

人物資料庫的建置：以「臺灣歷史人物傳記資料庫」為例

The Taiwan Biographical Database (TBDB): An Introduction

張素玢 Su-bing Chang
國立臺灣師範大學臺灣史研究所
Graduate Institute of Taiwan History
National Taiwan Normal University, Taiwan
109682@ntnu.edu.tw

摘要

人物是歷史學研究的重要基礎，舉凡人物的個性、家庭背景、經歷、社會階層，甚至於整個社會的階層流動、婚姻與政治網絡等都是歷史研究的議題。「中國歷史人物傳記資料庫 (The China Biographical Database, CBDB)」可謂目前與歷史人物研究相關最具規模的資料庫之一。目前臺灣已有的人物傳記資料多以數位文件的形式存在，而須用人力從大量人物傳記中爬梳、彙整資料。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，並運用前述工具協助歷史人文學者進行研究。

臺灣歷史人物傳記資料庫 (Taiwan Biographical Database, TBDB) 的建置計畫，源自中國歷代人物傳記資料庫 (China Biographical Database, CBDB) 的啟發，期望能提供臺灣歷史人物傳記資料的文本探勘、社會網絡分析工具及相關軟體服務，梳理錯綜複雜的人際網絡，開拓臺灣史的研究視野。基於建置者對資料的掌握度及著作權的取得，建置團隊首先選擇四位人文學者成員共同編纂的《新修彰化縣志·人物志》做為文本來源，其體例統一、考訂嚴謹，含括各時代、各類型人物，是 TBDB 良好的試金石。除了借鏡 CBDB，建置 TBDB 的重點在於依照臺灣的時空環境，設計符合臺灣歷史的人物屬性類別。目前初步完成檢索服務、人物地域分布及社群關係分析等功能，開放學界利用以除錯優化。目前資料內容已從彰化縣人物擴及臺北市、澎湖縣、臺中市、南投縣等地區。

本演講主旨在於描述「臺灣歷史人物資料庫」現階段所收錄之人物特性，闡述系統架構以及初步成果。此外，也說明建置「臺灣歷史人物資料庫」過程中的經驗和未來發展方向。

Abstract

Personage is an important kind of entities in the study of history. Comprehensive understanding of personage biographies is beneficial for researching into historical events. In the digital era, many personage biographies are available in digital formats; as a result, it is time-consuming and labor-intensive for researchers to explore invaluable findings from massive personage biographies. Facing this situation, researchers may be helped to utilize the information efficiently with information technologies.

The idea of developing Taiwan Biographical Database (TBDB) was inspired by China Biographical Database (CBDB) project, and aims to provide digital tools, such as text mining, social network analysis, and other related tools, for analyzing complex social networks and broaden research visions in the study of Taiwanese history. Considering the participants' knowledge of texts and the ability to acquire copyright, the team chooses the "Treatise of Historical Figures" of the New Edition of Changhua Local Gazetteer, compiled by four of the team members, as the base text. There are several advantages of this choice. First, the biographies in the Treatise are written according to the same format; second, the biographies are rigorously verified; third, they include people from all periods and all walks of life. Therefore, they should be a good starting point for the TBDB project. Besides learning from CBDB's model, the emphasis in developing TBDB is to build attributes for historical figures that are suited for Taiwan's historical contexts. Currently, we have built initial versions of basic search functions, a map which shows the geographical distribution of historical figures, and social network analysis tools. They are now open to the public in order to optimize and debug. In the future, we will continue to increase both the quality and quantity of the database and also develop new analysis tools.

This speech introduces the development of a text retrieval and mining system for Taiwanese historical people -- Taiwan Biographical Database (TBDB). It describes the characteristics of personages in TBDB, highlights the system architecture and preliminary achievement of TBDB. Finally, this talk elaborates on the lessons learned through the creation of TBDB, and the future plans.

致謝

本研究為科技部研究計畫的部分成果，計畫編號為：MOST 106-2420-H-003-010 和 MOST 105-2420-H-003 -016，上述整合型計畫的參與學者為柯浩仁、謝順宏、李宗翰、李毓嵐、李昭容、顧雅文等，共同建置了「臺灣歷史人物傳記資料庫」(TBDB)的基礎，特此致謝。

參考文獻

- 張素玢, 2018.04, 〈以《新修彰化縣志·人物志》建置臺灣歷史人物傳記資料庫(TBDB)的構想和做法〉, 《彰化文獻》22, 彰化:彰化文化局, 頁 15-31。
- 張素玢、李宗翰、李毓嵐、李昭容、顧雅文、柯皓仁、謝順宏, 2018.10, 〈從 CBDB 到 TBDB: 以《新修彰化縣志·人物志》為試金石〉, 《數位典藏與數位人文》2, 頁 91-115。
- 謝順宏、張素玢, 2018.06, 〈臺灣歷史人物文本檢索與探勘系統之建置〉, 《圖資與檔案學刊》92, 頁 67-87。
- Bol, P. K., Hsiang, J., & Fong, G. (2012). Prosopographical databases, text-mining, GIS and system interoperability for Chinese history and literature. In *Proceedings of the 2012 International Conference on Digital Humanities*.
- Digital Humanities Network (n.d.) *About*. Retrieved from <https://www.digitalhumanities.cam.ac.uk/about/about-page>.
- Fuller, M. A. (2015). *The China Biographical Database User's Guide*, Revised Version 2.0. Retrieved from https://projects.iq.harvard.edu/files/cbdb/files/cbdb_users_guide.pdf.
- Harvard University (n.d.). *Home. China Biographical Database Project (CBDB)*. Retrieved from <https://projects.iq.harvard.edu/cbdb/home>.
- Liu, C. L., Huang, C. K., Wang, H., & Bol, P. K. (2015). Toward Algorithmic Discovery of Biographical Information in Local Gazetteers of Ancient China. In *29th Pacific Asia Conference on Language, Information and Computation (PACLIC 29)*, Shanghai, China, October 30 – November 1, 2015.
- Martin, L. (2016). The university library and digital scholarship: A review of the literature. In Mackenzie, A. & Martin, L. (ed.) (2016). *Developing Digital Scholarship: Emerging Practices in Academic Libraries*. Facet Publishing.
- Sie, S. H., Ke, H. R., & Chang, S. B. (2017, November). Development of a text retrieval and mining system for Taiwanese historical people. In *Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC)*, 2017 (pp. 56-62). IEEE.