



**陳志銘**  
**Chih-Ming Chen**

陳志銘，目前為國立政治大學圖書資訊與檔案學研究所教授兼圖書館館長、社會科學資料中心主任，過去曾擔任國立政治大學圖書資訊與檔案學研究所所長、政大圖書館副館長、政大電算中心副主任等行政工作，曾獲得科技部吳大猷先生紀念獎、政大特聘教授、政大學術研究特優獎、政大學術研究優良獎等學術殊榮，過去合計已發表超過 70 篇 SSCI 及 SCI 等國際期刊論文，也出版過「DSpace 開放源碼數位典藏系統建置理論與實務」及「創新數位學習模式與教學應用」二本專書，主要研究專長包括數位學習、數位閱讀、數位人文、數位圖書館、大數據資料探勘、人工智慧與智慧型網際網路系統。目前擔任政大圖書館及圖書資訊與檔案學研究所共同發行「圖資與檔案學刊」主編，以及國立臺灣大學圖書資訊學系發行之「圖書資訊學研究」(TSSCI 收錄) 編輯委員，Elsevier 發行之 Computers & Education: Artificial Intelligence 國際期刊編輯委員，也曾擔任過科技部資訊工程學門規劃委員，以及資訊教育學門、圖書資訊學門複審委員。

個人網頁：<http://dlil-info.dlll.nccu.edu.tw/>

Chih-Ming Chen is currently a Distinguished Professor in the Graduate Institute of Library, Information and Archival Studies at National Chengchi University, and he is also the University Librarian at National Chengchi University Library, Taipei, Taiwan. He received B.Sc. and M.Sc. degree from the Department of Industrial Education at National Taiwan Normal University in 1992 and 1997, and received Ph.D. degree from the Department of Electronic Engineering at National Taiwan University of Science and Technology in 2002. His research interests include digital library, digital humanities, e-learning, data mining, machine learning, and intelligent agents on the web. He has received many awards for his outstanding researches.



**張鐘**  
**Chung Chang**

張鐘，國立政治大學圖書資訊與檔案學研究所碩士，目前為華人文化主體性研究中心資訊工程師，研究興趣為數位圖書館、數位典藏、數位人文、數位學習，並發表多篇數位人文相關期刊與研討會論文，過去曾參與過羅家倫數位人文研究平台、明代文集數位人文研究平台、馬來西亞人物誌數位人文平台等建置及多項數位人文工具開發工作。

個人網頁：[http://www.ccstw.nccu.edu.tw/team/m/ 張鐘 /](http://www.ccstw.nccu.edu.tw/team/m/)

Chung Chang received B.Sc. degree from the Department of Information and Library Science at Fu Jen Catholic University in 2016, and obtained his master degree from the Graduate Institute of Library, Information and Archival Studies at National Chengchi University in 2018. He is currently an IT engineer of the Research Center for Chinese Cultural Subjectivity at National Chengchi University, Taipei, Taiwan. His research interests include Digital Humanities, Digital Library, and E-learning.



**徐志帆**  
**Chih-Fan Hsu**

徐志帆，國立政治大學圖書資訊學與檔案學研究所碩士，目前為國立政治大學圖書資訊與檔案學研究所專任助理，具有開發數位人文、數位學習平台經驗。

個人網頁：<https://billxu0521.github.io/>

Chih-Fan Hsu received his master degree from the Graduate Institute of Library, Information and Archival Studies at National Chengchi University. His research interests include Digital Humanities, Digital Library, and E-learning.



**林嘉培**  
**Chia pei Lim**

林嘉培，馬來西亞華人，國立政治大學哲學系碩士，拉曼大學中華研究院中文系學士。目前為國立政治大學華人文化主體性研究中心計劃專任助理。研究興趣：數位人文、數位典藏、海洋史、沃格林研究（種族）

Chia Pei Lim received his Master degree from the Department of Philosophy at National Chengchi University in 2019, and received Bachelor's degree from the Institute of Chinese Studies at University Tunku Abdul Rahman in 2014. He is currently a Project Assistant in the Research Center for Chinese Cultural Subjectivity at National Chengchi University, Taiwan. His research interests included Digital Humanities, Digital Archive, Maritime History, and Eric Voegelin's on Race Issues.

# 馬來西亞人物誌數位人文平台之建置與應用

## Development and Application of Digital Humanities Research Platform for Biographies of Malaysia Personalities

陳志銘

Chih-Ming Chen

國立政治大學圖書資訊與檔案學研究所教授兼圖書館館長

Professor & University Librarian, Graduate Institute of Library, Information and Archival  
Studies at National Chengchi University

張鐘

Chung Chang

國立政治大學華人文化主體性研究中心資訊工程師

IT Engineer, Research Center for Chinese Cultural Subjectivity  
at National Chengchi University

徐志帆

Chih-Fan Hsu

國立政治大學圖書資訊與檔案學研究所專任助理

Research Assistant, Graduate Institute of Library, Information and Archival Studies  
at National Chengchi University

林嘉培

Chia pei Lim

國立政治大學華人文化主體性研究中心專任助理

Research Assistant, Research Center for Chinese Cultural Subjectivity  
at National Chengchi University

### 摘要

馬來西亞人物誌數位人文平台為國立政治大學華人文化主體性研究中心與馬來西亞興安會館、馬來西亞馬來亞大學華人研究中心合作建置，此一平台以「馬來西亞興化群賢錄」作為典藏內容，採用開放原始碼的數位典藏系統 Omeka S 為基礎發展系統功能，結合近讀 (Close Reading) 及遠讀 (Distant Reading) 兩個面向功能，作為本平台發展數位工具

的基礎。近讀為透過細讀文本全文解讀內容，能對文本內容進行深入的理解，因此本平台提供友善圖文閱讀介面，並搭配多個輔助文本閱讀工具，例如實體命名 (Name Entity Recognition) 等功能，幫助研究者理解人物誌中的內容；遠讀則是透過文本分析方法，幫助研究者從宏觀的角度解讀文本，除關鍵詞分析外，另提供地理資訊、社會網絡分析等功能輔助使用者掌握人物誌中人物關係與地理位置分布狀況，有助於加速解讀文本效率，探討隱含在人物誌中的脈絡。本文介紹了開發馬來西亞人物誌數位人文平台的第一階段成果。除了優化該平台的功能和使用者界面外，我們未來的工作還將繼續開發更多新功能，例如個人資源檔案，分層主題分析工具，觀點遷移和年代劃分的分析工具。並嘗試應用此數位人文研究平台來協助數位人文科學的教學。

**關鍵詞：數位人文、Omeka S、馬來西亞人物誌、人物社會網絡分析、地理資訊系統**

## **Abstract**

The Research Center for Chinese Cultural Subjectivity of National Chengchi University co-developed a digital humanities research platform for Biographies of Malaysia Personalities (DHRPBMP) with Malaysian Chinese Research Center of Malaya University. We adopted “The Malaysia Henghua Personalities” as basic corpus for the DHRPBMP and chose Omeka S, which is a new generation open source of institutional repository system, as the foundation for the development of DHRPBMP. The DHRPBMP provides several basic functions including a convenient and user-friendly reading interface with the support of an automatic text annotation tool based on name entity recognition (NER), linked data (LD), metadata search, full-text search, geographic information system(GIS), and character social network analysis to support digital humanities research. The developed DHRPBMP aims to assist digital humanists to efficiently read or analyze the contents of The Malaysia Henghua Personalities through distant reading and close reading. This paper presents the first stage outcomes of developing the DHRPBMP. In addition to optimizing the functions and user interface of this platform, our future work will also keep on developing or integrating more new functions, such as personal resource archives, hierarchical topic analysis tool, observation tool of viewpoint migration and age division (OTVMAD) as well as try to apply this digital humanities research platform to assist digital humanities' instruction.

**Keywords:** Digital Humanities, Omeka S, Biographies of Malaysia Personalities, Character Social Network Analysis, Geographic Information System

## 壹、緒論

隨著網路與資訊科技的快速發展，數位化已經成為全世界的發展潮流與趨勢，而隨著數位科技的進步，已逐漸改變了傳統人文領域的研究方法，為人文研究創造新的思維。在過去，人文學者多半以紙本的形式進行人文研究，現今則透過紙本數位化後建立各種不同的數位典藏資料庫，再加上發展可以分析數位典藏資料庫文本的數位工具輔助研究，已對人文研究領域的研究模式與方法產生巨大的改變與影響。換言之，透過資料庫輔以進行人文研究，已逐漸成為人文學者考量的研究模式之一 (Hockey, 2004)，也逐漸發展為「數位人文學」之新興研究領域。面對此一數位時代，人文研究者也開始積極思考，要如何運用資訊科技輔以人文研究，來幫助尋找在前數位時代中難以觀察的現象、無法想像的議題與無法進行的研究 (項潔、涂豐恩，2011)。

數位人文學為數位技術與人文學科融合的新興研究領域 (Drucker, 2013)，它包括有系統的使用人文學科的數位內容和工具，並使新的教學和研究成為可能。在數位人文領域的研究中，Correll(2014) 指出細讀 (Close Reading) 的鏈接是支持遠讀 (Distant Reading) 假設的重要方法，近讀係透過細讀文本全文來解讀內容，能對文本內容有著更深一層的理解；遠讀則是透過文本分析方法將文本進行抽象化之概念抽取，幫助人文學者從宏觀的角度解讀文本，例如檢索結果分布之視覺化，可以輔助人文學者掌握重要關鍵詞彙隨著時代、體裁及文章等的分布狀況觀察，有助於加速解讀文本效率，探討隱含在文本中的脈絡。在 Jänicke, Franzini, Cheema 與 Scheuermann(2017) 等人的研究中也提及，在使用遠讀視覺化工具時，若可以直接鏈結到來源文本，對於人文學者來說非常重要。目前所發展的數位人文分析工具的準確度仍難達成百分之一百，因此數位人文分析工具應該要扮演的角色是輔助人文學者研究，而不是取代它。因此，數位人文分析工具應該有可以顯示內容摘要或者抽象層次的遠讀，同時也要有能夠鏈結遠讀細讀的功能，如此會更便於研究人員對於分析文本的解讀 (Moretti, 2016)。

社會網絡分析為遠讀的分析方式之一，能協助人文研究者找出隱含在大量史料、文本中複雜的網絡關係。哈佛大學與中央研究院歷史語言研究所及北京大學所合作之中國歷代人物傳記資料庫 (CBDB)，就是透過累積大量的中國歷代人物傳記資訊，利用群體傳記學 (prosopography) 的方式輔以人文研究 (中國歷代人物傳記資料庫，2020)。目前已有許多國內外學術機構、學術圖書館陸續開發出支援不同研究取向的數位人研究平台及數位工具。例如國家圖書館與國立政治大學合作發展之明人文集古籍數位人文平台，提供視覺化分析、自動標註工具和社會網絡分析功能 (Chen & Chang, 2019)；國立政治大學圖書館於

開發之「羅家倫先生文存數位人文研究平台」，除提供視覺化分析、文本標註等工具外，亦根據「羅家倫先生文存」文本之特性，並結合概念史和資訊技術發展觀點變遷與年代劃分工具（陳志銘、張鐘與徐志帆，2020）；中央研究院數位文化中心開發之數位人文研究平台允許研究者上傳文本與權威詞、手動與自動標記文本等功能，並有統計圖表、文字雲、社會網絡分析圖等視覺化分析工具（中央研究院數位文化中心，2020）；臺灣大學數位人文研究中心「DocuSky」平台，能讓使用者上傳文本，建立個人化的雲端資料庫，並提供了文本標記、文本探勘、視覺化、時空分析等可輔助人文研究者進行文本解讀之數位工具（杜協昌，2018）。

馬來西亞為除兩岸外，華人族群比例最高的國家之一，2019 年國立政治華人文化主體性研究中心研究成員組團至馬來西亞進行訪查，並與馬來西亞馬來亞大學華人研究中心於 2020 年合作開發「馬來西亞興化群賢錄數位人文研究平台」，平台中支援繁體中文、簡體中文檢索，提供近讀及遠讀兩種模式工具，如文本標註、時空分析及社會網絡分析等，能幫助使用者快速理解馬來西亞華人的發展脈動。

## 貳、數位人文研究平台發展與功能現況

要藉由數位科技輔助人文學者進行文本研究與分析，首要條件是需具備數位化之文本資料，其次是數位人文分析工具。因此，數位人文的發展可以分為兩個階段，第一個階段是「數位典藏」（Digital Archive），即是將原先的紙本進行全文數位化；第二階段即是發展數位人文分析工具。

數位典藏計畫除了將人文資料進行永久保存外，也在網路上開放學術利用，為各領域的人文研究者帶來易取得而豐富的研究資料，從而提高了人文研究者的研究效能。此外，「中國歷代人物傳記資料庫 (CBDB)」係由哈佛大學、臺灣中央研究院歷史語言研究所及北京大學三個機構合作建置，該資料庫收錄超過 470,000 筆中國歷代人物傳記和譜系資料，研究者可以透過此資料庫檢索歷史人物之籍貫、官職等資料，並提供 API 介接的方式讓數位人文研究者使用數位工具取得其中的資料（中國歷代人物傳記資料庫，2020）。然而，現代的數位人文研究者已不再滿足於建立單純只具典藏功能的資料庫，而是開始建置各式各樣的數位人文研究工具，法鼓山的「CBETA 數位研究平台」除了提供漢文佛經的線上閱讀介面之外，亦提供「對讀文本編輯」、「部類、年代、作者與譯者統計分析」、「詞彙前後綴詞、分布地圖」等功能，供佛學研究者參考（Tu, Hung, & Lin, 2012）。DocuSky 數位人文學術研究平台是由

國立臺灣大學數位人文研究中心與資訊工程學系數位典藏與自動推論實驗室所開發，係一個針對數位人文研究需要所開發的平台。此一平台由「機構導向資料庫」轉向「個人導向資料庫」，讓研究者能透過工具自行建立資料庫（翁稷安，2016；謝博宇，2016）。研究者除了可以上傳自身所典藏的文本外，平台亦提供數位工具讓研究者能夠對建構的文字庫進行統計分析，幫助其查找資料、進行內文比對，以及進行字詞相關統計等，增加研究者對於文本內容的掌握（杜協昌，2018）。此外，「羅家倫先生文存數位人文研究平台」則是由為國立政治大學社會科學資料中心所發展，主要典藏「羅家倫先生文存」合計共十二冊之數位化版本，並且為了因應研究需求，平台亦提供掃描檔與對照全文、全文檢索與後分類、自動文本標註（Chen, Chen, & Liu, 2019）、觀點變遷分析，以及階層式主題分析等功能，以輔助使用者進行文本閱讀與多面向的文本分析。而「通用型古籍數位人文研究平台」則除了提供文本閱讀介面、全文檢索、雙詞檢索與自動標註等功能外，亦提供人物關係分析圖，可以自動從文本中識別人名，並透過視覺化方式呈現人物與人物之間的關係脈絡，協助人文學者理解文本中的人物社會網絡關係（Chen & Chang, 2019）。

綜合以上所述，人文學者已皆越來越仰賴於「工具」的輔助，並運用數位工具來解決過去的人文研究問題，而一個支援數位人文研究需具備近讀及遠讀兩種模式的工具來輔助人文學者進行研究，並且人機合作才能進行準確度高與有效產出之數位人文研究模式。

## 參、實體命名與社會網絡分析

命名實體識別（Named Entity Recognition, NER）是標註功能中常使用到的技術之一，可以識別文本中具有特定意義或代表性的實體，包括人名、地名、組織名、時間日期等專有名詞，亦廣泛應用於地理資訊系統（Geographic Information System, GIS）的地名識別、社會網絡分析（Social Network Analysis, SNA）的人物名稱識別上。在不同語種的文本分析上，英文的命名實體識別相關研究發展甚早，但由於只需考慮詞本身的特徵而不涉及分詞問題，因此實現難度相對較低。而在中文的分析上，因其內在的特殊性促使在文本處理時必須先進行詞法分析，相較於英文命名實體識別具備更高的挑戰性（Sun & Wang, 2010）。目前已有許多數位人文工具應用實體命名技術發展相關輔助數位人文應用，例如荷蘭萊頓大學的MARKUS文本分析和閱讀平台即是一個線上文本標註工具，允許使用者上傳文本並選取所需的標註類型，系統會針對人名、地名、官名等專有詞彙進行自動標註，並提供使用者維基百科、中國歷代人物傳記資料庫（CBDB）、中國歷史地理地名資料庫（TGAZ）與漢典（ZDict）等資源之外部參照（Ho & Hilde, 2014）。Yohan（2014）等人提出了一個基於規則的命名實體

識別和分類系統，可以自動識別文本中的命名實體，並為它們提供對應於泰盧固語 (Telugu language) 的適當類別。Kestemont 和 Gussem (2017) 則在深度學習領域應用了分層神經網絡架構，以自動解決兩個序列標記的任務，包括詞性標記 (part-of-speech tagging) 和詞形還原 (lemmatization)。其研究結果顯示，在可接受的成效上與先前發表的研究成果相當。

人文研究應用社會網絡分析多半為進行文本脈絡探究，將社會網絡理論應用於文本的情節分析中，利用社會網絡分析將敘事性的文本中具有空間與時間結構、核心角色等關係轉換成網絡結構圖，隨著情節的發展，網絡圖也會有所轉變，能讓研究者輕易了解文本中關係與情節的變化 (Moretti, 2011)。Moretti(2011) 藉由莎士比亞的名著《哈姆雷特》的社會網絡圖找出關鍵人物 Horatio，並從 Horatio 的社會網絡關係探討故事的脈絡。趙葳 (2016) 也用此概念研究李劫的《大波》三部曲中的人物關係，以及故事的情節變化。許多國內外經典或著名文學，例如《聖經》、《魔戒》、《三國演義》等，都有許多學者運用社會網絡分析對文本中的人物關係、故事情節進行探討 (Duling, 2000；Ribeiro, 2015)。

網絡分析結合多來源文本也能呈現人物的生平交往或遷移過程，以及整個大環境時空脈絡。So 和 Hoyt(2013) 曾分析跨太平洋的文學社團與文學活動，從中整理出 1920 年代現代主義詩歌在全球傳播的路徑；劉吉軒 (2012) 分析臺灣海外左派刊物，用社會網絡圖呈現出海外臺灣人士的政治思想輪廓；金觀濤等人 (2016) 研究《新青年》雜誌中群眾觀念的變化，分析重要關鍵詞變化，並利用概念網絡圖呈現《新青年》雜誌從自由主義轉向馬列主義的過程。史丹福大學的「Mapping the Republic of Letters」網站呈現啟蒙時代知識份子透過書信交流所形成的社會網絡 (Stanford University, 2013)。Chen 等人利用社會網絡分析建構明代藏書家之間的社會關係模型，從中探索藏書家社群的核心人物及呈現社群中不同時期的傳承性和關聯性 (Chen, Chen, & Huang, 2018)。

目前國內外已有許多數位人文研究平台能提供相關社會網絡數據或分析功能協助研究者進行社會網絡分析，例如「中國歷代人物傳記資料庫」提供開放版本的資料庫供研究者下載，研究者能將資料匯入社會網絡分析軟體進行分析，以了解各人物間的社會關係，圖 1 為透過「中國歷代人物傳記資料庫」資料所繪製福建莆田於 1200 至 1250 年間取得進士學位者的社會網路關係 (中國歷代人物傳記資料庫，2020)；中研院數位人文研究平台內提供文本社會網絡分析功能，能讓研究者調整社會網絡圖中關聯強度、群間距離、關係距離長度；國家圖書館通用型古籍數位人文研究平台呈現明代文集中的史料人物關係圖，並透過人機互動方式，讓使用者能修正圖中的人物節點與網絡關係，進而提高史料人物關係圖的完整性及正

確性 (Chen & Chang, 2019)；「臺灣歷史人物傳記資料庫 (TBDB)」則是在，專注於臺灣人物類別、屬性與關聯資訊建置，開發人物資料檢索、勘考工具及網絡分析工具 (TBDB 臺灣歷史人物傳記資料庫，2020)，謝順宏等 (2018) 即使用此一資料庫內收錄的彰化縣人物傳記全文，配合自動分群技術探索人物的社會網絡關係，並產出關係網絡圖。綜合上述，社會網絡分析能幫助研究者釐清文本之中的人物關係、情境脈絡等，進而了解當時的時空背景，協助人文學者進行更深入的人物關係解讀。

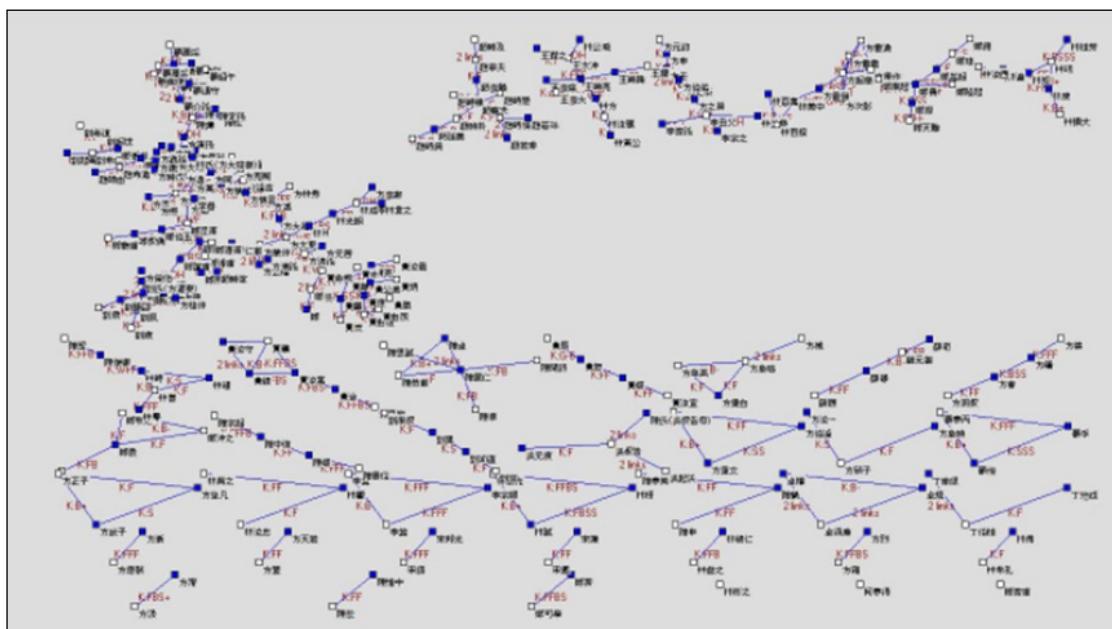


圖 1 福建莆田於 1200 至 1250 年間取得進士學位者的社會網路關係

資料來源：中國歷代人物傳記資料庫 (2020)

## 肆、馬來西亞華人發展與馬來西亞興化群賢錄

馬來西亞華人社會形成於十九世紀英殖民統治在馬來亞半島與北婆羅洲確立時期，大量華人因晚清中國國內政局動盪，被迫從南方省份經由英屬香港遠赴南洋謀生。馬來西亞華人社會網絡的形塑，最初係以草根階級（如農民、漁民）為主的移民為主，當地華人社團亦是由民間所組成，連同華文教育與華文報章被形容為馬來西亞華人社群的三大支柱。諸如興安會館等宗鄉親會館組織，除了以籍貫、地區作為地緣性組織之外，亦有以姓氏為主的血緣性會館組織。早期的會館組織部分為秘密結社組織而成，在華人移民社會初期凝聚同鄉的作用，負責建設與管理寺廟，舉辦祭祀活動，並且也設立私塾以進行教育，承繼華人文化傳統及其形式並將之紮根於南方。在英殖民政府逐漸加強對各類社團的管制以後，包含會館組織在內的華人社團也隨之適應環境的變化，在各不同時期作出相對應的調整，並且也起著代表

民間華人的訴求與當權政府的相關機構進行對話角色，以為華人爭取其權益，其中涉及包括經濟、社會和華文教育等議題。故會館組織及其人物發展軌跡，均能充份反映馬來西亞華人社會網絡是如何透過個別人物之間的個人關係、其職業身份、事件與行為之間，產生相互的關聯。

《馬來西亞興化群賢錄》一書由「馬來亞大學馬來西亞華人研究中心」與「馬來西亞興安總會」進行之馬來西亞興化人研究計畫成果之一，書中記載的八十五位興化民系人物介紹，涵蓋了政治、經濟、文教的數個領域，從他們的成長軌跡裡大多可以窺見到興化人家的文化印記，並反映馬來西亞華人當代史的各個時段發展脈絡。

## 伍、馬來西亞興化群賢錄數位人文平台

一般而言，支援數位人文研究的數位人文平台中，平台需要同時具備閱讀介面、資訊檢索、詞頻分析與資料探勘等面向功能，才能有效支援數位人文研究 (Widlöcher, 2015)。本研究發展之馬來西亞興化群賢錄數位人文平台也以上述方向為發展目標，主要開發四個輔助數位人文研究功能模組，包含資訊檢索與後設資料模組、閱讀介面模組、地理資訊模組及社會模組網絡分析模組。「馬來西亞興化群賢錄數位人文平台」採用開放原始碼的數位典藏系統 Omeka S 為基礎發展而成。Omeka S 已具有非常完整之資料的典藏、資料管理與數位策展功能，但是 Omeka S 缺乏友善、易用的閱讀介面，讓使用者難以同時閱讀掃描圖檔與對應之全文；文本檢索分析方面則對於中文語言的支援不佳，因此馬來西亞興化群賢錄數位人文平台針對 Omeka S 不足的地方進行改善及優化，並提供更進階的人物社會網絡分析功能，以輔助人文學者能藉由近讀及遠讀兩種解讀途徑，探討隱含在馬來西亞人物誌中的人物關係脈絡，讓此一平台使用上能夠更貼近人文研究者的需求。

平台首頁如圖 2 所示，系統架構如圖 3 所示，以下依序介紹各模組之功能與操作介面：



圖 2 馬來西亞興化群賢錄數位人文平台首頁

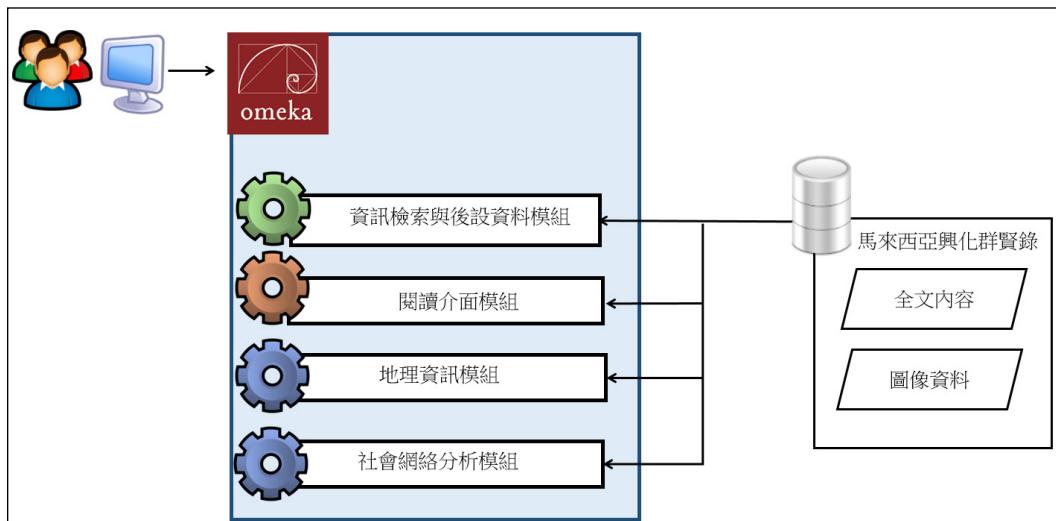


圖 3 馬來西亞興化群賢錄數位人文平台系統架構圖

## 一、資訊檢索與後設資料模組

「馬來西亞興化群賢錄數位人文平台」參照《馬來西亞興化群賢錄》一書中條目 (item) 的編排方式，係將書中每一位人物分別獨立成一個條目，以都柏林核心集作為主要後設資料欄位，如表 1 所示。資訊檢索部分針對中文語系檢索部分進行優化，將原本平台內建使用 MySQL 的檢索功能替換成 SOLR 檢索引擎，並支援簡體中文與繁體中文全文檢索和後分類分析功能，能加速使用者解讀效率，並能嘗試從不同面向分析檢索結果，目前平台提供稱

謂、關鍵字、作者等面向的檢索後分類功能，如圖 4 所示；本平台經熟悉馬來西亞文化之專家將書中的人物條目所屬之領域進行分類，建立各界群賢分類瀏覽功能，使用者可依據此一平台事先定義好之專業領域，進入該領域瀏覽相關的人物條目，目前本平台將書中的人物分為軍事、警界、法律、演藝、藝術、媒體、政治、經濟、文教、宗教等十大領域，如圖 5 所示。在資料交換上，本平台支援以 RESTful API 方式，提供 JSON、IIIF 等資料交換格式，可與其他數位典藏系統進行文字或圖像等資料交換，如圖 6 所示。後設資料模組方面，後設資料介面如圖 7 所示，點擊後設資料介面內關鍵詞，可查詢具有相同關鍵詞之條目。

表 1 後設資料範例

|      |                                                 |
|------|-------------------------------------------------|
| 標題   | 林金树                                             |
| 頭銜   | 拿督 DMSM 、AMN 、太平局紳 JP                           |
| 頁面開始 | 002                                             |
| 頁面結束 | 007                                             |
| 關鍵字  | 馬六甲、興安、興化、興安會館                                  |
| 生卒年  | 1919-2001                                       |
| 出生地  | 中國福建福清新唐鎮塘頭村                                    |
| 祖籍   | 中國福建福清新唐鎮塘頭村                                    |
| 所屬會館 | 馬來西亞興安會館                                        |
| 資料來源 | 馬來西亞興化群賢錄<br>The Malaysia Henghua Personalities |
| 撰寫者  | 林德順                                             |



圖 4 馬來西亞興化群賢錄數位人文平台檢索後分類介面



圖 5 馬來西亞興化群賢錄數位人文平台各界群賢介面

圖 6 RESTful API 格式

|                                                 |                                                              |
|-------------------------------------------------|--------------------------------------------------------------|
| 貢獻者 (Dublin Core)                               | 林德順                                                          |
| 主題 (Dublin Core)                                | 兴化                                                           |
| 時間跨度 (Dublin Core)                              | 1924-2013                                                    |
| 頁面開始 (Bibliographic Ontology)                   | 142                                                          |
| 頁面結束 (Bibliographic Ontology)                   | 147                                                          |
| 頁數 (Bibliographic Ontology)                     | 6                                                            |
| 關鍵字 (A vocabulary for biographical information) | 馬共<br>萬隆<br>中共<br>基命<br>泰國<br>抗日<br>武裝鬥爭<br>難民<br>英勇<br>和平協議 |
| 祖籍 (Friend of a Friend)                         | 福建省福清县江兜                                                     |
| 所屬會館 (Friend of a Friend)                       | 马来亚共产党                                                       |
| 引證 (Dublin Core)                                | 馬來西亞興化群賢錄 The Malaysia Henghua Personalities                 |

圖 7 後設資料介面

## 二、閱讀介面模組

「馬來西亞興化群賢錄數位人文研究平台」希望使用者在使用本平台閱讀文本的過程中，能像是閱讀真正的書籍一般便利與舒適，故採用掃描影像檔與文本共同顯示之閱讀介面，如圖 8 所示。左邊頁面為掃描之影像檔，右邊為與影像檔對應之全文，讓使用者於解讀過程中能進行圖文交互參照閱讀，可使用滑鼠滾輪或卷軸直覺性的下拉進行閱讀或頁面切換。此外，閱讀介面中也提供多項功能輔助使用者解讀文本，圖 9 為閱讀介面功能選單，排版設定支援使用者根據自身需求自由地調整字型大小、字距寬度、行距高度等；圖片設定允許使用者切換影像檔之顯示狀況；最後是本平台根據「馬來西亞興化群賢錄」之文本特性所發展的文本人名自動識別功能。



圖 8 文本閱讀介面



圖 9 文本閱讀介面功能選單

「馬來西亞興化群賢錄」書中所收錄條目以人物為主體，並記錄該人物之生平記事，本書中提及之人物在馬來西亞華人社會具有相當程度重要性，而重要人物不單僅限於書中所收錄之 85 位人士，亦有可能隱藏於描述此 85 位人物的內文之中，因此有必要從非結構化的文本中抽取潛在的人物資訊。依據以上目的，本平台特別開發文本人名自動識別功能，處理流程如圖 10 所示。首先將文本透過 BERT 將文本自然語言向量化，並透過實體命名 (NER) 辨識技術進行人物實體命名分析，將辨識出之人名，以顏色直接標註於文本介面中，點擊人物可查看該人物於其它文本中所出現之段落，進行文本內部參照，加速使用者掌握該人物於平台內的分布狀況或整體脈絡資訊，標註人物查詢如圖 11 所示。

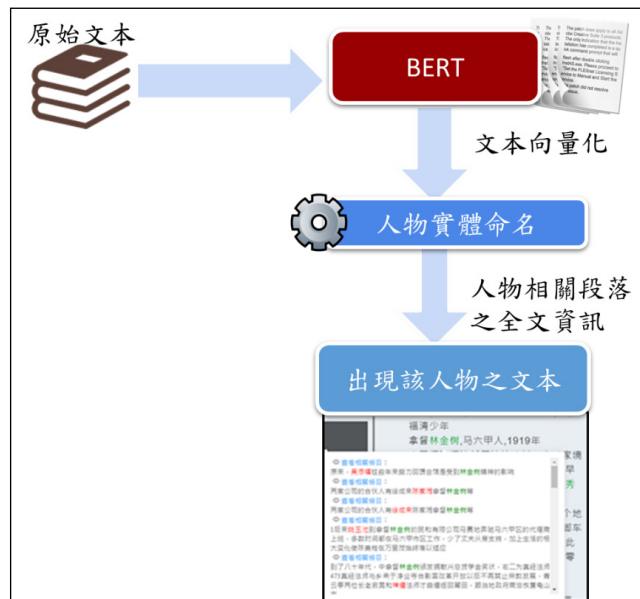


圖 10 文本人名自動識別功能流程

期,少年林金树先是到同乡长辈黄文秀

◎ 查看相關條目：  
原来, 吴添福这些年来努力回馈会馆是受到林金树精神的影响

◎ 查看相關條目：  
两家公司的合伙人有徐成来陈家湘拿督林金树等

◎ 查看相關條目：  
两家公司的合伙人有徐成来陈家湘拿督林金树等

◎ 查看相關條目：  
1后来姚玉池到拿督林金树的民和有限公司马赛地奔驰马六甲区的代理商上班，多数时间都在马六甲市区工作，少了丈夫从旁支持，加上生活的极大变化使陈美桂在万里茂始终难以适应

◎ 查看相關條目：  
到了八十年代，中拿督林金树颁发捐献兴总奖学金状，右二为真经法师473真经法师与乡亲于净业寺合影国改革开放以后不再禁止▼  
有色,无形中交际圈子也扩展了不少。

圖 11 標註人物查詢

### 三、地理資訊模組

近年廣泛應用於各個層面的地理資訊系統 (Geographic Information System, GIS)，可以打破純粹的文本分析限制，得以對文本進行「空間分析」(spatial analysis)，以利於進行基於空間資訊之觀察、發掘、思考與分析社會現象 (鄧志松、裘怡婷，2009)。「馬來西亞興化群賢錄」一書中，詳細紀載關於每一位人物詳細之地理資訊，本平台除了將人物之籍貫、出生地等資訊記錄在後設資料欄位外，也結合地理資訊系統，將該人物之地理資訊以視覺化方式呈現，如圖 12 所示。使用者亦可以自由切換不同圖層樣式進行空間資訊解讀，此一功能亦使用於群賢瀏覽功能中的各領域頁面下，可以幫助使用者快速掌握該領域人士之地理資訊分布，如圖 13 所示。



圖 12 地理資訊模組介面



圖 13 群賢瀏覽政治領域相關人物之籍貫地理資訊分布

#### 四、社會網絡分析模組

如前述段落所提及之「馬來西亞興化群賢錄」中的人物實體資訊，對於使用者解讀文本是相當重要的線索，因此如果能更進一步判斷這些人物實體之間是否具有潛在的關係屬性，

進而產生人物關係網絡，將更具輔助人物關係解讀應用價值。因此，本平台發展社會網絡分析模組，建立「馬來西亞興化群賢錄」中的人物網絡，並從人物實體間的文本內容推斷其關係屬性為何，最後以視覺化方式呈現「馬來西亞興化群賢錄」中的人物社會網絡關係圖，功能流程如圖 14 所示。本功能採用流水線 (Pipline) 方式抽取人物關係，首先與閱讀介面中的人名識別功能流程相同，將文本透過 BERT 轉換成詞向量後，進行人物實體命名辨識，作為人物關係抽取的前置作業，接著將文本內辨識出的人名實體，成對輸入本平台預先訓練好的人物關係分類器，進行人物關係預測，目前分類器將人物關係初步分為夫妻、手足、親子、朋友、師生、親戚、工作及“無關係”共八類，若預測結果非“無關係”類別，則將結果存入資料庫，並以視覺化方式呈現於介面中。社會網絡分析關係介面如圖 15 所示，圖中所呈現的資訊為該條目下抽取出的人物關係網絡，使用者可點擊網絡圖中的人物節點，直接進行人物全文搜尋，此一功能希望輔助使用者透過快速掌握整體「馬來西亞興化群賢錄」中的人物關係，發掘隱藏於其中的人物關係脈絡。

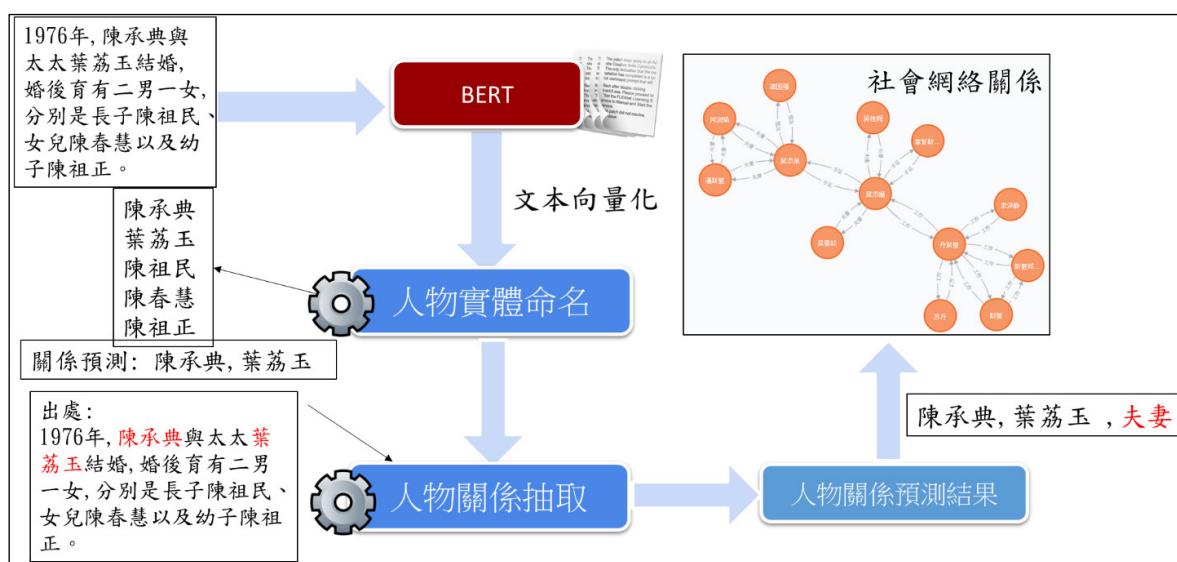


圖 14 馬來西亞興化群賢錄社會網絡分析流程圖

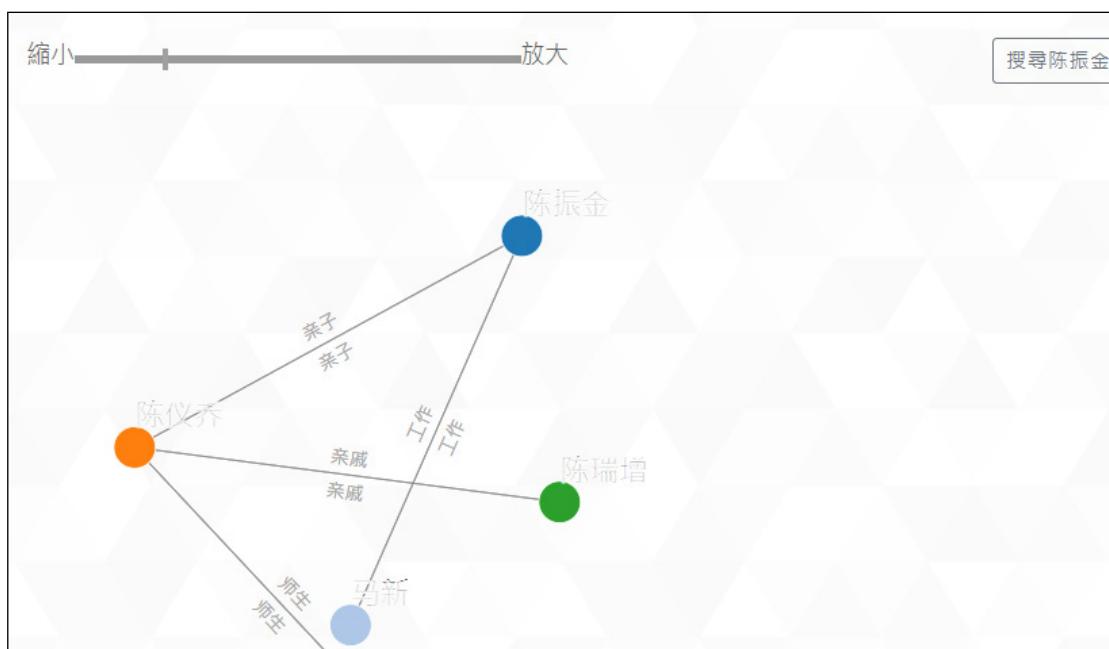


圖 15 人物社會網絡介面

## 陸、馬來西亞興化群賢錄數位人文平台研究案例分享

本節將分享其中一位研究者使用本平台進行探討馬來西亞華人發展，作為本平台解讀說明案例，該作者為某國立大學哲學系碩士學歷，研究興趣為馬來西亞華人政治發展探討，以下簡述該研究者之解讀過程。

該研究者首先透過平台中的各界群賢的分類功能進入軍事分類，該分類下有陳平、涂耐冰等人，使用者閱讀文本內容發現陳平與涂耐冰二人都與抗日事業和中共有關。

從陳平條目下人物關係社會網絡圖，研究者知悉陳平、馬共與中共領袖鄧小平之間存在一定關聯，如圖 16 所示，經細讀文本及外部查詢相關得知，陳平為馬來西亞共產黨之重要人物，並擔任過馬共總書記。約 1960 年代，毛澤東掌權時期，陳平受到時任總書記的鄧小平鼓勵，給予資金及武裝援助，對抗馬來西亞政府，而至約 1980 年代，鄧小平掌權，因「戰略轉移」為由，宣布停止輸出革命，停止對東南亞共產游擊隊的一切支援，馬共最後缺乏中共的支援和國際冷戰結束之際，於 1989 年停止所有軍事行動和平解散，陳平也流亡海外。從陳平的條目中，可以一窺中共對馬共從事游擊戰的支持到轉而冷漠之過程。

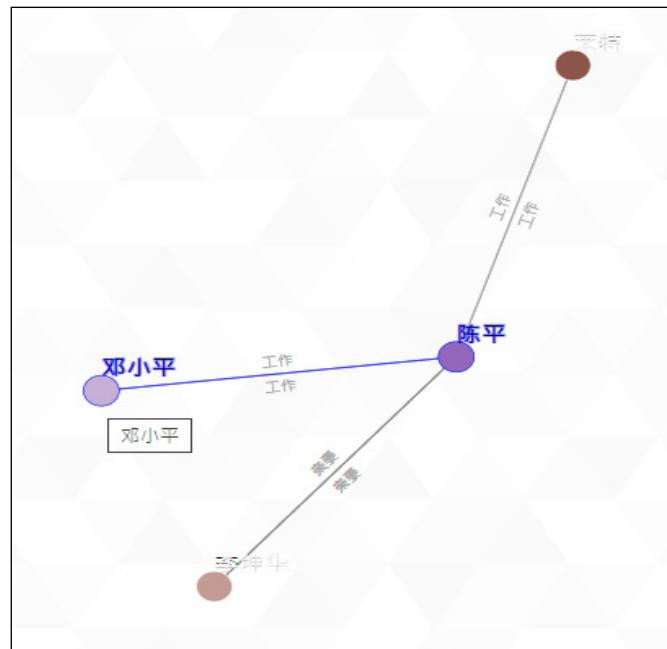


圖 16 陳平與鄧小平之人物關係網絡

至於涂耐冰，研究者一開始先透過人物社會網絡圖，發現他與二戰前後時期著名的陳嘉庚有關聯，如圖 17 所示，從涂耐冰條目的全文得知，1938 年時任南洋總會陳嘉庚會長、涂耐冰則是擔任南洋總會的秘書長一職。此外人物社會網絡圖中出現一位名叫黃慶昌人士，和陳嘉庚、涂耐冰相連，經細讀文本後，1939 年，陳嘉庚發出南洋總會第六號公告，號召成立南僑機工團回大陸服務，主要的計畫執行就是涂耐冰、黃慶昌兩人，而兩人之間的關係於砂勞越日報的創刊及古晉興安會館的成立上亦有重要的影響力。

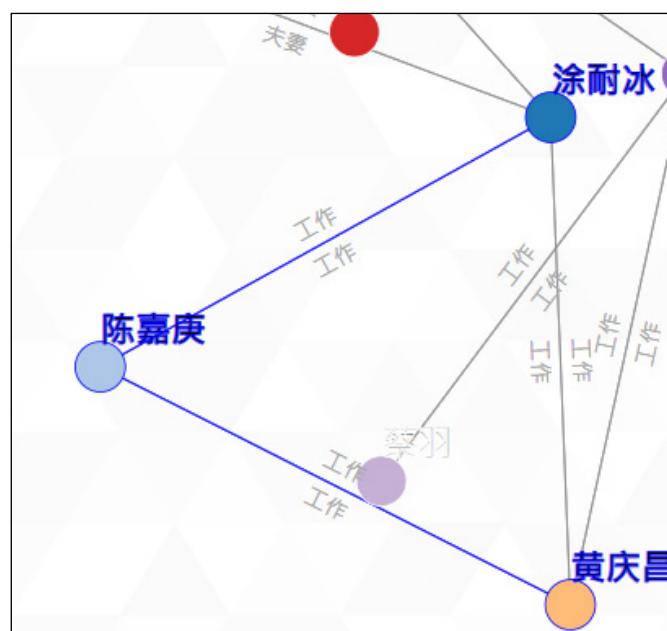


圖 17 涂耐冰、陳嘉庚與黃慶昌之人物關係網絡

透過此研究案例可以看出本平台所提供之功能，能幫助研究者找出隱藏於文本中之脈絡，如黃慶昌雖不是條目中的人物，但經由社會網絡分析功能，發現此人在南洋抗日的歷史上，也是相當重要的角色。而本平台也有助於研究者從南洋的角度出發，從各方面探討南洋華人發展與近代中國之間的關係。

除上述研究案例外，仍有許多潛在的研究議題，從本地研究的角度，可以關注個別籍貫人士所從事的經濟活動進行彙整，以對其他籍貫的人士做比較，這涉及馬來西亞獨立前後華人經濟活動與職業分配。如腳踏車行業、礦業等，多數商業活動都是由華人所從事。宗教方面，多數禪師活躍於清末民初時期下南洋，對各位禪師活動經歷的觀察與整理，透過地理資訊模組及社會網絡分析模組輔助，也許可梳理出佛教在南洋傳承與發展的路徑，而陳秉忠傳教士的帶領大批興化地區教友經新加坡至馬來西亞的詩巫地區開墾作業，及作為新馬地區移民的先驅者之一，他如何於扎根並融入當地，這些都是值得探索之議題。

## 柒、結論與未來方向

「馬來西亞興化群賢錄數位人文研究平台」除了提供友善的閱讀介面、輔助文本解讀工具，亦可藉由地理資訊模組或社會網絡分析模組的輔助，從宏觀的角度，以兼具時空及人物關係方式探索文本與人物關係脈絡。此外，本文亦透過平台研究案例分享提出應用此一平台進行人物關係之可能解讀模式，也提出幾點可進一步探討的馬來西亞華人人物研究議題。未來將持續蒐集、擴充馬來西亞華人相關史料，也進一步強化以下平台之功能發展：

### 一、鏈結資料形式串聯平台內部及外部資源

若能將數位人文平台中內部所典藏龐大、複雜、分散且具關聯的文本內容進行相互鏈結，並從中計算與該文本具有相似概念的內容資訊，將具有較高相似度的文本內容以鏈結方式，呈現給使用者，亦是輔助數位人文研究之重要技術發展議題。若能如此，不僅能幫助使用者取用相關聯的文本資訊，亦能節省其尋找關聯資訊所花費的時間。

此外，透過研究解讀案例可以發現，人文研究者在解讀資料時，時常需要參照其他的相關資源，以協助其理解文本中的脈絡或進行驗證。因此，未來平台將整合 Chen 和 Liu(2019) 所發展之自動標註工具，以鏈結資料 (Linked Data) 方式，自動的從多個外部來源資料庫取得資料，讓使用者在閱讀時，能即時參考針對閱讀文本之不同來源補充資訊，讓使用者更快速理解文本的內容。

## 二、知識圖譜之建立

在大量的史料與文本內容中，往往蘊含著複雜的實體網絡關係，這些實體可能包含人物、機構、事件與地點等。在過去研究者欲釐清其中的脈絡，往往需要耗費極大的時間與精力。目前本平台只初步進行的人物與人物關係之間的抽取，未來預計將實體命名的類型拓展至機構、事件、時間等，並抽取各實體彼此之間的潛在關係，整合多維的網絡關係，並透過資訊視覺化技術建立完善的知識圖譜，並且隨著文本中時序的推移，知識圖譜中所呈現的內容也會有所變化。

## 三、具個人化之平台收藏庫與分析功能

本平台預計發展個人化之平台收藏庫與分析功能，開發我的收藏庫 (My Collection) 功能，克服文本內容與工具綁死之限制，允許使用者將感興趣之條目加入個人收藏庫，同時支援上傳自身所典藏之文本，結合平台本身已具有的豐富內容，以及使用者個人擁有的內容，建立個人化的文獻集，並可根據自身需求，佈建合適分析個人收藏庫的數位分析工具，如此將能發展符合出人文學者需求且具彈性之數位人文平台。

## 致謝

此平台建置係由教育部高教深耕計畫特色領域中心之「華人文化主體性研究中心」經費補助 ( 經費代碼 109H21)

## 參考文獻

中央研究院數位文化中心 (2020)。數位人文研究平台。上網日期：109 年 10 月 11 日，檢自：  
<https://dh.ascdc.sinica.edu.tw/member/>

中國歷代人物傳記資料庫。CBDB querying and reporting system -online。上網日期：109 年 9 月 18 日，檢自：<http://db1.ihp.sinica.edu.tw/cbdb/help/systemintro.html>

杜協昌 (2018)。DocuSky：個人文字資料庫的建構與分析平臺。數位典藏與數位人文 , 2, 71-90。

金觀濤、邱偉雲、梁穎誼、陳柏聿、沈錦坤、劉青峰 (2016)。觀念群變化的數位人文研究 -- 以《新青年》為例。數位人文：在過去、現在和未來之間。

翁稷安 (2016)。從「機構導向資料庫」到「個人導向資料庫」：數位人文下一階段的可能發展。第七屆數位典藏與數位人文國際研討會，臺北。

陳志銘、張鐘、徐志帆 (2020)。羅家倫先生文存數位人文研究平臺之建置與應用。數位典藏與數位人文 , 5 , 73-115。

項潔、涂豐恩 (2011)。導論—什麼是數位人文。從保存到創造：開啟數位人文研究，9–28。

趙薇 (2016)。「社會網路分析」在現代漢語歷史小說研究中的應用初探——以李劫人的《大波》三部曲為例。數位人文：在過去、現在和未來之間。

劉吉軒、柯雲娥、張惠真、譚修雯、黃瑞期、甯格致 (2012)。以文本分析呈現海外史料政治思想輪廓。數位人文要義：尋找類型與軌跡。頁 83–114。

鄧志松、裘怡婷 (2009)。GIS 中國經社指標網路資料庫：建置與推廣。地理資訊系統季刊，3(4)，15–22。

謝博宇 (2016)。以 DocuSky 為核心的工具開發與建置。第七屆數位典藏與數位人文國際研討會，臺北。

謝順宏、柯皓仁、張素玢 (2018)。自動分群應用於傳記人物關係建立。「第九屆數位典藏與數位人文國際研討會」發表之論文，法鼓文理學院。

Aparecido Ribeiro, M., Vosgerau, R. A., Larissa Pereira Andruchiw, M., & Ely de Souza Pinto, S. (2015). The complex social network from The Lord of The Rings. ArXiv E-Prints, 1606, arXiv:1606.02610.

Chen, C. M. & Chang, C. (2019). A Chinese ancient book digital humanities research platform to support digital humanities research. *The Electronic Library*, 37(2), 314–336.

Chen, C. M., Chen, Y. T., & Liu, C. Y. (2019). Development and evaluation of an automatic text annotation system for supporting digital humanities research. *Library Hi Tech*, 37(3), 436–455.

Chen, K. C., & Chen, B. Y., & Huang, C. C. (2018). Who was the key figure? A social network analysis of Suzhou book collector groups in Ming Dynasty through a digital humanities approach. *Journal of Educational Media & Library Sciences*, 55(3), 243–284.

Correll, M., Alexander, E., Albers, D., Sarikaya, A., Gleicher, M. (2014). Navigating reductionism and holism in evaluation. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pp. 23–26.

Drucker, J. (2013). Intro to digital humanities: Introduction. UCLA Center for Digital Humanities. Web available at [http://dh101.humanities.ucla.edu/?page\\_id=13](http://dh101.humanities.ucla.edu/?page_id=13), Retrieved January 28, 2018.

Duling, D.C.(2000). The Jesus Movement and Social Network Analysis: (Part II. The Social Network). *Biblical Theology Bulletin*. Vol 30, Issue 1, pp. 3 -14

Franco Moretti. (2011). Network theory, plot analysis. *New Left Review*.

Ho, H. I. B. & Hilde, D. W. (2014). MARKUS text analysis and reading platform. Retrieved September 11, 2020, from <https://dh.chinese-empires.eu/markus/beta/>

Hockey, S. (2004). The history of humanities computing. *A Companion to Digital Humanities*, 3-19.

Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2017). Visual text analysis in digital humanities: Visual text analysis in digital humanities. *Computer Graphics Forum*, 36(6), 226–

250.

Kestemont, M. and Gussem, J.D. (2017). Integrated sequence tagging for medieval Latin using deep representation learning. *Journal of Data Mining and Digital Humanities, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages*. available at: <https://jmdh.episciences.org/3835>

Moretti, G., Sprugnoli, R., Menini, S., & Tonelli, S. (2016). ALCIDE: Extracting and visualising content from large document collections to support humanities studies. *Knowledge-Based Systems*, 111, 100-112.

So, R. J. & Hoyt L. (2013). Network Analysis and the Sociology of Modernity, *boundary 2*, 40(2), 147-182.

Stanford University (2013). Retrieved September 16, 2020, from <http://republicofletters.stanford.edu/>

Sun, Z. & Wang, H. (2010). Overview on the advance of the research on named entity recognition. *New Technology of Library and Information Service*, 26, 6, 42-47.

TBDB 臺灣歷史人物傳記資料庫 (2020)。TBDB 臺灣歷史人物傳記資料庫。上網日期：109 年 9 月 11 日，檢自：<http://tbdb.ntnu.edu.tw/index.jsp#about>

Tu, A., Hung, J.-J., & Lin, Y.-H. (2012). Building a text analysis platform for Chinese Buddhist text -An example based on CBETA and Tripitaka Catalog Projects. Presented at the PNC 2012 Annual Conference and Joint Meetings, UC Berkeley School of Information, USA. Retrieved from <http://joeyhung.info/publications/>

Yohan, P.M., Sasidhar, B., Basha, S., A.H. and Govardhan, A. (2014). Automatic named entity identification and classification using heuristic based approach for Telugu. *International Journal of Computer Science Issues*, Vol. 11 No. 1, pp. 173-180.