
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202304.00656

Development and Application of a Thematic Database for World Chinese Studies: A Case Study from East China Normal University (Postprint)

Authors: Deng Luxiang, Xu Xin

Date: 2023-04-01T16:02:48+00:00

Abstract

[Purpose/Significance] Through the construction of a specialized database, this study addresses issues in contemporary global Chinese Studies, including the difficulty of accessing extraterritorial knowledge about China, insufficient attention to public opinion, and neglect of interpersonal relationships. [Method/Process] From the stages of data collection, processing and integration, and retrieval and utilization, this paper discusses the design scheme of the global Chinese Studies specialized database, and takes the East China Normal University Global Chinese Studies Database as an example to introduce its data collection, processing and integration, and retrieval and utilization. [Result/Conclusion] Through the construction of the global Chinese Studies database, it can promote the discovery of important figures, important institutions, important research fields, and interpersonal relationships in extraterritorial China studies. Furthermore, the database provides a data foundation for scholars to conduct deeper and more multi-dimensional related research.

Full Text

Construction and Application of the World China Studies Thematic Database: A Case Study from East China Normal University

Deng Luxiang¹, Xu Xin^{1,2}

¹Social Survey and Data Center, East China Normal University, Shanghai 200241

²Faculty of Economics and Management, East China Normal University, Shanghai 200062

Abstract:

[Purpose/Significance] This study aims to address persistent challenges in contemporary World China Studies by constructing a thematic database, including difficulties in accessing overseas knowledge about China, insufficient attention to public opinion, and neglect of inter-scholar relationships.

[Method/Process] The paper discusses the design of the World China Studies thematic database across three stages: data acquisition, processing and integration, and retrieval and utilization. Using the East China Normal University World China Studies Database as a case study, it introduces the specific implementation of data collection, processing, integration, and retrieval.

[Result/Conclusion] The construction of the World China Studies thematic database facilitates the discovery of important scholars, institutions, and research fields in overseas China studies, as well as relationships among scholars. Furthermore, the database provides a data foundation for scholars to conduct deeper and more multidimensional research.

Keywords: World China Studies; Thematic Database; Think Tank; Public Opinion

Classification Number: G250

DOI: 10.13266/j.issn.0252-3116.2021.07.005

In recent years, China's rapid development has attracted worldwide attention, with the international community's focus on China continuing to grow. This has led to a surge in global knowledge production about China, as foreign media, scholars, and institutions increasingly study Chinese issues ranging from macro-level social, economic, and political topics to specific aspects of daily life in particular historical periods. These research efforts aim to acquire important knowledge and intelligence about China to gain advantages in cooperation and competition with China. However, due to information asymmetry, ideological differences, political confrontation, and other factors, these research outcomes often carry biases that damage China's international image and hinder the creation of a friendly international development environment. This problem became particularly prominent during the COVID-19 pandemic: while China achieved major successes in fighting the epidemic and could provide valuable experience and important support for other countries, Western societies produced harsh voices, with some politicians and media ignoring facts and deliberately smearing China and its anti-epidemic experience, hoping to launch a new round of "Cold War" to contain China's development. Against this backdrop, systematically analyzing and studying global knowledge about China—knowing both ourselves and our opponents—is essential to avoid being constrained by others in international cooperation and confrontation and to safeguard our interests.

1 Overview of World China Studies Research

Foreign scholars' research on China primarily encompasses two themes: (1) studies of Chinese history, culture, and language, traditionally called "Sinology"; and (2) studies of contemporary Chinese issues, covering politics, economics,

military, society, and other aspects, known as “China Studies” [1]. “Sinology” was once the mainstream of overseas Chinese studies, but this situation was overturned after World War II. With the founding and rise of the People’s Republic of China, research on contemporary Chinese issues has gained increasing attention from scholars worldwide, evolving from academic concerns into important topics in political, diplomatic, military, and economic fields, with research outcomes significantly influencing policymaking in their respective countries. Foreign scholars’ research on China provides new ideas, methods, and findings for domestic academic research on related issues. However, it must be acknowledged that among these research outcomes exist many questionable, far-fetched, or even erroneous and deliberately distorted studies. For valuable new ideas, methods, and findings, we should adopt their strengths to solve domestic problems; for questions, we need timely responses to clarify doubts; for attacks, we require precise and forceful counterarguments. This requires domestic scholars to pay attention to overseas research on Chinese issues and understand its dynamic progress and viewpoints.

Regarding the study of overseas Chinese research, academia has various terms including Overseas Chinese Studies, Foreign Chinese Studies, Extraterritorial Chinese Studies, International Chinese Studies, and World Chinese Studies. Zhu Zhengmin proposed that Overseas Chinese Studies refers to foreign scholars’ comprehensive research on China, covering not only Chinese history but also literature, philosophy, politics, economics, culture, society, military, and other aspects [1]. Liang Yi defined Foreign Chinese Studies as research by foreign scholars on modern and contemporary Chinese issues in the fields of philosophy and social sciences [2]. The concepts of Extraterritorial Chinese Studies and International Chinese Studies are similar to the former two. In recent years, the Shanghai Academy of Social Sciences has proposed the concept of “World Chinese Studies,” defining it as comprehensive research on China conducted both overseas and domestically [3], covering the aforementioned concepts of “Sinology” and “China Studies” while also including indigenous Chinese studies. World Chinese Studies has gradually developed into an emerging discipline aimed at promoting interaction between overseas and domestic research to truly help build a Chinese discourse and knowledge system. Correspondingly, World China Studies refers to domestic counter-research on global Chinese studies.

In the traditional “Sinology” field, Zhu Zhengmin introduced Paul Cohen’s formation of the “China-centered approach” in Chinese historical research and noted Cohen’s observation that Chinese scholars’ counter-research appears asymmetrical compared to the scale of overseas Chinese studies [4]. Yan Shaodang analyzed numerous Japanese Chinese studies scholars and their works, arguing that Chinese researchers should reflect on foreign researchers’ cultural contexts, literary historiography concepts, and textual authenticity, and adjust their academic perspectives [5]. Wu Yuanyuan, based on previous research, divided Chinese scholars’ attitudes toward overseas Sinology into four stages: contempt and praise based on academic confidence in the late Qing and early Republican period; adoption and correction after the 1920s; criticism of

implicit ideologies and methodologies after 1949; and blind worship of the West after the 1980s [6]. These studies show that research on traditional overseas “Sinology” has primarily focused on specific researchers and their works using close reading and in-depth analysis—traditional humanities research methods.

As contemporary Chinese issues have gained increasing attention in other countries, domestic scholars have gradually recognized the importance of counter-research on overseas studies. Zhu Zhengmin noted that John Fairbank’s establishment of modern American Chinese studies marked not only an important turning point in American Chinese research but also a watershed for global Chinese studies [1]. Wei Haisheng, Tang Zhouyan, and Zhang Shuhua argued that telling China’s stories well and disseminating China’s voice can influence overseas Chinese studies scholars and thereby affect the world’s correct understanding of China [7]. Liang Yi pointed out in her research that in overseas Chinese studies, think tanks have gradually replaced university research institutions in influencing governments [2].

The emphasis on contemporary Chinese issues has led to significant changes in the content and subjects of World China Studies. Research content has shifted from focusing on history and literature to emphasizing politics, military, diplomacy, society, and economics. Research subjects have transformed from being scholar-dominated to involving joint participation by government, military, business, media, and the public. Moreover, the countries where research subjects are located have expanded from developed countries and neighboring countries like the UK, US, France, and Japan to a broader range.

2 Challenges in World China Studies Research

Despite the massive volume of global Chinese studies and knowledge production, domestic scholars conducting counter-research still face the following dilemmas:

2.1 Limited Recognition of Research Importance and Value

Currently, nearly 100 countries have established considerable-scale Chinese studies research institutions worldwide. These institutions monitor China’s dynamics through various channels to serve academic research and national decision-making. However, domestic scholars’ counter-research on overseas Chinese studies remains more introductory than analytical, and more translational than research-oriented [8]. This has caused some scholars to unconsciously become transmitters of Western academic perspectives, losing academic autonomy and independence. This contradicts our vision of building a discourse system with Chinese characteristics. The root cause is the failure to truly recognize the significant value of tracking and assessing overseas Chinese research dynamics. Academically, World China Studies can draw on excellent foreign research methods and findings to promote domestic research on related issues, identify questions and errors, and provide targeted clarifications. In terms of serving national decision-making, World China Studies is a state-level endeavor and

a prerequisite for mastering the initiative in setting major international agendas, building a Chinese discourse system, and seeking consensus for common development, holding important significance.

2.2 Difficulty in Obtaining Comprehensive and Timely Overseas Chinese Knowledge

Most domestic World China Studies research revolves around textual interpretation of specific individuals, works, statements, or news reports. Against the backdrop of rapidly spreading information in the internet age, this approach faces two major problems: first, the inability to obtain the latest information in a timely manner, compromising the timeliness of research content and preventing timely, forceful responses to overseas questions, erroneous viewpoints, and accusations against China, which inevitably causes adverse effects; second, the difficulty in obtaining comprehensive overseas Chinese knowledge, making it impossible to construct a systematic and complete panoramic view of overseas perceptions, leading to potentially generalistic research conclusions that cannot guarantee academic objectivity or effective decision-making support.

2.3 Insufficient Attention to Public Opinion

Think tanks play important functions in providing internal policy consultation and advice, guiding media and the public, and externally expanding national influence and discourse power. The public opinion they reflect significantly impacts government governance, socioeconomics, and military diplomacy in various countries. Meanwhile, news media and social networks carry public opinion characterized by wide dissemination and broad influence. With accelerated globalization, foreign public opinion views and perspectives on Chinese issues discussed and disseminated in international political confrontation and exchange have become increasingly influential, affecting everyone from ordinary foreign citizens to government decision-makers. Notably, domestic World China Studies research still primarily emphasizes the significance of academic research, with less research serving practical needs, particularly regarding overseas think tank outcomes, which does not match the current significant influence of think tanks.

2.4 Neglect of Inter-Scholar Relationships in Overseas Studies

Among scholars conducting Chinese studies overseas, some wield considerable influence, such as John Fairbank, Ezra Vogel, Benjamin Schwartz, Paul Cohen, Yoshihiro Murayama, and Kōjirō Yoshikawa. These scholars significantly impact academic research and political decision-making in their countries regarding China, and many maintain close relationships—for instance, Paul Cohen served as a researcher at Harvard University's Fairbank Center for Chinese Studies and was deeply influenced by Benjamin Schwartz. The original intention of conducting World China Studies, beyond learning from perspectives and supplementing deficiencies, includes engaging in dialogue and exchange on issues of concern in overseas Chinese studies to guide correct understanding of

China. This requires us to not only identify key figures in overseas Chinese studies but also find pathways to transmit China's voice to these key individuals, ultimately influencing decision-making and public perception in their countries through these key figures.

Against the backdrop of rapidly changing international situations and recurring anti-China waves, the severe problems listed above urgently require effective solutions to prevent disadvantages in international cooperation and competition due to inadequate information access.

3 Design of the World China Studies Thematic Database

3.1 Construction Significance

The advent of the big data era has greatly enriched the information resources available for World China Studies. However, big data is also characterized by large scale and rapid production, diverse types, uneven quality, and low overall value density, making traditional information methods and technologies inadequate for retrieving, processing, and analyzing massive heterogeneous data. Consequently, "data intelligence" based on intelligent technologies for processing, analyzing, and mining massive data has emerged. It should be noted that learning and mastering "data intelligence" methods and technologies is relatively difficult, causing many World China Studies researchers to feel daunted when facing big data. This is precisely where thematic databases prove valuable.

Thematic databases are specialized data resource integrations constructed for specific disciplines, fields, individuals, events, or other objects [9]. They represent innovative means and important pathways for developing and utilizing humanities and social sciences data resources, playing significant roles in providing academic materials, supporting strategic decision-making, and inheriting traditional culture [10]. Characterized by concentrated information, in-depth content, and high value density, thematic databases cater to the specific needs of particular disciplines, fields, individuals, or events and their researchers. The World China Studies thematic database collects, analyzes, processes, stores, and displays various research outcomes on China from around the world. Its construction helps scholars more conveniently access comprehensive data resources within this thematic area to conduct research, representing both recognition of the importance and value of research in this field and an important measure to promote its healthy development. On one hand, scholars can conduct research directly based on resources in the thematic database, eliminating the pain of acquiring and integrating massive multi-source heterogeneous data resources. To address insufficient attention to public opinion, integrating think tank, news media, and social network information resources into the World China Studies thematic database can fill this gap. To remedy neglect of inter-scholar relationships, the database construction can specifically focus on individuals, discover and visually display relationships among scholars, help plan pathways for transmitting China's voice based on these relationships, and support precise targeting

or forceful responses to public opinion crucial for national decision-making. Additionally, comprehensive coverage of overseas Chinese studies knowledge can help scholars from different disciplinary fields join World China Studies research, bringing new interpretations and utilization methods to domestic academia.

3.2 Construction Approach and Key Challenges

The construction of humanities and social sciences thematic databases typically involves multiple stages including resource acquisition, processing and integration, and retrieval and utilization [11]. The construction of the World China Studies database needs to address current pain points and challenges in World China Studies research at these essential stages.

3.2.1 Data Acquisition Data resources are the core of the database. Scientifically and reasonably selecting data sources and determining acquisition strategies determines whether data resources can effectively support various counter-research efforts in the World China Studies field. The emphasis on contemporary Chinese issues requires that the World China Studies thematic database's data resources include not only academic research outcomes but also focus on information resources from think tanks, news media, and social networks. Academic research outcomes are typically distributed in various academic databases with rich and standardized data fields. Information from think tanks and news media is widely distributed and highly heterogeneous. Social network information usually clusters on a few major platforms and features brevity. A single acquisition strategy cannot accommodate these widely sourced and highly differentiated information resources, making data acquisition one of the key challenges in World China Studies thematic database construction.

Therefore, different data acquisition strategies need to be formulated for different categories of information. For academic research outcomes, effective search terms can first be selected for retrieval in academic databases, followed by data crawling using web scraping tools based on search results. For think tank and news media information, a combination of general engines and focused crawlers can be adopted: first, appropriate keywords are selected to form a thesaurus, general search engines' breadth-first crawling strategies are used to collect specific internet information, and then vertical search focused crawlers are employed for deep crawling of high-priority authoritative websites to achieve effective control over information acquisition scope and direction. For social network information, depth-first search methods can be used: first collecting post or topic titles, links, authors, and other information, then entering posts or topics to collect reply information, links, authors, and subsequently entering replies to collect nested replies and authors, thus traversing all collectible relevant data.

3.2.2 Data Processing and Integration The data processing and integration of the World China Studies thematic database includes data cleaning and data aggregation.

Key aspects of data cleaning include deduplication, cleaning of garbled fields and formats, and alias unification. One of the purposes and original intentions of World China Studies research is to engage in dialogue and exchange on key issues of concern in overseas Chinese studies. As humans are the subjects of dialogue and exchange, they are particularly important. During data cleaning, special attention should be paid to handling cases where one person has multiple names. English name organization forms are diverse, including full names, names without middle names, middle name initials only, abbreviations of all but the surname, and separators (dots or hyphens) between surname and given name. How to clean and present personal name data is a challenge. Therefore, during data processing and integration, personal names need to be unified, mapping multiple aliases of the same person.

Currently, data resource aggregation is generally transitioning from external feature-based to content-based aggregation, from coarse-grained to fine-grained aggregation, and from single-dimensional to multi-dimensional aggregation [12]. Numerous technologies and methods are available, including thesauri, ontologies, linked data, bibliometrics, folksonomy, and social network analysis [13], with some scholars proposing new methods such as knowledge supernetwork theory-based semantic aggregation models for data resources [12]. These methods have their own advantages and disadvantages and need to be selected based on specific resource conditions. Moreover, World China Studies research focuses on three key elements: scholars, issues, and viewpoints. How to integrate diverse types and structurally different information into the database and effectively present information related to scholars, issues, and viewpoints is key to World China Studies thematic database construction. Considering the importance of humans as the subjects who raise questions and express viewpoints, data integration should focus on individuals, associating, organizing, and presenting individuals with their relevant actions (such as publishing academic papers and think tank articles, social network posts and replies) to facilitate subsequent research on identifying issues and refining viewpoints.

3.2.3 Data Retrieval and Utilization Retrieval effectiveness is mainly reflected in accuracy and efficiency. Traditional keyword-matching retrieval suffers from problems such as “faithful expression,” “expression differences,” and “vocabulary isolation” [14], affecting retrieval result accuracy. Moreover, lack of inference during retrieval increases the difficulty for users to acquire knowledge through search, reducing retrieval efficiency. To address this, many scholars have proposed semantic retrieval concepts, such as using semantic technologies like entity recognition, entity classification, relationship extraction, knowledge storage, and knowledge reasoning to achieve flexible semantic retrieval [11], which can effectively help users quickly obtain needed resources.

In terms of utilization, relatively superficial information can be analyzed and presented through intuitive visualization interfaces, such as key scholars, important institutions, inter-scholar relationships, mapping between scholars and

issues, and mapping between scholars and viewpoints, enabling relevant scholars and decision-makers to consult them intuitively. This allows them to invest limited time and energy into more complex and in-depth research problems, thereby reducing their burden. Currently, numerous visualization methods and technologies are available, such as relationship networks, multi-layer networks, and knowledge graphs, which can be selected based on specific data conditions and user needs.

3.3 Platform Architecture Constructing the World China Studies thematic database requires solving the following key problems: (1) acquiring and mining overseas Chinese studies scholars and their issues and viewpoints, and multi-dimensionally presenting the full picture of overseas Chinese studies; (2) providing diverse retrieval functions through a portal website to facilitate searching for scholars, outcomes, and viewpoints; (3) visually displaying complex relationships among scholars through visualization technology to help scholars and decision-makers more conveniently formulate information transmission pathways and mechanisms. Its architecture from bottom to top mainly includes infrastructure, technical support, database, data application, and service presentation, as shown in [Figure 1: see original paper]:

- (1) **Infrastructure.** The infrastructure of the World China Studies thematic database includes servers, databases, basic networks, etc. The database can select CentOS as the operating system for its security, low maintenance, and stability; PostgreSQL relational database can be chosen for its open-source nature and rich functionality.
- (2) **Technical Support.** The technical support for the World China Studies thematic database includes modules for data acquisition, data cleaning, data retrieval, and visualization, using technologies such as Python web scraping, Java Web services, and front-end visualization.
- (3) **Database.** The World China Studies thematic database can aggregate multi-source, multi-type information from academic research, think tanks, news media, and social networks, establishing associations between individuals and various types of information to lay the foundation for deeper issue and viewpoint extraction and information dissemination pathway planning.
- (4) **Data Application.** Data applications in the World China Studies thematic database include data viewing, retrieval, and visualization functions. These functions present data to users in graphical or textual ways and provide better display through selection and retrieval methods.
- (5) **Service Presentation.** All data application functions are presented through the World China Studies thematic database portal website.

4 Construction and Application of the East China Normal University World China Studies Database Platform

As one of the earliest Chinese research institutions to conduct counter-research on overseas Chinese studies, East China Normal University built the World China Studies thematic database in 2019 to support policy decision-making and promote the further development of World China Studies. This database is an information resource database that collects, analyzes, processes, stores, and displays scattered resources such as individuals conducting Chinese studies outside China and their related papers, works, reports, news, and think tank outcomes.

4.1 Data Acquisition

4.1.1 Data Sources To simultaneously cover overseas Chinese research outcomes in both “Sinology” and “China Studies” domains, the thematic database selected Google Scholar and think tanks as its two major data sources. Google Scholar covers both “Sinology” and “China Studies,” while think tanks focus on contemporary Chinese issues in politics, military, society, economics, science, and technology.

4.1.2 Data Acquisition Strategy Unlike traditional data acquisition that directly collects papers and reports retrieved through keywords, this database adopts a different strategy to reduce the difficulty and pressure of scholar deduplication and cleaning in the scholar database. It first extracts scholar information from search results, then expands to collect scholars’ research-related papers, reports, news, and other resources.

First, search terms such as “Chinese military,” “Chinese investment,” “Chinese conflict,” “Chinese language,” “Chinese internet,” “Chinese censorship,” “Chinese population,” “Chinese government,” “Chinese traffic,” “Chinese agriculture,” and “Chinese industry” are used on Google Scholar to retrieve academic outcomes related to Chinese studies in fields like military, investment, language, and internet. Second, scholar information is extracted from these academic outcome lists to obtain scholar fields. Finally, scholars’ complete paper lists and citation counts are obtained, and papers related to China are screened from the lists using “China” and “Chinese” as search terms to serve as data sources for the paper database.

For think tanks, data is obtained through the following steps:

- (1) Referencing the “2018 Global Go To Think Tank Index Report” published by the University of Pennsylvania in January 2019 [15], the top 20 think tanks with search results are selected from the top 100 global think tanks list using “China” and “Chinese” as search terms, as shown in :

Think Tank Data Sources
(Brookings) (U.S.)

RAND Corporation (U.S.)
Chatham House (U.K.)
Korea Institute for International Economic Policy, LIEP (Korea)
Carnegie Endowment for International Peace (U.S.)
Wilson Center (U.S.)
Center for Strategic and International Studies, CSIS (U.S.)
Italian Institute for International Political Studies (Italy)
International Institute for Strategic Studies, IISS (U.K.)
Lowy Institute for International Policy (Australia)
SWP-Berlin (Germany)
French Institute of International Relations, IFRI (France)
Council on Foreign Relations (U.S.)
The Heritage Foundation (U.S.)
Peterson Institute for International Economics (U.S.)
Center for International Governance Innovation, CIGI (Canada)
German Council on Foreign Relations (Germany)
Human Rights Watch (U.S.)
European Council on Foreign Relations (U.K.)
Australian Institute of International Affairs, AIIA (Australia)

- (2) Visiting the official websites of these think tanks and using “China” and “Chinese” as search terms to identify relevant scholars. In October 2019, among 3,168 scholars from these 20 think tanks, 240 scholars (approximately 7.58%) were identified using “China” and “Chinese” as search terms.
- (3) Visiting scholar profile pages to obtain scholar fields, then using “China” and “Chinese” as search terms to screen scholars’ reports, viewpoints, news, blogs, and other resources for China-related content as data sources for the public opinion database.

4.2 Data Processing and Integration

Based on actual data conditions, processing includes paper deduplication, publication date standardization, personal name unification, null character processing, garbled field correction and format cleaning, and important count cleaning and statistical sorting. Through observation, scholar profile pages typically display full names, while other places often show names without middle names or with initials. Therefore, during data cleaning, names on profile pages are matched with author names in papers and think tank articles to add mapping relationships. For scholars with multiple null fields or no papers or think tank outcomes, considering their likely marginal position in Chinese studies research, such scholars are directly removed.

During database construction, preliminary research was conducted on Google Scholar and think tank data sources, and combining the field conditions of collected papers and public opinion data, scholar database tables, paper database

tables, and public opinion database tables were designed, with detailed contents shown in , , and , respectively constructing three sub-databases: scholar database, paper database, and public opinion database. To establish associations between individuals and information, paper database tables and public opinion database tables are linked to scholar database tables through the scholarid field. During data insertion, based on the relationships between database tables, individuals' IDs (scholarid field) are located and used as filler fields when inserting paper and public opinion data, thereby achieving interconnection among the three sub-databases.

After cleaning and organization, the numbers of scholar, paper, and public opinion information entries are shown in :

Numbers of Scholars, Papers, and Public Opinion Entries After Cleaning

Unit: entries

Scholars: 240

Papers: 2,847

Public Opinion: 1,234

Scholar Database Table

Paper Database Fields

Public Opinion Database Fields

4.3 Data Retrieval and Utilization

4.3.1 Query and Retrieval The World China Studies thematic database supports both keyword retrieval on the database homepage and keyword retrieval or classified browsing after entering the scholar database, paper database, or public opinion database.

In the scholar database, browsing by scholar source and name initials is supported. Scholar details include name, position and institution, research field, source, relationship graph with other scholars, citation count visualization, and lists of China-related papers or public opinion. Clicking paper titles redirects to the corresponding paper detail page in the paper database; clicking public opinion article titles redirects to the corresponding public opinion detail page in the public opinion database.

In the paper database, browsing by publication date and citation count is supported. Paper details include title, author, abstract, journal, publication date and volume/issue, page range, publisher, and citation count visualization.

In the public opinion database, browsing by public opinion category is supported. Public opinion details include title, author, category, publication date, full text, and original link.

4.3.2 Identification of Important Scholars The World China Studies thematic database visually displays the top 10 scholars by citation count: M.F. Zhou, M. Santosh, B.M. Jahn, B. Naughton, J.G. Liou, M.H. Bond, B. Popkin,

F.L. Wu, J. He, and D. Shambaugh. Since scholars from think tank sources lack citation information, the top 10 are all from academic papers. The annual changes in these scholars' publication numbers and citation counts are shown in [Figure 2: see original paper].

Analysis reveals that these scholars' research fields mainly concentrate in geology (4 scholars), economics (1), social psychology (1), health and nutrition (1), urban development and governance (1), epidemiology (1), and Asian international relations (1), showing large field span. Notably, geology not only has many scholars but also high research output and very high citation counts.

4.3.3 Visualization of Scholar Relationship Graphs The World China Studies thematic database visually displays relationship graphs among scholars. [Figure 3: see original paper] takes Harvard University scholar Alastair Iain Johnston as an example, showing his connections with other scholars and connections among scholars associated with him. For instance, Johnston is associated with A. Carlson, J.J. Suh, and P. Katzenstein, while Carlson, Suh, and Katzenstein are mutually connected, suggesting these four scholars may conduct joint research.

4.3.4 Identification of Important Research Fields The World China Studies thematic database visually presents research fields of World China Studies scholars, as shown in [Figure 4: see original paper]. Analysis of scholars and their fields reveals that Economics, International Relations, International Business/Trade, Political Economy, Machine Learning, Climate Change, and Epidemiology are key focus areas for overseas scholars. Notably, while 4 of the top 10 cited scholars work in Geology, this represents only a small portion of overall overseas Chinese studies. The reasons may be: (1) think tank scholars' focus areas differ significantly from academic scholars, with academic scholars focusing more on scientific issues while think tank scholars focus more on economic and political issues; and (2) although geology researchers have high individual output, their numbers are smaller than scholars in economics and politics, resulting in lower total output in these fields.

4.3.5 Identification of Important Research Institutions Analysis of institutional affiliations of academic scholars in the World China Studies thematic database reveals hot institutions for Chinese studies research, including University of Hong Kong, University of California, City University of Hong Kong, Hong Kong Polytechnic University, National University of Singapore, University of Southern California, Australian National University, Hong Kong University of Science and Technology, and Monash University. Scholars focusing on Chinese studies research can continuously monitor these institutions to track research progress, as shown in [Figure 5: see original paper].

The database also ranks and visualizes the 20 think tank data sources by scholar numbers, as shown in [Figure 6: see original paper]. Brookings, RAND Corpo-

ration, and Chatham House rank top three, with their Chinese studies scholars accounting for 39.58% of the total from all 20 think tanks, indicating these three think tanks have many scholars focusing on Chinese issues and require special attention.

5 Conclusion and Outlook

Constructing the World China Studies thematic database addresses long-standing challenges in contemporary World China Studies research at the stages of data acquisition, processing and integration, and retrieval and utilization, acquiring, integrating, and sharing knowledge resources on overseas Chinese studies. This holds major significance for both academic and practical levels.

The East China Normal University World China Studies Database represents an attempt to address current World China Studies research challenges through thematic database construction. In this practice, we attempted to elevate think tank public opinion to the same level of importance as academic research while simultaneously attending to both traditional “Sinology” and contemporary Chinese studies without rigidly separating them. This meaningful attempt allows researchers to conduct studies from more perspectives, methods, and dimensions based on database data. However, the database still has shortcomings requiring improvement: (1) The search terms used in data acquisition have significant limitations, which is very one-sided given the breadth and diversity of Chinese studies research, resulting in a large gap between current data volume and actual Chinese studies research volume. This requires systematic and in-depth investigation of World China Studies research topics to design more comprehensive, systematic, and focused search term lists to present the full picture of overseas Chinese studies and avoid missing important topics. (2) Data sources are singular—for instance, only Google Scholar for academic research and only 20 think tank institutions, particularly lacking news media and social networks as important sources of public opinion, which may cause deviations in identifying important research institutions and individuals. Therefore, data sources need supplementation, focusing on overseas mainstream media with wide audiences and influence, and social networks with large user bases and broad coverage. For academic research, non-English resources, especially from neighboring countries like Japan and Korea, should be included. (3) Multi-level and multi-angle data association, analysis, and visualization are limited. Obtaining deeper information or conclusions still requires researchers to conduct mining and analysis independently, with limited direct decision-making support effectiveness. Therefore, more analytical functions and visualization presentations need to be developed and applied.

The identification of these issues points the direction for future continuous optimization and improvement of the World China Studies thematic database, enabling it to serve academic research and national decision-making more accurately and effectively.

References

- [1] Zhu Zhengmin. New Reflections on Historiography Theory and History of Historiography—Discussion on the Relationship with Overseas Chinese Studies[J]. *International Sinology*, 2012(2): 15-24.
- [2] Liang Yi. Xi Jinping's Important Discourse on External Propaganda and Overseas Chinese Studies[J]. *World Socialism Studies*, 2018(9): 4-14, 95.
- [3] Shanghai Academy of Social Sciences World China Studies Institute. *Frontiers of World China Studies Theory*[M]. Shanghai: Shanghai Academy of Social Sciences Press, 2016: 3.
- [4] Zhu Zhengmin. Professor Paul Cohen's Qing History Research[J]. *Journal of Jiangxi Normal University (Philosophy and Social Sciences Edition)*, 2004(6): 25-31.
- [5] Yan Shaodang. Reflections on Overseas Chinese Studies[J]. *Exploration and Free Views*, 2007(2): 33-36.
- [6] Wu Yuanyuan. A Century of Chinese Scholars' Criticism of Overseas Sinology and Its Implications[J]. *Theoretical Studies on History*, 2019(4): 112-124, 159-160.
- [7] Wei Haisheng, Tang Zhouyan, Zhang Shuhua. Facing a New Era of Overseas Contemporary China Studies[J]. *Social Sciences Abroad*, 2019(5): 21-30.
- [8] Zhang Xiping, Shi Zhiyu, Romanov, et al. China's Development over the Past 70 Years and Overseas Chinese Studies[J]. *Social Sciences Abroad*, 2019(6): 61-73.
- [9] Sun Jianjun. Research on Key Technologies and Methods for Humanities and Social Sciences Thematic Database Construction[J]. *Journal of Information Resources Management*, 2020, 10(5): 11.
- [10] Shi Yanping, Li Yang. Research on the Construction and Application of Linked Data Models for Humanities and Social Sciences Thematic Databases[J]. *Modern Intelligence*, 2019(12): 19-27.
- [11] Jiang Ting, Sun Jianjun. Research on Deep Semanticization of Humanities and Social Sciences Thematic Databases[J]. *Journal of Information Resources Management*, 2020, 10(5): 12-22.
- [12] Mao Ping, Ju Xiaohong. Research on Data Resource Aggregation for Humanities and Social Sciences Thematic Databases Based on Knowledge Supernetwork Theory[J]. *Journal of Information Resources Management*, 2020(5): 38-47, 54.
- [13] Ma Hongjia, Li Jie, Shen Yong. Research on the Integration Trend of Digital Resource Aggregation Methods[J]. *Information and Documentation Services*, 2015(5): 24-29.
- [14] Dong Hui. Information Retrieval Based on Ontology and Digital Libraries[J]. *Journal of the China Society for Scientific and Technical Information*, 2003(6): 648-652.
- [15] 2018 Global Go To Think Tank Index Report[EB/OL]. [2020-09-10]. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1017&context=think_{tanks}.

Author Contributions:

Deng Luxiang: Manuscript writing;
Xu Xin: Conceptualization and review.

Abstract:

[Purpose/Significance] Trying to build a thematic database of World China Studies to fix the problems including hard to get the knowledge of China from overseas, paying insufficient attention to public opinions, and ignoring the correlation of scholars in World China Studies. [Method/Process] Introduce the design of the World China Studies thematic database in the aspects of data acquisition, data processing and integration, as well as data retrieval and utilization. And then did a case study of the World China Studies Thematic Database in East China Normal University. [Result/Conclusion] Building the thematic database of World China Studies can greatly facilitate the locating of important scholars, important institutions, hot fields, as well as correlations among scholars. In addition, the database could also provide data sources for other in-depth and multidimensional research.

Keywords: World China Studies; Thematic Database; Think Tank; Public Opinion

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.