚清民國時期，在全球化的影響下“展覽”的觀念在中國形成。受西方賽珍會、博覽會的影響，中國收藏家手中秘不示人的書畫，從最初的私家雅集走向以展覽會的形式公開展覽；同一時期珂羅版的印刷技術成熟，私人收藏又以圖錄的形式獲得展示的公共性。在中西文化交流過程中，大量的中國畫被西方收藏家收藏，並因早期西方藏家對中國藝術的認知，構建了西方式的中國畫審美標準，這些藏品也成為後世西方學者研究中國書畫史的最初範本和材料。

本研究聚焦美國佛利爾美術館的創始人、美國最大的中國畫收藏家查爾斯·朗·佛利爾（Charles Lang Freer, 1856-1919）與中國近代最大的書畫收藏家虛齋（龐元濟，字萊臣，號虛齋，1864-1949）之間的收藏傳奇。以 1915 年虛齋挑選藏品參加美國巴拿馬太平洋博覽會的參展圖錄《中華歷代名畫記》（英文書名：Biographies of Famous Chinese Paintings from the Private Collections of Mr. LC Pang, 1915）和 1916 年虛齋為佛利爾購藏中國畫定制的圖錄《唐五代宋元名畫》（英文書名：Antique Famous Chinese Paintings collected by P'ang Lai Ch'en, 1916）為研究對象，並根據佛利爾美術館 1915 年和 1916 年的購藏檔案、佛利爾與虛齋的書信往來為研究內容。

本研究共涉及 190 件作品，169 件標註為高古時期作品。由於五代兩宋大部分的古代書畫作品，極少有作者親筆命定的“畫名”，大部分連作者名字都不落款。而大部分所謂傳世之作的“標題”和“題名”多為鑑藏者根據作品的內容、題材、情節意會的概括，並記錄於書畫著錄之中，這些題名體現了鑑藏者個人認知與品味，也體現了時代的審美標準。《林泉高致》中對畫題的定義有兩個方向：一是畫的“題名”，即畫的名字；二是畫的“主旨”，即畫的內容、喻意和情節的總括。因此對於收藏家鑑藏品味的研究，從藏品的畫題入手研究是一個極佳的切入點。畫題寥寥幾字，概括了畫作的主體、內容、情節等。通常將畫題詞句進行拆分，便可以得到主題和題材。本研究從圖錄“畫題”入手，用“主題聚類”技術多維度進行“畫題名類統計”研究，通過“主題詞”提取、聚類和語義分析，用數據來呈現收藏品味。畫名單字頻次研究結果：1915 年虛齋在參展巴拿馬太平洋博覽會的圖錄中展示給西方的中國畫主題是“山、雪、林、秋、風、

* 杭州師範大學，Email: artxuying@163.com

** 山東大學歷史文化學院歷史學系，Email: brianacwu@163.com

江”，英文翻譯所呈現的情況則依次是“Scene, landscape, mountain, snow, birds, autumn, horse, villa, tree”；這符合虛齋收藏的體系，山水是其收藏的主體。佛利爾從兩部圖錄中收藏的 34 件作品的集中的主題則是“雪、寒”，英文畫名呈現的則是“mountain, snow, winter, scene, horse, elephant, birds, music, fairy, tree, plant, girl”。此外研究中還分析了作品的年代、畫家、材質、捲軸冊的類型、賦彩與水墨形式的比例。

通過對圖錄藏品數據的系統分析呈現：以佛利爾為代表的西方收藏家的中國畫審美趣味；以虛齋為代表的近代紳士群體，在民族文化主義的背景下，在全球化的語境中，展示給西方的中國畫的概貌。通過中西方收藏史中兩個典型人物的收藏互動案例研究，揭示在中國傳統社會發生轉型的時期，書畫是如何被紳士階層作為承載民族認同的文化名片，被西方所欣賞和接納，並在之後的中國畫收藏中定格了“經典”的審美範式。

現代白話文「基因組」初探：
《紅樓夢》、《安徽俗話報》和《新青年》
語言風格統計研究

鄭文惠*、余清祥**、何立行***

摘 要

過去文學史習慣以五四運動為分界，將漢語書面語由文言而白話的轉變視為一場「文學革命」的產物，關鍵則是以五四啟蒙報刊——《新青年》雜誌為主要展演場域。然而，近年來晚清研究的興盛使學界開始注意到有必要重新檢討此一論斷。根據學者蒐羅整理，五四運動前二十年刊行的白話報刊約有三、四百種，另有不標示為白話報刊但文字淺易的刊物三百種以上，以及文言報刊所附白話欄目附張五十餘種。這些刊物遍布於中國各地，其語言風格是否直接影響白話文的興起，是值得探究的問題。

五四健將對晚清白話書寫現象並非未曾察覺，而是刻意做出區隔。周作人將晚清白話文與五四白話文的差異歸結為：晚清人是將古文「翻作白話寫出來」，五四人則是「話怎樣說便怎樣寫」；以及「古文是『老爺』用的，白話是『聽差』用的」兩點。胡適則認為晚清文人最多「有意主張白話」而非「有意主張白話文學」，在文學承傳上跳過晚清上接《紅樓夢》、《水滸傳》¹。此一提法當然有其歷史背景和論述策略，但實際上五四白話的血緣是更近於晚清白話或古典白話呢？除了質化的分析解讀之外，或許我們也可以採用數位人文研究方法一探究竟。本研究團隊曾以統計方法探究《紅樓夢》的寫作風格，又採用監督學習（supervised learning）、非監督學習（un-supervised learning）的統計方法，並借用生物多樣性的概念，探勘《新青年》文言、白話兩者內部生態系的差異，尋找具有成為客觀區辨指標潛力的語言表徵，以建立判讀文、白語言差異的指標，藉以考察《新青年》初期與晚期文體的變化。此次進一步借用生物學「基因組」的概念，以統計方法探索現代白話文在語言風格上的承傳和變異，並分析《新青年》與古典、晚清白話文間的關聯。

本研究以古典白話、晚清白話和五四白話中三份代表性文本作為研究素材：《紅樓夢》、《安徽俗話報》（1904-1905）、《新青年》（1915-1926）。其中《安徽俗

* 政治大學中國文系教授，Email: whcheng123@gmail.com

** 政治大學統計系教授，Email: csyue@nccu.edu.tw

*** 清華大學中文系專案經歷/兼任助理教授，Email: lillianlho@gmail.com

¹ 詳見胡全章《清末民初白話報刊研究》，北京：中國社會科學出版社，2011。

話報》在五四運動前所刊行的三、四百種白話報刊中深具典型性與代表性意義：《安徽俗話報》是同為《新青年》主編陳獨秀於 1904 年 3 月 31 日在蕪湖為開民智而創辦的，係安徽最早創刊的白話文報紙，因其語言通俗、內容豐富，傳播範圍廣、社會影響大，成為清末白話報發展高峰期的代表。《安徽俗話報》辦報的主要任務是為了把各處的事體、各項淺近的學問，用通行的俗話演出來，以傳播知識，啟蒙大眾，號召大眾參與社會變革。因此，除了使用白話文以作為開化大眾的腳本，以擴大傳播範圍外，也實質考量訊息的多層級傳播方式，以強化社會影響力，充分展現了大眾化報刊的社會屬性與文化功能。再者，由於《安徽俗話報》的主編陳獨秀及同人與安徽地緣的文化助力，而得與《新青年》聲氣相通，均使以《新青年》為核心的新文化運動及白話文運動與《安徽俗話報》有著內在深層的關聯。本研究借用生物學「基因組」的概念，當能有效分析《新青年》與古典白話及晚清白話文《安徽俗話報》之間的關聯。

在生物學上，基因組的定義是「包含在該生物的 DNA 中的全部遺傳訊息」，但實際上遺傳訊息無法由單一基因組提供，必須透過多個個體的協調與組織。由於每一個體皆不相同，生物家族間 DNA 的承傳變異關係會隨著比對資料的擴大而增修，當比對的個體到達一定數量時，科學家可以大致破譯一個物種的遺傳訊息，勾勒基本的基因圖譜；但若持續進行研究，則能持續發現更多基因變異。此一概念運用在語言風格的數位研究上意味著：本研究成果能對三個文本間的相似和差異有較清楚的認識，並作為將來進一步深究三種白話文承傳變異關係的基礎。

關鍵字：白話文、語言風格、基因組、新青年、安徽白話報、紅樓夢

《全唐詩》音樂詩的數位人文研究：以弦樂器為主

鄭文惠*、顏靜馨**、梁穎誼***

摘要

從先王雅樂之始，中國即有八音（金、石、土、革、絲、木、匏、竹）之屬的樂器。《詩經》時期，樂、舞與文學已然高度結合；歷經秦漢，迨至隋代，中原文化與異文化幾度遇合而開啟文化混融的契機。依據研究，普遍認為漢樂與胡樂的融合，至宋代大致完備。從樂器史角度而言，宋·陳暘《樂書》已將樂器依使用場域、起源、演化等標準進行歸類。若深入檢視《全唐詩》所收郊廟詩、樂府詩後，可發現遲至唐代中晚期可能已奠定後世樂器的類型、形制、演奏手法等，甚至殿堂禮樂在「樂—詩」的結合上，除承襲前代特色外，更融入異族色彩。唐代設梨園、教坊，其所創作之燕樂、法曲，將傳統音樂文化帶入另一個革新的輝煌時期，無論世俗常民或文人社群，音樂也成了生活中的一部分。在廟堂與民間，無論祭祀、朝會、宴飲，乃或平居、農耕、織作，音樂於唐代如同茶、酒一般和生活相接相融。對文士而言，身在朝廷或居處江湖，音樂在生命中不可或缺；對民間來說，作為謀生的技藝，為了滿足消費對象，習藝者必得提升技藝、改良器樂，進而也促進了樂器與音樂的變化發展。證諸《全唐詩》四萬餘首詩¹，包含樂器詞的詩歌近二萬首，適足以說明唐代音樂文化的興盛與流行。

唐詩作為抒情自我的再現，抒陳著詩人的思感與聞見，縮合著現實經驗與理想投射；唐詩傳唱與音樂結合的型態係後世研究當時樂器發展、音樂文化的重要材料。是以本研究由《全唐詩》所見樂器入手，選定弦(絃)、管、打擊類樂器常用樂器名詞與其它如卷葉、嘯、麗譙、三楚等為關鍵詞，藉由 R 統計軟體進行第一階段文本篩選，再聚焦於弦樂器。主要在於漢字一字多義的特性，諸如打擊樂器中的金、石、節，管樂器中的管、竹、和等關鍵詞須大量人工介入方能進行判斷。然而，弦樂器中除了做為總名的絲、弦須較多的人工處理外，其它相關名詞定義清楚，對於數位技術的嘗試具有優勢。因而，本研究先聚焦於弦樂器的數位人文研究。

弦樂器在傳統文化中歷史悠遠，樂曲的傳嬗在時間洪流中也不斷被淘洗，往往詩人或借況己境，多所寓興，或挪作國家治亂的象徵，即使樂曲名稱也多成為

* 政治大學中國文系教授，Email: whcheng123@gmail.com

** 台中科技大學通識中心兼任講師、中正大學中國文學系博士生，Email: xing3325@gmail.com

*** 政治大學中國文系博士後研究員，Email: htiosc85@gmail.com

¹ 《全唐詩》中「組詩」多作一首計，如王建〈宮詞〉一百首，視同一組；但如樂府詩〈昔昔鹽〉、〈胡笳十八拍〉，個別詩人如李白〈古風〉，劉長卿〈竹枝詞〉等卻拆成一首一首，因此，《全唐詩》數量四萬餘首，目前並非是有效的計數。

文化隱喻的載體。從象徵文人雅士的琴、瑟到坊間充滿異族風情的琵琶、胡琴，無不承載著遠古以來天人交流、先王治道、聖賢情志、貞女行操等道德表徵或文化記憶；且相較於其它類樂器，文化融合的軌跡鮮明。因而，如琵琶作為唐代流行的樂器，除是流行時尚的示現外，隨著安史之亂，也多作為胡漢頹頹或家國離亂的象徵，大小胡笳曲也雜糅了詩人對外族政治力滲入的不安。樂器、樂曲及音樂文化實質上與抒情小我及家國大我緊密相關，甚至涉及詩詞文體的衍異、城市文化的發展，及文化抵抗、權力斡旋等跨文化的交流與協商。

緣此，本文沿續前年結合數位技術與分子鏈研究方法，持續拓進主題與意象叢的研究，聚焦於《全唐詩》中的弦樂器及其共現的詞叢，考察如「琴」，經常和酒、書組詞：琴尊、琴樽、琴觴、琴壺、琴杯、琴爵、琴酌、琴酒，而這些組詞大多出現在閒居、貶謫或退隱生活情境中；鶴琴共同組詞，除指涉古琴曲名外，通常和隱逸或神仙傳說共現，道士的比例高於僧眾，神仙傳說則西王母、琴高、宓妃出現比例最高……等等。透過關鍵詞及其共現詞彙的分析，主要在於凸顯主題、情境與意象叢的結構，一方面勾勒樂器、樂曲與使用情境的關係；一方面深層考掘樂器出現的具體情境和唐代音樂文化、城市文化、工藝技術、胡漢關係、儒釋道思想觀念……等等的連結，並觀察在郊廟祭典儀式、樂府詩，或文士的集會詩歌中，樂器因應不同場合的配置、用途的差異，以凸出樂器配置的獨特性與功能屬性。此外，也凸出文人雅集宴樂或園林隱逸空間中的音樂生活及其深層的文化底蘊，包括琴、瑟、琵琶……等如何承載文化記憶且具體展演於生活中而具有一定的表徵意義與作用，尤其詩人在耳目之娛外如何賦予更多形而上的思考與作用，進而提昇弦樂器或音樂文化的精神境界；甚至樂器裝飾所呈現樂器等級及其所表徵的身分階層，適足以印證唐代與波斯等國的文化交流事實……等等。為使研究更立體更具說服力，本研究也援引隋唐禮樂志、會典，及西域交通史、中國樂器等相關資料作為輔證。

關鍵字：《全唐詩》、音樂詩、弦樂器、數位人文、文化交流、琵琶

論壇：中國古代詩歌格律發展史的數位研究

本組 panel 以台灣大學中文系蔡瑜教授為召集人，蔡教授主持科技部數位人文專題研究計畫「漢詩音韻分析系統」，開發了「漢詩格律分析系統」，其研究團隊於本 panel 分別發表〈漢詩格律分析系統的開發：以格律研究為導向的數位工具案例〉、〈中國古代詩歌古、新體格律發展之「自然合格率」理論初探〉、〈數位方法模塑出的漢詩格律發展史觀——以初唐詩學為例〉等三篇論文，以呈現階段性的開發與研究成果，並透過系統開發過程中，關於詩歌文獻和語音聲韻兩種學術型資料庫的整合、資料結構的調整、自然語言分析、演算法與效能測試、使用者介面設計等等，與中國文學研究者之間所產生的互動與辯證關係，從而提出中國古代詩歌在中古時期新體詩格律發展之「自然合格率」理論，嘗試突破傳統格律與詩歌研究的瓶頸。

另一方面，像這樣針對中國古典詩歌格律的研究需求，開發專業分析功能的數位工具或系統，早期有中央研究院文哲所林玫儀教授的「詞學韻律數據庫」和北京大學中文系李鐸博士的「《全唐詩》分析系統」與「《全宋詩》分析系統」。這些系統的設計皆是將數位文本與詩歌體式分析規則相結合，進一步分析和統計唐詩、宋詩、唐宋詞的詩體、句式和押韻方式，通過大數據所呈現出來的趨勢，探索中國古典詩歌格律發展的歷程。本組 panel 的召集人所主持的「漢詩音韻分析系統」研究計劃，在開發初期亦多方借鑒前賢的經驗。

職是之故，本組 panel 邀請了北京大學中文系杜曉勤教授發表〈使用計算機工具軟件輔助研究中國古代詩歌格律演變史的案例和心得〉，杜教授自 1995 年的博士論文開始，便關注中國古典詩歌格律的發展，2004 年至 2011 年歷經 8 年所打造「中國古典詩文聲律分析系統」，可以說是中國古典詩歌格律發展史領域當中以數位人文作為研究方法的前驅。該系統的設計因學術機緣與研究者的興趣有別，與召集人蔡教授所開發的「漢詩格律分析系統」，無論在資料庫的運用、資料結構、使用者介面，乃至於實務應用與理論開展，可資交流與對話，共同探索中國古典詩歌格律發展史之數位研究，其中可能面對的議題、方法與未來發展。

關鍵字：漢詩、格律、自然合格率、漢詩格律分析系統、中國古典詩文聲律分析系統

使用計算機工具軟件輔助研究

中國古代詩歌格律演變史的案例和心得

杜曉勤*

摘 要

「中國古典詩文聲律分析系統」為 2004 年中國國家社會科學基金項目「中國古代詩歌聲律資料庫的建立及中近古時期詩歌律化過程研究」的核心研究成果，該系統的研製主要分為三個步驟：（一）以李珍華、周長楫編撰的《漢字古今音表》（中華書局 1999 年）為依據，建立「中國中上古漢字音韻資料庫」。（二）建立「中國古典詩歌文本資料庫」，收錄《詩經》、《楚辭》、逯欽立輯校《先秦漢魏 晉南北朝詩》、徐陵編《玉台新詠》、清彭定求等編《全唐詩》、宋人郭茂倩編《樂府詩集》諸多中上古詩歌文獻，共 17 種 1182 卷 900 多萬字。（三）研製「中國 古典詩歌聲律分析系統」軟件，用 Delphi 語言針對「中國古典詩歌文本資料庫」中的詩歌作品快速標注四聲和平仄，自動標識句式、聯式、篇式，並統計句、聯、篇、韻的聲律格式。運用這套系統，爾後陸續針對特定的中國古代詩歌體式進行分析與統計，發表了〈沈約所賞魏晉五言詩的詩律標準〉、〈吳聲西曲與永明詩律 形成之關係〉、〈盛唐“齊梁體”詩及相關問題〉等多項專題研究，以批量的四聲 自動標注和八病標識、統計技術，配合精心校勘的詩歌文本、音韻系統多音字的 辨識、傳統文獻考證和文學史知識，來突破過去的研究瓶頸，對中國古代詩歌格 律演變過程提出客觀而有效的觀察與解釋。

* 北京大學中文系古代文學教授，Email: dxq@pku.edu.cn

漢詩格律分析系統的開發：

以格律研究為導向的數位工具案例

鄧賢瑛*、王昱鈞**

摘要

「漢詩格律分析系統」是因應漢詩詩歌研究者的需求，而開發出來的一個線上系統。本文主旨在提供「漢詩格律分析系統」至目前累積的開發經驗、所遇見的問題以及尋找解決路徑的過程，闡述如何採用現有的技術與資料的基礎之上，開發出合於人文學者所需要的研究工具。

傳統的漢詩格律研究重視文獻的梳理，並且需要透過不同的格律模型，對大量的詩歌文本進行探討作為佐證。但常因為資料過多、過於複雜的格律模型難以人力駕馭，而造成龐大的人力與時間負荷。「漢詩格律分析系統」的開發目的，即在於通過電腦提供的龐大計算能力，期能減輕研究者的負擔，呈現漢詩格律發展歷程中精微的變化，以利研究者發現問題。

由於電腦並不理解何謂「漢詩」、「格律」這些概念，因此「漢詩格律分析系統」在開發初期的重心即置於：

- 一、如何教電腦讀入詩歌文本
- 二、如何教電腦何謂詩歌格律
- 三、如何改善分析效率
- 四、如何提供使用者數據

為了解決前述之問題，本系統首要在於建立一個合用的詩歌模型，此模型應能正確的解析使用者輸入的字串，而得出一首詩歌應具有各種屬性，諸如：每句長度、每個字的音韻資料、句數、聯數等等各種基本資訊，並且能夠針對漢字一字多音、多形的特性，提出相對應的背景資料。

其次，為了能讓研究者對詩歌進行各種角度的觀察，必需建立適用的格律模型，此模型擔負的工作主要有二項：一是在一首詩歌內文所有音韻組合的可能性當中，挑選出符合最多格律的音韻組合，二為在決定好音韻組合之後，提供使用者分析結果。

在建立好詩歌與各式的格律模型之後，即發現若為提供線上即時運算服務，詩歌文本的長度為攸關效率的關鍵。因此，我們進一步改善了原有分析音韻組合

* 臺灣大學中國文學系計畫專任助理，Email: hyteng@ntu.edu.tw

** 法鼓文理學院佛教學系助理教授，Email: albyu35@gmail.com

時所採用的樹狀結構，對詩歌文本進行分段處理，由原本僅能分析五言絕句，達到分析五言排律的目標。

最後，本系統必需讓研究者能快速地儲存資訊，因此進一步改善使用者介面，提供匯出功能讓使用者能保留數據與分析結果；並且開發自動化程式，對多首詩歌進行批次的處理與分析結果匯出。

本系統開發至今，已針對研究者的需求，初步建一個用計算機處理詩歌格律的流程。除了完成格律分析系統的提供人文學者所需之研究工具外，於數位人文領域的研究上，我們亦初步嘗試導入機器學習模型，以詩歌格律學者們所整理之格律規範作為特徵，訓練出用以辨識格律系統之模型以預測詩歌之格律屬性，並分析其正確率，嘗試開發詩歌格律研究之新的可能性。

中國古代詩歌古、新體格律發展之

「自然合格率」理論初探

羅珮瑄*

摘要

「漢詩格律分析系統」依據現今流傳的古籍文獻中所記載自六朝永明時期至初唐的詩學，統理其中具有規範性質的格律內容，架接了「漢詩文獻資料庫」與「小學堂漢字古今音資料庫」之中古聲韻材料，建構出沈約四病、元兢四病、元兢調聲三術、劉滔二四字不同、五言略頌等諸項格律條件，該系統不僅可以擷取、排列語音資料，並且可隨使用者所設定的格律規範提供自動選音的建議，再進行各項格律分析與量化數據，提供研究者進行專題研究。

所謂格律分析與量化數據的具體做法，即列出該詩在各項格律條件下的合格與不合格次數，並計算其合格率、不合格率。由於每項格律條件的施作方法並不相同，有的是單句施作、有的是兩句一聯施作，最為複雜的可謂是格律發展漸趨成熟之後四句一循環的施作，是以計算合格、不合格次數與比例的時候，必須細緻考慮、設計公式。

另外，考慮到漢字特有的方塊字、一字一音、多字同音的特性，具有自然對仗的韻律，漢詩詩人創作時若是隨機選擇用字，有時候也會符合這些人為後設的格律條件規範，那麼如何透過統計數據來辨識、理解詩人的創作意識？如何解釋數據所呈現出來的各種消長變化？面對各項格律條件規範，合格率必須達到多少才能證明詩人創作的意圖？

職是之故，本系統設計「自然合格率」的計算公式，作為體現詩人創作意圖的標準之一。所謂「自然合格率」意指詩人創作時隨機選擇用字的聲調，其結果能自然合乎某項格律規範的機率。每項格律規範因其規範的性質、聲調選擇的數目、施作的次數，而有不同的「自然合格率」。而每首詩於某項格律規範的合格次數，除以該項格律規範的施作總次數，則是該首詩於該項格律規範的「合格率」；

「不合格率」可類推，即每首詩於該項格律條件的不合格次數，除以該項格律條件的施作總次數。凡該格律條件的合格率高於自然合格率者，可視為作者有意識地符合該格律規範；低於自然合格率者，則視為作者缺乏該格律規範的意識。

本文一方面探討「漢詩格律分析系統」中建立格律分析與量化數據的流程，一方面提出自然合格率的理論，以數位人文和大數據的視野，為中國古代詩歌格律發展歷程中，古體與新體的變遷和創作意識，提出新的觀點。

* 政治大學中國文學系博士候選人，Email: winnie72@gmail.com

數位方法模塑出的漢詩格律發展史觀：以初唐詩學為例

蔡瑜*

摘要

在中國詩學史上初唐是格律理論最蓬勃興盛也極富原創性質的階段，它承繼在齊永明的語言自覺與詩體革新之後，持續試鍊漢語漢字做為詩歌載體的多元途徑；它對「永明體」的斟酌與變創，更直接促成了律體詩的成熟。初唐近百年的格律發展呈現出一條持續變動推進的軸線，其中重要的理論建構與聲律實驗，同具時代與詩學的意義，皆是歷史生成不可或缺的部分。就其質性而言，如何勾勒其發展的圖像無疑是最貼近其歷史定位的研究進路，也是理解中國詩歌格律美學如何成立的最重要環節。數位方法在此一課題上的運用，提供學者多元的格律樣式與充分的數據，正有助於呈現複雜紛陳的現象，以探究此一時期典律競爭的歷程，展開理論分析→格律構擬→量化統計→理論與作品交叉分析的研究進程，而在此進程中理論建構與作品實踐之間具有相互解釋與彼此修正的關係。

格律的形成有賴於理論與作品實踐兩方面的同步到位，本文在格律理論梳理上以日僧空海《文鏡秘府論》為主要依據，再輔以中土及日本先後時期的文獻，進行格律條件的構擬。在作品實踐上則分從總集及別集兩個部分檢核，在總集上以現今僅存的初唐詩選集《翰林學士集》、《珠英學士集》作為格律檢核的對象，再輔以初唐四傑、文章四友、沈佺期、宋之問等初唐重要詩人的格律分析。其中崔融是《珠英學士集》的編集者，同時也是《唐朝新定詩格》的作者；李嶠則總領《三教珠英》的編修，同時又是深具規範作用的《雜詠百二十首》的作者，二人的著作尤其具有交叉比對的意義。透過考察這些密切相關的共時性的詩學事件，希望進一步探討兩個重要格律議題：一、初唐處於格律發展階段，其時古體與新體的界限該如何畫分？二、初唐所援用的新體格律究竟經過怎樣的變動發展的過程？

其具體步驟如下：

其一，依據文獻資料梳理出不同時期的聲律理論及相應的調聲規範，於數位系統中建立精確的檢核標準，作為相互比較的定點。具體而言，包括齊永明沈約所開創的「八病」；初唐元兢《詩髓腦》基於調整「八病」而提出修正的「元兢四病」，以及其所開創的「調聲三術」。在這些格律定點下，建立個別規範自身整體條件的檢核、不同規範間交錯配搭的檢核機制。

其二，導入自然合格率及合格率的計算機制，所謂「自然合格率」意指詩人

* 臺灣大學中國文學系教授，Email: yutsaitw@ntu.edu.tw

創作時隨機選擇用字的聲調（如以四聲為別選其一或以平仄為別選其一），其結果能自然合乎某項格律規範的機率。每項格律規範都因其規範的性質、聲調選擇的數目、施作的次數，而有不同的「自然合格率」。而每首詩於某項格律規範的合格次數，除以該項格律規範的施作總次數，則是該首詩於該項格律規範的「合格率」。經由「自然合格率」與「合格率」的比較，可以作為詩人是否遵行某項格律規範以及嫻熟程度的判準。

其三，以「拈二」的自然合格率計算，作為區分古體與新體的判準。由於初唐時期是唐人格律的摸索期，嚴格的古體新體之別未見明文確立，即或可能已具有相應的概念，也並不容易從作品的形制立判其屬，若非經過嚴謹的格律檢核，實難定論。本文認為要解決這個難題，宜建立一個相對客觀的標準，元兢「調聲三術」的「拈二」已具有「平仄二元化」及「黏對」的概念，是律體建立的里程碑。故在方法上應先進行整體五言詩的通檢，再以「拈二」的「自然合格率」做為區分此期古體與新體的標準。因自永明以來，第二字相對的概念已相沿成習，第二字相粘則是唐人嶄新的發明，是唐代新體意識的代表觀念，故以「拈二」的「自然合格率」為判準，未達自然合格率者屬古體，超過自然合格率者屬新體。本文透過建立客觀標準，解決格律發展初期古體新體難以分判的難題，以此篩選出適合的研究對象。

其四，經由詩作在各項格律條件的「自然合格率」與「合格率」的比較，進一步分析各項格律條件在初唐時期的施用，如何經歷過一段選擇、實驗、熟練、捨棄的過程。雖然每一個詩人的具體狀況可能存在差異，但已展現出一個基本的傾向。如規範句末字的「上尾」、「鶴膝」被延續下來，第二字「平頭病」的避忌則最先被更嚴謹的「拈二」徹底取代，「蜂腰」因規範第二字與第五字，也呈現逐漸被棄守的趨勢，「拈二」則成為最主要的調聲法則而日趨純熟。

本文在數位方法的輔佐下進行格律項目與作品實踐的交叉分析，盡可能避免「一以貫之」與「以後律前」的研究進路，努力發掘初唐格律發展過程的重要細節，在格律條件的取捨之間繼續探索格律美學的動態發展。

從檔案類電視節目探索

優化數位典藏用戶情感體驗的新路徑

王露露*、徐擁軍**

摘要

自 2016 年年底以來，黑龍江衛視和央視先後播出的《見字如面》和《朗讀者》作為一批滿足公眾文化訴求和懷舊情結的代表性節目，在大陸掀起了文化情感類綜藝節目的收視熱潮。據豆瓣網的數據顯示，《見字如面》和《朗讀者》的觀眾評分分別為 9.0 和 8.6，超過或者持平了同時期熱播的《極限挑戰》《火星情報局》等娛樂類綜藝節目。這兩檔電視節目都以名人書信或歷史文學作品等文化性較強的檔案元素作為節目的內容定位，通過構建一個基於共同情感結構、價值觀念和社會認同的浸入式人文記憶場景，使得原先主要通過個人查閱檔案來實現的私人式懷舊，轉變為可在公共情感空間內集體實現的文化熏陶和記憶緬懷儀式。檔案類電視節目是檔案信息資源的一種社會化、通俗化和娛樂化的利用方式，既反映出了當前社會各界對檔案記憶建構功能和情感價值的關注，又反映出檔案可以一定程度上滿足公眾強烈的人文情感訴求。筆者認為，研究此類電視節目，一方面，為挖掘數字檔案資源的情感價值開闢了一個新的研究視角；另一方面，為優化數位典藏的用戶情感體驗提供了一個新的實現路徑。

通過前期的文獻調研，發現媒體傳播和人文社會學科領域的幾位知名學者對檔案類電視節目有較為深入的研究，譬如，美國歷史記憶研究學者沃爾夫·坎斯特納（Wulf Kansteiner）認為：「所有的記憶媒介，尤其是電子媒介，既不是單純地反映集體記憶，也不是單純地決定集體記憶，但是它們卻不可擺脫地被捲入集體記憶的建構和演變之中，注意到這些是至關重要的。」大陸傳播學學者張紅軍和朱琳認為：「電視綜藝節目在建構集體記憶符號體系時通過組織‘性格化的觀眾’形成‘假想的我們體系’，在修飾實踐中遵循著‘觀眾不是簡單的信息接受者，而是節目的參與者和意義的共享者’之受眾理論，最終完成對其心理置換後的意義共享和共識達成。」在前期研究成果的基礎上，筆者認為，還可以圍繞人文教育意義、情感價值開發、情感體驗昇華、檔案真實性與情感性的把控、檔案部門與節目製作方的合作機制等新主題進行更為深入、切實的探索與研究。

本文主要採用文本分析和數據分析方法，以新浪微博、豆瓣網、嗶哩嗶哩彈幕網和 Acfun 彈幕網等大陸知名視頻評論或社交網站為主要調查平台，試圖通過

* 中國人民大學資訊資源管理學院，Email: wanglulu555@163.com

** 中國人民大學資訊資源管理學院，Email: 20070081@ruc.edu.cn

分析《見字如面》《朗讀者》《我在故宮修文物》《檔案》《故宮》等熱門檔案類電視節目的觀眾評論、評級評分等相關內容，總結近年來檔案類電視節目在檔案資源數字化服務中所扮演的重要角色和所發揮的獨特功能，以及此類節目在主要類型、內容定位和演繹方式等方面的特點和優勢，從而發現公眾對檔案價值關注焦點的轉移和檔案類電視節目對情感記憶的建構路徑，最終為檔案記憶情感價值的實現提供有益的參考。這種研究方法也具有較強的可行性和科學性，一方面，調研平台均為大陸主流社交網站或者視頻評論網站，觀眾活躍度高且評論內容不受約束，較為真實地反應了觀眾對電視節目的直觀感受；另一方面，基於前期初步調研發現，用戶評論內容具有較強的思想性、意見導向性和情感表達性。譬如，網友對《見字如面》進行了如下評價：「目前最愛黃永玉和曹禺部分，反復看了好幾次！信本身傳達的那種赤誠的人性和友誼，特別讓人感動」那句「醒來啊，麥克白，把沈睡趕走！」擊中了我，很具有一種與自我抗爭的振奮人心的氣質。」「《見字如面》閱讀的信件大多來自歷史，當代書信較少，但毫無疑問，還原度很高，網友們樂於看到這樣的節目。」網友對《朗讀者》進行了如下評價：「也許是因為每次朗讀都以真實可循的故事代入，每篇文章朗讀時都有作品當事人或者故事當事人在場，這種感染力在眾目睽睽之下逐現端倪」。

美國歷史電影學者羅伯特·伯戈因（Robert Burgoyne）曾說：「歷史電影幫助形構數百萬人的思想，在銀幕上看到的描述對公眾的歷史主體觀的影響經常比書本大得多。」以檔案類電視節目為代表的大眾傳媒的發展代表了大眾文化意識形態的演變，同時它們正在以無法想象的滲透力影響並塑造著國家想象和社會記憶。因此，有必要關注檔案類電視節目對社會記憶的建構功能及其衍生的情感價值，研究如何建立一個基於客觀歷史真實又不失人文情懷的社會記憶表徵體系，從而幫助公眾樹立起符合社會價值導向和滿足情感需求的科學歷史主體觀。電視節目，尤其是固定週期播送的文化類綜藝節目，在情感激發、知識普及和儀式感營造方面具有獨一無二的優勢，所以，此類節目在豐富和落實數位典藏的文化教育功能方面的作用也非常值得關注。

個人歷史文本事件分析工具及其於

明代倭寇研究之應用案例

白璧玲*、吳承翰**、蔡融易***、蔡宗翰****、范毅軍*****

摘要

將自然語言處理技術運用於數位化歷史文本之分析，已成為數位人文研究的一項重要課題。考量歷史事件研究之實際需求，若能將歷史文本之段落按照事件類型進行分類，對研究者之資料整理效率將有大幅的提升。本團隊去年研發出非監督式文本事件分群技術，並將之應用於《明實錄》之事件分類上，得到很好的效果，並刊登於 2016 年數位人文國際研討會(Pai, Lai et al. 2016)及 DH 2017 研討會(Tsai, Lai et al. 2017)，成功展示了文本事件分類技術的具體應用範例。由於此技術基於深度學習領域之 Paragraph2vec 算法(Dai, Olah et al. 2015)，無需依賴斷詞程式即可產生表徵段落語意之向量，因此大大提升本技術於歷史文本研究之應用面。

由於個別研究者對事件分類的應用有所差異，因此，事件分類的基準，如事件分類的精細度等等，必定隨之不同。在收集人文研究者的意見後，我們認為極有必要開發個人使用之文本事件分類工具，以供個別研究者針對自身資料整理與研究需求，調整事件分類之基準。此外，在非監督式文本事件分群技術中，調整事件分類基準的方式為設定分群數量；若群數少，則代表單一事件類型涵蓋大範圍，反之則代表單一事件類型涵蓋小範圍。在本事件分類研究工具中，研究者可自由調整此參數，觀察群數為何值時，整體的分群效果最接近研究需求。

考量上述研究者的運用需求外，在本團隊所開發網頁版《明實錄》事件分類工具的使用者反饋意見中指出，檢視各群的主要事件類型與分群品質，以決定是否採用此次分群結果之工作，相當耗時耗力。因此，在個人版事件分類工具中，我們亦加入一些解析文本所涉及時間與空間範疇為基礎的功能。主要將各文本段落以模版比對出時間資訊，並轉換為西元年，以及利用中央研究院建置之「中華文明之時空基礎架構」(CCTS, <http://ccts.sinica.edu.tw>) 平台及歷史地名資料庫，

* 中央研究院人文社會科學研究中心地理資訊科學研究專題中心博士後研究，Email: lingpai@gate.sinica.edu.tw

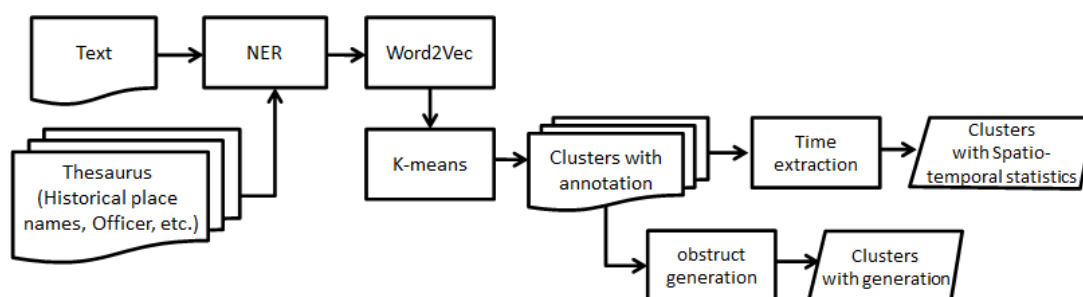
** 國立中央大學資訊工程系碩士研究生，Email: tfssai@gmail.com

*** 國立中央大學資訊工程系碩士研究生，Email: jungyitsai@gmail.com

**** 通訊作者，國立中央大學資訊工程系教授、中央研究院人文社會科學研究中心地理資訊科學研究專題中心副研究員，Email: ttsai@csie.ncu.edu.tw

***** 通訊作者，中央研究院歷史語言研究所研究員、人文社會科學研究中心地理資訊科學研究專題中心執行長，Email: mhfanbbc@ccvax.sinica.edu.tw

與史語所「人名權威資料庫」(http://archive.ihp.sinica.edu.tw/ttsweb/html_name/)，來擷取歷史地名、人名等，進行詞頻統計，並進行地圖連結。如此，基本上即可結合地理資訊系統(GIS)來呈現歷史文本所述人、事、時、地等特性，從區域尺度整合各項研究資源；進一步則透過代表性字詞的擷取，建立文字雲，產生各分群之摘要，讓使用者可以快速瀏覽各分群所具有的特有屬性。凡此皆有助於梳理文本、呈現其內容之梗概，從而延伸研究課題，或與相關研究相互呼應，拓展研究視野。本研究實作文本分析流程圖如下：



在系統初步開發完成後，本跨領域團隊之歷史研究人員，便透過此個人版事件分類系統，針對《明實錄》編年史料所述倭寇相關事例進行分析。經過參數調整後，事件分群功能將輸入文本段落分為三群，再檢視各群之高頻人名、地名、段落時間分布、以及文字雲後，我們判讀群 1 內容多為倭寇侵擾與官軍剿寇戰事

相關；群 2 與攻防船艦往來與海防事務相關；群 3 與防倭官員派任與賞罰等行政事務相關。由觀察結果可知，群 1 所見沿海各省皆為倭寇侵擾，例如浙江海門、乍浦、嘉興，及鄰近的南京松江府金山一帶，多有事例記載，時間多在嘉靖年間，而洪武至正統年間各時期亦見相當數量的記載。倭寇侵擾海域向北可及於朝鮮一帶，在《明實錄》中相關事例頗多，例如永樂年間浙江定海衛即追擊倭寇達「東洋朝鮮國義州界」，可見衛所海防任務不限於守禦其海域轄區範圍。群 2 所見浙江沿海衛所佈署與築城時間甚早，洪武年間大致已奠下基礎，九衛之中，由倭寇事件相關文本資料分析可探知，時間最早者為的海寧衛，《明實錄》即述其為洪武中因倭寇擾浙江海鹽縣地臨海地區而設，同時設澈浦、乍浦二千戶所備禦，至正統年間進一步部署，於臨海縣創築城池。以地理位置而言，位於舟山群島的昌國衛尤為浙江海防要衝，於洪武十七年(1384)改所為衛，其後因轄境瀕海居民「嘗從倭為寇」，於洪武二十年(1387)廢昌國縣，遷民為寧波衛卒，並徙昌國衛至象山縣西南天門山(今東門島)，此外，謫戍昌國衛者，「咸出海捕倭以功贖罪」，凡此皆顯示倭寇為患甚烈。倭寇來襲具有時節特性，船隻數量則不定，在嘉靖年間可達二、三百餘艘，而浙江、福建沿海諸衛建造專禦倭寇之船艦，就《明實錄》所見曾達六百餘艘；其造船與相關修防事宜所需經費與人力負擔之繁重，在相關史料中多有反映。群 3 所見嘉靖年間倭患趨於嚴重，增設巡視都御史，巡視浙江並兼管福建漳泉一帶軍務，而倭寇事件涉及之相關人物，及因防禦倭寇功過而升遷或遭貶謫之事，也多見記載，例如太倉衛指揮僉事翁德，因剿平倭患，「獲倭寇九十二人得其兵器海艘」，升為指揮副使，海道副使陶大年則因備

倭失職而降級，及「都指揮等官來熙等革任」。總之，透過使用本研究開發個人文本事件分類系統之各項文本分析功能，如文本各群集之文字雲、高頻人名、地名、段落時間分布，並對照相關文本段落進行解讀，有關明代倭寇事件特性皆可概要地呈現出來。

本研究以個別研究者之需求出發，設計一套歷史文本事件分析工具。與其他個人數位人文文本分析工具相較，本系統著重支援事件與時間空間連結之研究。藉由系統自動分群運算結果，使用者可評估適用的分群數量，並參照系統對於文本內容涉及時間、空間及代表詞彙之辨識、擷取與統計結果，取得文本事件之時空範疇與內容特性，進而透過事件相關文本段落之對照，來輔助研究論述，亦可進一步結合 GIS 空間分析方法，以及運用大比例尺地形圖來進行細部研究。

為驗證本研究工具的有效性，我們挑選倭寇為主題，邀請團隊內人文研究人員使用本工具進行研究。根據實際使用經驗，本工具可協助研究人員大幅節省挑選與分析資料時間。預期文史研究者運用本研究建置之個人版事件分類暨文本分析系統，其專業背景相對可自行發揮的空間較大，操用上也更具彈性，而透過具體的使用經驗，相信可從中發掘更多適用的分析功能需求，促使本系統持續改善輔助歷史研究之效能，以推展數位人文方法於歷史研究應用的可能性。

關鍵字：個人化歷史文本分析工具、事件分群與段落摘要產生、文本自動標註、自然語言處理、地理資訊系統

Development of A Wearable System For Evaluation of Meditation

Pei-Lin Liang^{*}, Bo-Jau Kuo^{**}

Abstract

Introduction: Meditation is recommended as a stress management technique which can induce a calm and relaxed state of mind. As meditation demonstrates its significant advantages, researchers have been trying to find a more scientific way to study the mechanism. Since the 1990s, meditation has applied to multiple mental health conditions, and interest in meditation has grown significantly over the past decades. Previous studies have suggested that meditation can reduce stress, blood pressure, insomnia symptoms and improve parasympathetic nervous system function as well as cognitive function. A number of studies indicated the influence of meditation practice on autonomic function such as breathing patterns, heart rate, skin conductance and blood volume pulse. Despite a larger number of papers and theoretical proposals, little is known about the long-term impact of meditation on the brain. Due to the different styles and forms of meditation, studies of meditation still need for further study. The purpose of this study was to seek further evidence of electroencephalogram (EEG) alterations to determine whether meditative states can be determined from other states of consciousness. Among various EEG rhythms, it was noticed that variation of alpha power provided an index of meditative state transition from the resting state of normal consciousness to meditation. **Hypothesis and aims:** In this study, we focused on the neurophysiological changes of meditation, in particular effects observed in EEG and heart rate variability (HRV) experiments. We hypothesized that wearable devices can record and provide users the personal meditation quality index and improve the positive effects of meditation. Moreover, this study aimed to develop a widely accepted meditation quality index for all meditators. **Materials and Methods:** 12 experienced meditators (EXP, more than one year of meditation experience) and 6 novice meditators (NOV) were recruited in this study; EEG and HRV signals of all participants were continuously recorded before, during and after meditation. The EEG and HRV monitors (KY Lab, Taiwan) used in this research can detect not only brain electrical activity but

* National Yang-Ming University, Email: brightpauline@hotmail.com

** National Yang-Ming University, Email: tbjkuo@ym.edu.tw

also heart rate of the participants. Participants with personal or family history of psychiatric disorders, smoking habits, alcohol addiction and drug abuse were excluded. Informed written consent was obtained from all participants, and the experimental protocol was approved by the Ethics Committee of National Yang-Ming University. A typical experiment consisted periods of 5-minute resting state with eyes closed (baseline), 3-minute breathing state, 30-minute meditative state, 5-minute stretching state and 5-minute resting state after stretching. All participants were comfortably seated in a sound-attenuated room, and the experiment lasted for 50 minutes. We analyzed the physiological phenomenon of variation in the time interval between heart beats, and it was measured by the variation in the beat-to-beat interval using time domain and frequency domain methods. The EEG and electrocardiography (ECG) signals were truncated into successive 64-second (4096 points) time segments with 50% overlapping. For each time segment, Hamming window was applied to attenuate the leakage effect. Our algorithm then estimated the power density of the spectral components based on fast Fourier transform. The resulting power spectrum was corrected for attenuation resulting from sampling and the Hamming window. Low-frequency power (LF, 0.06–0.6 Hz), high-frequency power (HF, 0.6–2.4 Hz) and the ratio LF/HF of the RR spectrogram were quantified. They were logarithmically transformed to correct for the skewness of the distribution. HRV has been categorized into high-frequency (HF) and low-frequency (LF) ranges according to its frequency. HF is equivalent to the well-known respiratory sinus arrhythmia and is considered to represent vagal control of heart rate. LF is jointly contributed by both vagal and sympathetic nerves. The ratio LF/HF is considered by some investigators to mirror sympathovagal balance or to reflect the sympathetic modulations. For example, HF has been shown to decrease in diabetic neuropathy, whereas LF/HF is sensitive to postural change and mental stress. As for EEG, we quantified the theta power (6–10 Hz), alpha power (10–13 Hz), beta power (13–32 Hz) and gamma power (32–64 Hz) of the EEG spectrogram for each 16-s time segment. The theta, alpha, and beta powers of the EEG are dominant during the waking state. Beta power and alpha power have been associated with alertness. Gamma power serves as a gain control for mental processing, enabling postsynaptic potentials to integrate and direct downstream networks to bind the elements of sensory processing into a perceptual object. The baseline value of EEG and HRV was deducted from the other parameters value to exclude the individual difference. Results: Regarding the EEG activity, our data indicated that beta and gamma powers decreased from the breathing training till the meditative training in all participants; EXP had a significant decline in breathing and meditative states as compared to baseline. As for theta power, EXP and NOV had different trends; EXP had a decreasing trend during meditative training, whereas NOV had an increasing trend at

the same period. Alpha power in both groups gradually increased from the breathing training till the meditative training; additionally, EXP had a significant decline during the breathing training as compared to NOV. Results showed that alpha/beta and theta/beta ratios had significantly increased during the meditative training in both groups; increase in alpha/beta ratio may represent the relaxation of the participants, and increase in theta/beta ratio represents the potential biomarker for attentional control. It has been suggested that alpha/beta and theta/beta ratios may be a down-regulation of mental processing activities while remaining fully conscious. As for the results of HRV, it showed no significant main effects of overall autonomic nervous activities in both groups. R-R interval of ECG had a significant increase in EXP during meditative training, which indicated the slower heart rate as compared to the baseline. EXP had a significantly decreasing trend in sympathetic activities which started from meditation till meditation ended. Moreover, significant increase in sympathetic activities was showed in EXP as compared to NOV, which may indicate that meditation task can balance the sympathetic activities. For parasympathetic activities, the effect of meditation task was significant, and there was a lower HF after the meditation task compared to baseline. Discussion: Since beta waves are often occurred in conscious state like decision making, talking, judgement and problem solving, decreasing trend in beta power observed in the present study may represent a decrease in active thinking. Also, a previous longitudinal study in NOV has observed the same phenomenon with our findings. Moreover, it is suggested that increased theta power during meditation reflect focused attention for the contents of the mental experience. Consistent with the subjective description of meditation as a relaxed but alert state, it is likely that results of HRV during meditation can be viewed as balanced control between sympathetic activities and parasympathetic activities. Conclusions: Our findings indicated that a state of meditation is characterized by reduced high-frequency brain wave processing and a reduced trend in sympathetic activities along with the slower heart rate. EEG signals from meditators distinguished the meditative state from other states of consciousness, with the increase in alpha/beta and theta/beta ratios, and this suggests that meditative states might be different from higher cognitive and mentally focused states. Our results substantiated the idea that long-term meditative training induces the changes in the electrocortical activity of the brain, and also suggested that monitoring patterns of physiological variables may indicate the dynamic changes of inner experiences during meditation practice. This study could provide a more precise investigation into the nature of meditation experiences and a more accurate comparison of meditation states with other relaxed conscious states.

Keywords: Electroencephalogram, Heart Rate Variability, Meditation

Customer Choices, Charging Policy And Social Welfare Analysis in A Duopoly Music Industry

Yi-Shin Lin*

Abstract

Digital music platform allows music industry sales pipeline moved from the physical path to a virtual path, along with digital music platform also incorporates digital marketing (Digital Marketing) to create more profit sources. This study was set up digital platform from the recording industry to provide digital merchandise architecture to explore digital platforms how through membership fees, music download fees and advertising fees income sources to find new profit margins in the case of consumer behavior change. And try using Hotelling (1929) to construct digital music platform operators and members and advertisers interaction between regional model, and sub-game perfect equilibrium (Sub-Game Perfect Equilibrium) and push-back method (Backward Induction) two-stage solving and comparative static analysis, by analyzing the results of operations and marketing strategy advice of digital music platforms in the industry.

The results of this study showed that the platform industry to earn profit under the premise of the platform industry to meet the basic utility of consumers is more important than meeting the basic utility of advertisers, in other words, the platform industry if it can't meet the basic utility of consumers, it is difficult through advertising revenue to make up for this shortfall. Therefore, we recommend that if faced with the dual oligopoly situation, without creating product differentiation or development of market positioning, target market and target market (that is, the marketing field STP), if just from the 4P platform is not easy to run and create long-term market prospects and space.

Keywords: Membership Fee, Download Fee, Advertising Expense, Duopoly Industry, Digital Music

* Associate professor, School of Business, Wenzhou University. Email: 20170007@wzu.edu.cn

DocuSky : a DH Platform in the Making

Hsieh-Chang Tu*、Nung-Yao Lin**、I-Mei Hung***、Jieh Hsiang****

Abstract

Our poster will introduce and promote DocuSky, it's a digital humanities research platform and developed by the Research Center for Digital Humanities and CSIE Digital Humanities Lab. Our ideas are hoping to further cooperation with researchers, we are working on developing an integrated research environment for humanities researchers. We hope the researchers not only enjoy the rich collections of digital resources, but they are also able to search, access, organize, and analyze those collections by themselves ideas, and even discover new research topics freely without any constraints from manpower, space, time and system setting. DocuSky is architected as an academic research platform under these ideas. DocuSky is a research platform designated for humanities researchers to coordinate diversified needs in different research stages, and provide them tools to build up personal cloud with international standards. Researchers are able to use these tools for text processing, building databases, tagging, text reorganizing and contextualizing, text-mining, GIS and visualization. Here, we will show through DocuSky's online operation, Welcome to use and advice.

* Research Center for Digital Humanities, National Taiwan University. Email: hsieh.chang@gmail.com

** Research Center for Digital Humanities, National Taiwan University. Email: nungyao@gmail.com

*** Research Center for Digital Humanities, National Taiwan University. Email: yimay0519@gmail.com

**** National Taiwan University. Email: jieh.hsiang@gmail.com

數位人文在清華大學之思行及兼述

清大「圖書館即時偵測服務環境」系統之研發初探

楊叔卿*、林怡瑄**、林宏曄***

摘要

數位人文全球化是二十年來之趨勢，正如 2016 數位人文與數位典藏研討會指出，自 1990 年代起，世界各國推動重要社會文化資產數位化，已將數位化定位為促進知識深化發展的重要機制，不僅致力於基礎建設的奠定和發展，進而鼓勵數位資源與技術在人文社會科學領域之研究教學中的應用和創新。

臺灣清華大學在這波數位人文全球化趨勢中起步甚早，早在 1990 年代，清華大學已成立跨院系所的團隊，整合各面向之人力資源，以創新思維投入數位人文的創思發展與應用，例如，當時清大已在網路上建置推出首座大型的「清蔚園虛擬博物館」及「清蔚園網路知識園」，傳佈新型式特色的數位人文知識；再者，在數位人文內容或議題的設計呈現屢有新創的展現。例如，曾針對當時社會上爭議性極大的核能發電議題等提出新穎的數位人文設計與內容再現，以清大擁有全國獨一無二的核能領域之研究為本，把冷僻艱深的核能學術研究成果，藉由數位人文的創意之轉化，對社會上多元對象以深淺不同的數位型態進行知識傳佈。當時最著名的案例即是，針對兒童核能教育之需，研發出網絡上的第一本多媒體兒童互動動畫 (animated) 電子書等，藉由「小花歷險記」故事中的主人翁~一條擬人魚角色的成長，探索核能生態環境之影響與認知，嘗試透過數位科技的特性與潛力，對不同年齡階層的社會大眾進行不同形式的數位核能知識之傳播。過去十數年來，不斷藉由進展中的前瞻科技，致力於網路虛擬世界中大小分眾數位人文之創新與素養提升、知識的普及深化內容之創新，與試圖把古今學術研發成果以多媒體多樣貌形式傳佈，廣受國內外重視與歡迎。

此外，及至近年來隨著科技軟硬體的快速發展、小型化的進展與普及，行動化數位時代來臨，科技幾乎隨時隨地急遽地影響及改變人們生活的各層面。特別是近年來隨著手持載具，如智慧型手機 iPhone、iPad 等人手一至多機的高度普及，復加上以人為本需求及設計思考理念的興起、受重視及推廣，與對於人文社會創新的渴望，如何透過跨領域教學課程的設計、施行，以培育數位創新人才、整合創新與實踐，皆是近年來推動數位人文的新趨勢與發展方向。

* 國立清華大學學習科學與科技研究所，Email: scy@mx.nthu.edu.tw

** 國立清華大學，Email: julialin822@gmail.com

*** 國立清華大學，Email: s105002504@gapp.nthu.edu.tw

清華大學為因應此發展新方向，於 2017 年二月結合科技部數位人文研究專題資源，及在學校的配合支持下，克服諸多困難，在學習科學與科技研究所首次推出「數位人文專題」課程，由五位教師進行跨領域的共時教學，並加入業師之數位大數據應用分析等題材，致力於數位人文的創新思維，開發潛力與培育人才之實踐。該首推之課程不僅為高等教育教學實踐帶來新的視野，也創立了數位人文跨域課程的新典範。在十八週的課堂中，每週五位來自人文、設計及資訊背景的學者與業師齊聚一堂，共時教學，透過每週不斷地對話與互動，希望能夠激發不同背景研究生的數位人文創新思考、研究能量的持續發展，冀望持續茁壯成熟，進而實踐知識的多元與深化發展。

所以，本論文除了前半部簡要省思回溯清大自二十年前即已開始在數位人文的投入施行外，接下來論文的另一部分擬兼述在該首推的數位人文跨域共時教學課程完成後的一小部分成果，期在此分享融入設計思考 (Design thinking) 觀點初探清華大學「圖書館即時偵測服務環境」研發之思考實踐，本專題是該課程期末學生小組合作之一項專題，亦是本課整合學界、業界跨領域對數位人文思考後多元觀點培養學生的一項成果展現。設計思考是一種思維模式，IDEO 的 CEO Tim Brown 在 Harvard Business Review (哈佛商業評論) 定義它為，以人為本的設計精神與方法，考慮人的需求、行為，也考量科技的可行性，也是本課程探討數位人文體現，引領人文創思之一項依憑。

清華大學於 2013 年建置完成一棟創新的圖書館，它以使用者為中心整合數位科技之建構，該館名為：「學習資源中心-旺宏館」，是一棟複合機能建築，地下一層、前棟四層、後棟七層，空間功能包含：圖書館、大型教室、國際會議廳、清華行政中心，以及人才招募中心。隨著科技發展，該館處處可體現數位科技融合於館內之新穎巧思，提供個人化或群組便捷線上數位資料使用，及多媒體資源各式軟硬體新穎的服務展示，與數位學習應用，相較於傳統到圖書館借閱圖書，具有極高的創新與便利性，旺宏館不再強調傳統的「藏書」功能，而是提供整合線上資源虛實混合的開放共同學習環境 (Learning commons)。其功能強調兩大方向：1. 互動，可以進行團體的分享與討論，區域包含逗留點、清沙龍、討論室、團體室和簡報練習室；與 2. 自學，可以在安靜的環境自學研讀，包含各樓層之閱覽席、夜讀區、安靜閱覽區、研究小間與電腦共學區，是具有多元學習環境的圖書館，空間開放且景觀開闊。

依據 2016 年的統計，該圖書館平均每小時有 214 人次入館，而服務的讀者，包含學生 16,547 人、教職員工 3,074 人、以及其他辦有借書證的讀者 40,640 人，該數據顯示，清大總圖書館使用人次多，且服務對象不只限於學校師生。因而，目前圖書館的服務環境維管之考量，需包含：背景情境，如圖書館內溫度、噪音音量等，會與不同區域的使用者和使用人數，以及不同時間點有關；而背景情境又與使用者的舒適度有關係。即不同區域、不同時間會呈現不一樣的氣氛。例如：學校期末考試週，會有很多學生前往圖書館進行複習，某些人多的區域不只是溫度會提升，紙筆書寫、電腦打字都會提高噪音音量，這樣可

能會造成學生更大的壓力感。因此，要如何改進以使用者為中心的優良服務，在數位人文的理念與應用設計思考的原理下，是否使用者可以在進入圖書館前，先知道各個區域的氣氛，選擇自己適合的理想位置，提升其互動、自學的效率，是本研究之目的，因而，擬探討如何應用手持數位科技即時互動特性融入圖書館服務環境系統之研發。

所以，在系統的研發上，本研究問題包含以下兩大面向：1. 如何透過行動裝置的特性應用可設計出一個符合使用者對清華大學圖書館的服務環境之情緒感受向度？2. 如何有效率地科技應用與介面呈現讓使用者可以視覺直觀的感知圖書館當下情境環境以作即時有效之場境判斷選擇？

為了深入及有效設計與施行本研究，探討了下列面向的文獻，作為規劃及落實本研究之基礎與依憑。

一、圖書館之服務環境：由於電子資源取得便利，圖書館的功能不再只重視豐富的視聽圖書資源，而是學習和教學的溝通場所，圖書館的服務包含：1) 提供安全、舒適和安靜。2) 免費。3) 提供與他人在同個學習環境的場所。4) 提供學習、探索、查詢和重現的機會。及 5) 給予選擇和意外收穫的機會。清華大學的圖書館被定位為數位資源中心，「以讀者為主的資源利用與交流平台」，提供的服務，例如：各層的閱覽區具有安靜、舒適，且提供與他人在同個學習環境；二樓的電腦供學區提供探索、查詢；而一樓的清沙龍和「知識集 新知，主題展」的展覽區，時時提供了不同主題的內容和演講，是讀者意外收穫的機會。清華大學圖書館滿足使用者的需求，使其持續使用圖書館。

二、服務場景及環境心理學的因素考量：環境心理學是指研究人和環境相互作用，環境影響人的想法、行為，人也會改變環境。而服務場景 (Servicescape) 是一個模型來解釋在服務的環境中使用者的行為，以設計環境讓使用者產生預期的反應 (Booms；在服務場景之模型中的因素，Baker (1987) 認為可分為三類：1. 潛在因素 (Ambient factors)、2. 設計因素 (Design factors) 與 3. 社會因素 (Social factors)。潛在因素是指環境特質，例如：圖書館內的溫度、噪音、氣味等，會影響行為和情緒。若圖書館無法提供或是沒有達到一定程度的潛在因素，會對使用者的行為有中性或是負面影響，而採取規避行為 (Avoidance behavior)，可能會換座位或是減少來圖書館的次數。設計因素為使用者的視覺觀感，可分為美感，如建築內、外的色彩；與功能，如配置、指標，比較容易塑造使用者正面的影響，和鼓勵顧客的趨近行為 (Advent behavior)。社會因素是在環境中互動的人，外表、行為和數量都會影響使用者的感受，正向或負向的影響都有可能。例如：服務台大排長龍，和多人一起參加圖書館活動，會對行為和情緒產生相反的影響。

其中，潛在因素與社會因素可能會根據圖書館不同區域、不同時間點而有所不同，例如：夏天會開啟較強的冷氣，但是人多的區域可能不會太涼爽、或是在期末考試的時段，閱覽區人數越多的地方可能會對使用者產生正面的感受。雖然設計因素容易塑造使用者正面的影響，但沒有潛在因素和社會因素改變得容易，

因此，本研究著重於了解使用者潛在因素和社會因素兩方面的感受。

三、圖書館環境數據之相關研究：透過定時測量圖書館不同分館之環境數據，包含：溫度、濕度、風速量，同時利用問卷了解閱讀者之基本資料：性別、年齡、身高、體重、衣著，與熱舒適度：溫度感覺指標、濕度感覺指標、風速感覺指標、環境舒適度滿意度，這兩方面的數據進行相關系統分析與交叉分析。提供了包含夏季男女生敏感度不同，與分館不同感受之建議等等。

四、色彩感知會影響情緒：色彩心理是透過視覺開始自知覺、感情而後到記憶、思想、意志、象徵等的過程，而情緒體驗(Emotional Experience)是個體透過主觀的感受、知覺及意識的情緒狀態，在心理組成上，情緒是接受了外在刺激所產生的情感評價，再喚起個體的心情與記憶，最後個體知覺到自己的情緒片段，才意識到情緒的產生的過程。情緒回應是根據個體對該物體的經驗，但是通常人們感知色彩的色彩情緒(Color Emotions)是高度相關。

因此，本研究依據 CIEL*a*b*色彩空間對應 Russell 環境情緒體驗環圖，與前述服務場景的模型因素，為清華大學圖書館設計出三個感受向度，透過行動裝置的應用與介面呈現：1) 溫度（熱的程度），為紅色；2) 噪音（安靜的程度），為藍色；及 3) 壓力（輕鬆的程度）為綠色，例如：溫度較熱（偏紅）加上較輕鬆（偏綠），對應到情緒體驗環圖為高興、興奮（黃色）；較安靜（偏藍），加上較輕鬆（偏綠），對應到情緒體驗環圖為放鬆、困倦（藍綠色）。

清大「圖書館即時偵測服務環境」系統設計：依據前述資料彙整與歸納後，本研究擬研發一系統平台及使用之開發工具，如下：

1. 系統平台：Android、iOS
2. 開發工具：Unity、Notepad++
3. 伺服器端：Apache、MySQL
4. 程式語言：C#、PHP

系統功能與使用情境：該系統包含兩個功能：一是透過手機介面呈現圖書館之氣氛，選擇使用者想要入座的區域；二是對於入座的區域進行使用者情緒感知互動回饋。因而，本系統包括：系統首頁；瀏覽樓層；選擇樓層；選擇區域；感受回饋；及感受回饋後之畫面回應。在「選擇坐位」功能上，使用者在進入圖書館前，瀏覽圖書館各樓層、各區域氣氛的顏色區塊，顏色區塊為 2 小時內的感受回饋統整，並會讓使用者對應到特定的情緒，因此可以選定欲入座的位置。另一方面在「感知互動回饋」功能上：使用者在使用該區域的位置時，可以對當下的感受進行回饋，系統會將 2 小時內蒐整的回饋之內容呈現在顏色區塊中。

綜言之，本論文分二部份，前一部份簡要回溯台灣清華大學自二十年前投入數位人文之努力，及另一部份述及近日清大進行數位人文整合教研之於教學創思與分享一項實踐之研發案例。目前清大「圖書館即時偵測服務環境」系統初步具體構思已完成及描述如上，其獨特處在於以往在環境偵測上只是紀錄客觀的數據，如溫溼度值、噪音的分貝，本研究具即時雙方互動，例如：假設在某一溫度的空調下，位於清華大學總圖書館的人普遍的感受。另外，感受向度未來也可以不受

限制，即由使用者來定義感受向度，回饋的向度也依使用者來決定。另外，相關研究指出，色票與實質物體色彩給予人的感受不盡相同，是一個情緒體驗，因此，未來可以建議服務場景中的設計因素可以做改變。此外，本系統另一特色擬以文字雲視覺設計方式呈現所蒐集的大數據，藉由文字雲中的文字大小與顏色提供趣味及快速直覺的數位資料解讀與互動。

期望完成該系統後，未來經由個人化手持裝置之環境服務系統可有主動、趣味性地長期收集清華大學圖書館使用者之主觀感受之大數據，進行統整及分析，達到數位人文之另一創思與實踐，不斷改善圖書館數位資源學習環境之人本需求。

關鍵字： 數位人文、清華大學、清蔚園、圖書館服務系統

英國非物質文化遺產類數字人文項目的實踐和啟示：

以 art-humanities.net 網站為調研核心

林佳萍*

摘要

近年來，一批數字人文研究機構，如國際數字人文組織聯盟（The Alliance of Digital Humanities Organizations）、數字人文學會（The Society for Digital Humanities）相繼成立。許多大學還設立了自己的數字人文研究中心，如美國斯坦福人文實驗室、倫敦國王學院人文計算研究中心等。計算機與網絡技術的應用逐漸深入到人文研究的各個領域，數字技術與人文研究的結合成了學者時下討論的熱門話題。

art-humanities.net 始建於 2008 年，是由倫敦國王學院的 e-Research 中心（Centre for e-Research, CeRch）開發和管理的一個數字藝術和人文學科的教學和研究網絡平台，並得到了英國藝術與人文研究理事會（Arts and Humanities Research Council, AHRC）的支持。旨在通過提供創建和使用數字資源的工具和方法及其具體項目信息來支持和推進藝術人文領域的教學和研究對數字工具和方式的理解和使用。art-humanities.net 是一個開放式的合作平台，任何對數字藝術人文領域的研究感興趣的組織、機構和個人均可以註冊成為會員進行學習和交流，同時它也是 Network of Expert Centres 網站的主頁。Network of Expert Centres 也是面向數字藝術人文領域研究和教學的專業中心的合作交流平台，其會員資格僅限於大不列顛和愛爾蘭的專業研究和教學中心，目前已有包括約克大學的 Archaeology Data Service (ADS)、貝爾法斯特女王大學的 Centre for Data Digitisation and Analysis (CDDA)、牛津大學的 Oxford e-Research Centre、劍橋大學的 Centre for Research in the Arts, Social Sciences and Humanities (CRASSH) 等在內的 22 個專業中心加入了其會員，網站累計上傳的項目已達 605 個，包括音樂、語言、歷史、舞蹈、戲劇、哲學、法律等 19 個門類，並按字母順序（alphabetical）、內容類型（content types）、學科門類（discipline）、和資助機構（funding bodies）4 種方式進行排序或組合。雖然這些項目的研究機構多為英國的教學和研究中心，但項目所涉及的具體研究對象有一部分屬於亞洲、美洲和歐洲的其他國家。

本研究的研究目標即以 art-humanities.net 網站為調研核心對英國高校相關科研機構開展的非物質文化遺產類數字化項目進行總體的梳理和分析，為中國大陸非物質文化遺產數字化項目的運行與發展提供思路。根據 2003 年聯合國教科文組織頒發的《非物質文化遺產公約》中的定義，“非物質文化遺產”即“被各

* 廣州中山大學資訊管理學院，Email: 1024527910@qq.com

社區、群體，有時是個人，視為其文化遺產組成部分的各種社會實踐、觀念表述、表現形式、知識、技能以及相關的工具、實物、手工藝品和文化場所。這種非物質文化遺產世代相傳，在各社區和群體適應周圍環境以及與自然和歷史的互動中，被不斷地再創造，為這些社區和群體提供認同感和持續感，從而增強對文化多樣性和人類創造力的尊重。在本公約中，只考慮符合現有的國際人權文件，各社區、群體和個人之間相互尊重的需要和順應可持續發展的非物質文化遺產。”

本研究將依據上述研究目標，循序展開：(1) art-humanities.net 網站項目梳理，在 605 個各類型的項目中手動挑選出信息完備的非物質文化遺產類數字化項目，(2) 概括整理挑選出的項目的基本情況和組織機構建設等信息，取得對項目的總體認識，(3) 著重針對不同類型的非物質文化遺產數字化項目的數據庫體系或平台建設中的資源加工和資源發布進行論述，(4) 總結英國上述數字化項目的經驗及對中國的啟示。研究方法方面主要採用內容分析法對 art-humanities.net 網站項目進行兩個層次的分析。第一，總體層面：梳理所整理的項目的基本情況和組織機構建設情況，包括起止時間、學科領域、年代、國家、材料來源、參與單位、資金來源等，構建總體視角；第二，個案層面：針對不同類型的非物質文化遺產數字化項目的數據庫體系或平台建設的資源加工和資源發布情況進行重點分析，包括平台架構與信息呈現、搜索引擎、元數據標準、(使用的方法及類別、軟件工具、數字資源創建、數據格式創建、創建內容類型、出版物、)，深入挖掘項目經驗。

以單字切割與水平排列方法 提升中文古籍影像的文字辨識能力

吳昱賢*、蘇冠宇**、王祥安***

摘 要

中文古籍的字體、排版、書寫方式會造成影像文字辨識的辨識率不佳，甚至無法辨識。本論文主要研究從中文古籍影像中切割單字及重新排列文字方向的方法，並實驗這兩個方法對影像文字辨識的影響，實驗結果顯示，本研究的方法能大幅提高辨識軟體對中文古籍的文字辨識率。

關鍵字：單字切割、水平排列、影像文字辨識

* 元智大學資訊工程學系，Email: ymmh123456789@gmail.com

** 元智大學資訊工程學系，Email: watermelo0326@gmail.com

*** 中央研究院數位文化中心，Email: sawang@gate.sinica.edu.tw

共現字分析工具於競選政見應用之初探

戴浩光*

摘 要

本文針對網絡上蒐集的 559 筆 2016 年立法委員競選政見，利用數位工具中的共現字分析（co-word analysis）技術，進行政見的主題分群（clustering）。

分析結果顯示，選舉政見可分成「改革」、「政治制度」和「兩岸主權」三個政治性議題政見，以及「社會與家庭關係」、「地區事務」和「民生議題」三個民生性議題政見。其中，前者佔總數超過 77%，可見政治議題是選舉中的焦點。再者，只有國民黨和民進黨候選人的政見能涵蓋 4 至 5 個議題，代表大黨能藉此覆蓋較多的議題和受眾。最後，結果顯示國民黨候選人較多偏好「兩岸主權」議題，民進黨候選人則較多偏好「政治制度」和「社會與家庭關係」議題。

本文示範了如何應用數位工具分析政治文本，亦顯示出數位工具有助理解選舉中的議題焦點，以及分析背後反映的社會狀況。

關鍵字：立法委員、政見、共現字分析、主題分群、文字探勘

* 國立臺灣師範大學東亞學系，Email: hugotai01@gmail.com

支援數位人文研究之 文本自動標註系統發展與使用評估研究

蔡明月*、陳志銘**、劉鎮宇***

摘要

本研究發展支援數位人文研究之「文本自動標註系統」，藉由 Linked Data 的概念匯集來自不同資料庫的資源，並加以整合後，替文本進行自動注解，讓使用者在解讀文本時能夠即時參照其他資料庫的資源，並提供友善的具文本標註之閱讀介面，以利於人文學者透過閱讀進行資料的解讀。本研究以實驗研究法比較本研究所發展之「文本自動標註系統」與「MARKUS 文本半自動標註系統」在支援人文學者進行文本資料解讀之閱讀成效與科技接受度是否具有顯著差異，並輔以半結構式深度訪談了解人文學者對於本研究發展之「文本自動標註系統」的看法及感受。

實驗結果發現，採用本研究發展之「文本自動標註系統」的閱讀成效高於「MARKUS 文本半自動標註系統」，但未達顯著差異；而「文本自動標註系統」之科技接受度顯著優於「MARKUS 文本半自動標註系統」。另外，「文本自動標註系統」閱讀介面簡潔明瞭，比「MARKUS 文本半自動標註系統」更適合閱讀，而閱讀介面是否易於使用與是否有用，是影響人文學者能否接受採用系統輔助數位人文研究的重要因素。此外，在兩個系統類似功能比較分析後也發現，「文本自動標註系統」在查詢詞彙功能、連結到來源網站功能及新增標註功能都比 MARKUS 文本半自動標註系統更為直覺易用，另外人文學者普遍認為斷句功能比自動斷詞功能更重要，鏈結來源資料庫則以萌典最有幫助。

關鍵字：數位人文、文本自動標註系統、中文自動斷詞、鏈結資料

* 國立政治大學圖書資訊與檔案學研究所特聘教授兼圖書館館長、社會科學資料中心主任，
Email: mytsay@nccu.edu.tw

** 國立政治大學圖書資訊與檔案學研究所特聘教授兼所長、圖書館副館長，Email:
chencm@nccu.edu.tw

*** 國立政治大學圖書資訊與檔案學研究所碩士班研究生，Email: mood51002@gmail.com

第八屆數位典藏與數位人文國際研討會論文集

8th International Conference of Digital Archives and Digital Humanities Conference Proceedings

研討會日期：2017年11月29日-12月01日

研討會地點：國立政治大學

研討會網址：http://www.aiecon.org/conference/DADH2017/index_ch.htm

舉辦單位（照筆劃順序）：

國立政治大學

國立政治大學人工智慧經濟學研究中心

國立臺灣大學數位人文研究中心

臺灣數位人文學會

- ※ 本論文集由第八屆數位典藏與數位人文國際研討會中之發表內容彙編，所有文章及圖片內容屬原作者所有
- ※ 本論文集內容受著作權法保護，任何形式轉載、重製、散布、公開播送、出版或發行本論文集內容應取得該著作財產權人同意或授權後，方得利用
- ※ 非經著作權人同意，不得轉載

2018年1月彙編