Issues in Science and Technology Librarianship

Tips from the Experts

Chinese Sci-Tech Journal Databases

Linda R. Musser
Head of the Fletcher L. Byrom Earth and Mineral Sciences Library
Penn State University
University Park, PA
lindamusser@psu.edu

Yurong Y. Atwill, PhD
Asian Studies Librarian
Penn State University
University Park, PA
yya2@psu.edu

Abstract

Western indexing tools have a documented English-language bias, which contributes to the challenges facing researchers publishing in their native language. There are actions that librarians can take to mitigate this bias such as including information on library guides about indexes that highlight resources not covered by traditional sci-tech indexing tools. This article introduces three academic journal databases from China that are available for North American and other institutions. With a focus on sci-tech journals, basic descriptions of coverage and functionality of these resources is presented. Significant quantities of Chinese sci-tech information are not indexed by Western sci-tech abstract and indexing tools but are available via the databases discussed.

Recommended citation:

Musser, L.R. & Atwill, Y.Y. (2021). Chinese sci-tech journal databases. *Issues in Science and Technology Librarianship, 99*. https://doi.org/10.29173/istl2622

Introduction

Over the millennia, science has been communicated in many languages and researchers, of necessity, were multilingual. At various times, a common language emerged – Greek, Arabic, Latin – to help bridge the language barriers in western science; Chinese was the lingua franca in east Asia for centuries. Throughout these periods, even with a common language available, scientists remained largely multilingual. During the modern period, the languages used to communicate science began to condense. By the mid-1800s, the primary languages of western science consisted of English, French and German, in roughly equal proportions (Gordin, 2015). Partly driven by political forces, the core languages of the sciences continued to evolve such that, by 1934, English-language serials comprised nearly half of the worldwide total, followed by significantly fewer but equivalent numbers of German and French serials, with smaller numbers of Russian, Italian and serials of other languages (Sherrington, 1934). There was a surge in use of Russian language in the 1950s and 1960s however, by the 1980s, English language had become the predominant language of science, a trajectory that continues today.

The major tools that index scientific research have long had uneven coverage of non-English-language publications. In 1967, Wood (1967) found that while nearly fifty percent of sci-tech literature was written in non-English languages, the coverage of the literature by sci-tech indexes reflected an English-language bias, with 82% of abstracts of English language publications. The high percentage of English language coverage in indexes continues today. For example, *Compendex* requires an English language abstract for inclusion and English language materials comprise approximately 90% of the database. The English language dominance in the *Web of Science* ecosystem is even more extreme at 95% (Vera-Baceta, 2019).

English is currently the common language used for scientific communication in this millennium and, as such, imposes extra burdens on researchers for whom English is a second (or third) language (ESL). Mastering reading and writing in a foreign language, in this case English, takes time, effort and practice — something that native speakers of English are largely exempt from. Not only must ESL researchers master reading in English but they also must be prepared to publish their findings in English or risk their work being overlooked by the mainstream. To this inherent inequity, add the fact that, if an ESL researcher chooses to publish in their native language, that publication is much less likely to be indexed in mainstream sci-tech indexing tools such as *Compendex* and *Web of Science*. If the researcher's native language utilizes non-Roman scripts, the chances of inclusion in a western indexing tool are reduced even further.

Such are the challenges for ESL researchers in China and elsewhere. Librarians can help mitigate this situation by advocating for better coverage of non-English language materials in sci-tech indexing tools and promoting existing tools that cover non-English resources. Inclusion of non-English indexing tools on sci-tech subject guides is another mechanism to reduce the impact of the English-language bias present in major sci-tech databases. This article examines both these aspects for Chinese resources.

Awareness of Chinese Sci-Tech Resources

Interest in Chinese journal databases to date has largely been driven by librarians and scholars in the humanities and social sciences. These databases are often multidisciplinary so, for those universities that already have access to Chinese databases, we wondered if the sci-tech librarians at those institutions were promoting them to their users.

We examined the library research guides of nine large academic institutions with significant online Chinese studies resources to determine whether the Chinese journal aggregators licensed at those institutions were mentioned or promoted to sci-tech users on relevant library guides. We found no mentions of any of the Chinese journal aggregators although, refreshingly, the University of Michigan had a separate library guide focused on Chinese resources in STEM (Fu, 2021).

We also examined the availability of Chinese journal aggregators at thirteen libraries with strong engineering programs and reviewed their engineering subject guides to determine if the engineering subject specialists were promoting access to the Chinese sci-tech literature to their users. Of the thirteen universities, nine had access to one or more of the Chinese journal aggregators but none promoted those resources on their engineering library guides.

Focus on China

China's impact on global research and development has been significant in recent decades, with China contributing approximately a third of total worldwide growth in R&D since 2000 (U.S. National Science Board, 2020). As measured by research publications, China has surpassed the United States in total quantity of sci-tech publications, as illustrated in Figure 1, and the National Science Board (2020) reports that China produces nearly twice as many articles related to engineering than the United States. While the publication output in China has risen, international recognition as measured by citations has lagged. Recognizing that the language barrier was contributing to the low citation rate outside of China for Chinese works, the Chinese government provided support to launch English-language journals (Wang, 2018) and, to improve indexing of works written in Chinese, inclusion of English-language citations and abstracts are now commonly required for publication in Chinese-language journals.

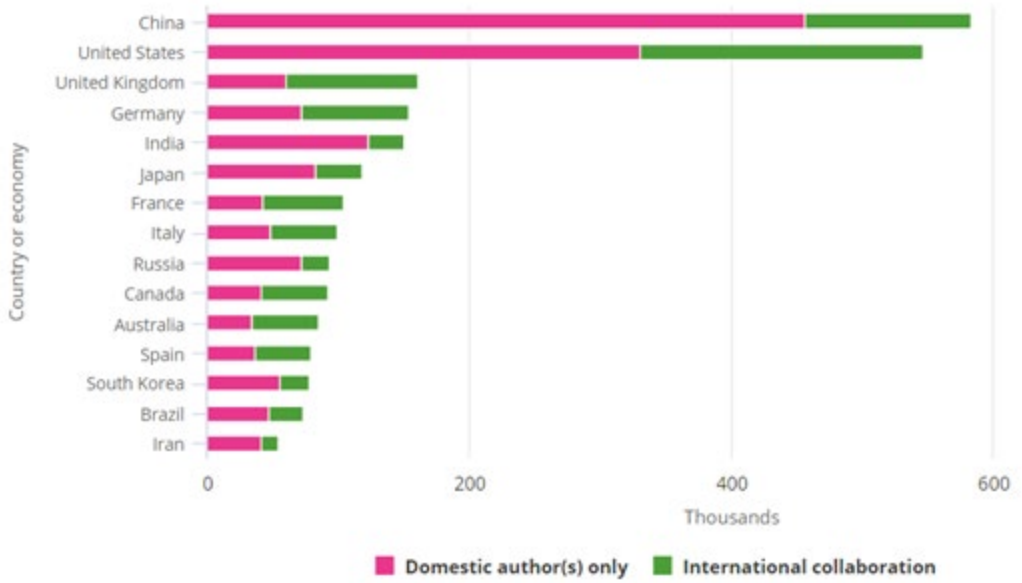

Figure 1: Authorship of sci-tech articles published in 2018, by country (U.S. National Science Board, 2020).

Traditional science and technology databases index some Chinese sources; however, in most cases, coverage remains limited. Previous works have highlighted or compared Chinese information resources but those with a sci-tech focus are almost nonexistent (Wang, 2006; Zhou, 2019). This work attempts to determine the extent of current coverage of mainland Chinese sources in selected Western science and technology databases and highlights additional databases that librarians can explore to address coverage gaps related to the Chinese science and technology literature.

Table 1: Number of sci-tech serials from China based on data from *UlrichsWeb Global Serials Directory*

Subject area	All active serials	Academically-focused serials	Academically-focused, refereed serials
engineering/ technology	3,574	2,816	675
earth/space/environmental sciences	1,333	1,136	374
biology	1,082	911	235
medicine/health	2,369	2,130	488
TOTAL	***8,358***	***6,993***	***1,772***

Background on Chinese Journal Aggregators

The major producers of academic journal databases in China are: China National Knowledge Infrastructure (CNKI), VIP Information Consulting Company (VIP), and Wanfang Data Company. CNKI, initiated by Tsinghua University, emerged in 1998 as a Chinese digital information resource and is one of the first Chinese companies to successfully market to the United States. Wanfang, established in 1993 and backed by the Institute of Scientific and Technical Information of China under the Chinese

Ministry of Science and Technology, was the very first corporation in mainland China with databases of Chinese information resources as its core business (Atwill, 2005). The VIP products have not been widely marketed to libraries overseas, therefore this study will exclude VIP offerings and instead include Superstar, a more recent entry to the Chinese journal aggregator market. Superstar Co. was established in 1993 and originally focused on the digitization of print materials, creating one of the largest Chinese digital book databases in the world (Baidu Baike, 2021). In recent years, Superstar created the Superstar Journals database and became the latest company to join the journal aggregator market. The China market is mature and, since the first three companies have largely fulfilled the internal-to-China market needs, Superstar has focused its journal database on the international market. Specifically, this article examines the sci-tech offerings of the *China Academic Journals* (CAJ) database of CNKI, *China Online Journals* (COJ) database of Wanfang and the *Superstar Journals* database.

Unique Characteristics of Chinese Journals and Journal Aggregators

The current practice of Chinese academic journals is to require authors to provide basic bibliographic information in both Chinese and English, e.g., title, author, keywords and abstract. Such practice allows e-resource aggregators to record basic bibliographic information in both languages, though journal issues from decades ago may lack such information.

Readers who are interested in older publications may notice that, for journals started a long time ago, there is usually a big gap between 1966 and 1978. During the Chinese Cultural Revolution, which ran between 1966 and 1976, most academic and research work was forced to cease. Education and academic research gradually returned beginning around 1978 and 1979. For example, *Journal of China Coal Society* published during 1964-1966, resumed in 1979 and continues to present day. While a majority of publications in social sciences and humanities also halted, a few might have "survived" by turning the focus towards revolutionary topics. However, sci-tech publications, and research work in general, simply ceased.

When discussing the quality of Chinese academic journals, core journals are often mentioned. Chinese core journals are viewed as similar to peer-reviewed journals in American academic and research fields. To be considered a core journal, a journal must be recognized officially, usually by the designated professional organization in the field. The core journals are listed in various publications, such as *A Guide to the Core Journal of China*, published by Beijing University Press (Zhu, 2012). This guide is periodically updated, usually every four years or so. The recognized core journals play an important role in academic fields for promotion or obtaining doctoral degrees in China.

Chinese Journal Aggregators

China Academic Journals (CAJ) Database of CNKI

When CNKI began building CAJ, full text coverage was 1994 to present. Over the years, more than 3,500 journals were retrospectively added back to the inaugural issue.

Currently there are over 8,233 (over 1,900 core journals) academic journals included in CAJ; 2,614 of these are sci-tech titles (excluding agriculture, medicine and health). Many institutions cannot afford the entire CAJ database but CNKI offers subscriptions to subsets of the database by subject. Regardless of the number of subject subsets subscribed, users may search the entire CAJ database to identify articles. Non-subscribers can search CAJ but access to full text is limited to the subscribed subject subset(s). Information on CAJ can be found on their website (CNKI, 2021).

China Online Journals (COJ) Database of Wanfang

COJ provides access to more than 8,200 academic journals published in China (of which 1,775 are core journals (Yu, 2016); over 3,000 of these are sci-tech titles (excluding agriculture, medicine and health). Users may search the database with both Chinese and English keywords, and a bilingual dictionary is available. For most journals, coverage is from 1998 onward, with full text. Like CNKI, Wanfang offers a choice of subscription to any one of eight subject sections. Wanfang encourages users to subscribe to the entire COJ database and offers big discounts for whole subscription. Details on the COJ can be found on their website (Wanfang, 2021).

Superstar Journals Database

Superstar Journals (also known as *Chaoxing* 超星期刊) contains around 7,400 full-text Chinese journals, of which approximately 2,400 are sci-tech titles, and more than 1,300 core journals. As the late comer among journal aggregators, Superstar offers a smaller number of journal titles compared to the other two suppliers, with coverage dating primarily from 2000. It offers an affordable price that contains access to multiple subjects. Details on *Superstar Journals* can be found on their website (Superstar, 2021).

Chinese Journal Aggregators Compared

Database Size

CAJ has the largest number of journals, core journals and full-text coverage. COJ is not far behind, and Superstar is third. There are titles covered by COJ and Superstar that CAJ does not have. Since Wanfang was established by the Institute of Scientific and Technical Information of China under the Chinese Ministry of Science and Technology, its initial focus was on science and technology; thus, COJ has the largest number of sci-tech journals among the three.

Full-Text Coverage

CAJ offers full text for all its journals starting from 1994 to present. It has been retrospectively adding earlier volumes and issues, offering full-text access from the initial issues for many journals. COJ full-text coverage ranges from 1998 to present. As a newcomer, Superstar full-text coverage is less than CAJ and COJ, with the majority of coverage starting after 2000. However, Superstar is retrospectively adding older volumes on a continuous basis.

Table of Contents Display

When searching a specific journal, both CAJ and Superstar present tables of contents for issues beyond their full-text coverage, i.e., article records are displayed even if full text is not available. COJ tends to display the years and issues of full-text coverage only. All three aggregators provide introduction and evaluation information about the journal, while COJ and Superstar's introductions often include history of the journal in detail.

Frequency of Updates

COJ updates its content twice weekly while CAJ and Superstar claim daily updating.

Pricing Options

In general, the CAJ database, being larger, may be more expensive than COJ or Superstar. CAJ and COJ offer options of subscription to subject subsets of their databases to cut down the cost. Price varies depending on the packages selected. Discounted price may be offered to subscribers who select multiple subsets or the entire database. As the newcomer, Superstar markets its journal database by offering a competitive price for the entire database without breaking down to subsets.

Overlap

A 2016 study measured the overlap between CAJ and COJ at 78%, (Yu, 2016), indicating a high rate of overlap of academic journals. Meanwhile, database producers have attempted to obtain exclusive rights over the years. In 2008, Wanfang obtained exclusive rights to 220 medical journals (of which 71 are core journals), while CNKI secured exclusive rights to 2,300 journals (of which 1,000 are core journals) (Tong, 2016). When exclusive rights are obtained by one aggregator all other aggregators lose the ability to host the full-text content, however, over time coverage may change.

Access to Searching

One little-known feature of these three aggregators is that they provide free searching to non-subscribers. Both CNKI and Wanfang allow users to conduct federated searches in their main platform across multiple databases, including their entire journal database plus other collections, e.g., theses and dissertations, proceedings, patents and so on. Users may select and view full bibliographic records, including abstracts, but full-text viewing and downloading are limited to subscribers only. Superstar's free search is limited to journals only and users may view the results list but not view the full bibliographic record. Table 2 provides links to the aggregators' websites with free searching options.

Table 2: Aggregators sites for free searching

Aggregator name	URL for free searching
CNKI	https://oversea.cnki.net/
Wanfang	https://www.wanfangdata.com.cn/
Superstar	https://qikan.chaoxing.com/

Interface Language

Both CAJ and Superstar offer Chinese and English interface options, while users may switch between two languages at ease. COJ offers a Chinese interface only; however, record displays in most fields usually contain both English and Chinese. Given that not all items have English abstracts or keywords, it is wise to perform subject searches using both English and Chinese terms. COJ includes a bilingual dictionary and a translation program such as Google Translate can be used to obtain appropriate search terms in Chinese. Table 3 illustrates the different results by performing a subject search using English and Chinese terms.

Table 3: Number of records retrieved in a subject search using English and Chinese terms, by database.

Search topic		China Academic Journals		China Online Journals		Superstar Journals	
English	Chinese	English	Chinese	English	Chinese	English	Chinese
Aerogel	气凝胶	9,388	12,140	12,003	11,160	2,140	3,431
Coal mining	采煤	58,231	32,186	30,522	41,093	53,715	51,499
Nanoparticles	纳米粒子	275,120	76,904	504,608	18,210	29,321	18,917

To help place these results in context, similar searches were performed in *Compendex* and *Web of Science* with the results displayed in Table 4. Searches were limited to journal document type and Chinese language. Searches were also performed using the author's country of origin equaling China; however, the search retrieved a preponderance of English-language commercial journal sources rather than Chinese publications therefore those results were not reported.

Table 4: Number of journal article records retrieved, overall and in Chinese, by database.

Search term or phrase	Compendex (journal articles)	Compendex (articles in Chinese)	Web of Science (journal articles)	Web of Science (articles in Chinese)
Aerogel	11,504	594	12,860	204
Coal mining	18,458	2,426	10,310	94
Nanoparticles	382,549	7,285	736,877	6,850

Chinese Sci-Tech Resources Beyond Journals

The three information providers profiled offer much more than just journal coverage. Both CNKI and Wanfang offer platforms designed for federated searching of multiple databases, free of charge. Wanfang covers journals, theses and dissertations, proceedings, patents, scientific reports, standards, legal regulations, gazetteers, and visuals. CNKI has offerings covering journals, theses and dissertations, proceedings, newspapers, yearbooks, monographic serials, patents, and standards.

Superstar is the largest Chinese ebook aggregator. Their book database, *Duxiu*, contains 3.5 million volumes, covering a majority of monographs published in China over the past 70 years (Superstar, 2021a). Users may request document delivery for 2.6 million volumes and a good number of those books cover sci-tech topics. Due to copyright limitations, *Duxiu* does not offer full-text online or immediate PDF download. Rather, subscribers have access to the book's table of contents and preliminaries from which to request document delivery, with the number of pages determined according to copyright regulations. The requested pages are then delivered to the user's email box, usually within minutes.

Conclusion

Clearly, there are opportunities to increase awareness of the availability of Chinese sci-tech resources among both sci-tech librarians and researchers. Given that searching is free in the three major journal aggregators, there are few barriers to exploring the wealth of hitherto overlooked Chinese research. The increasing availability of English language keywords and abstracts in Chinese databases, coupled with the ready availability of machine translation tools (e.g., Google Translate), searching and using resources in Chinese databases has become more practical for many librarians and researchers. This article focused on Chinese language sci-tech resources but similar challenges exist for other languages and countries. It is heartening to realize that free resources exist to facilitate discovery of non-English sci-tech publications. For example, SciELO provides free access to academic materials written in Spanish and Portuguese (SciELO, 2021). Sci-tech librarians can help combat the English-language bias inherent in traditional sci-tech indexing tools and contribute to reducing the inequities faced by ESL researchers by including and even promoting the use of non-English language sci-tech tools.

References

Atwill, Y. (2005). E-journals from China: Technical and collection issues. *The Journal of Academic Librarianship, 31*(6), 598-604. https://doi.org/10.1016/j.acalib.2005.09.001

Baidu Baike. (2021). *Superstar.* https://baike.baidu.com/item/超星/33315

Clarivate. (2021). *Chinese science citation database.* https://clarivate.com/webofsciencegroup/solutions/webofscience-chinese-science-citation-index/

CNKI. (2021). *CNKI index.* https://oversea.cnki.net/index/

Compendex. (2021). *Compendex source list.* https://www.elsevier.com/solutions/engineering-village/content/compendex

Fu, L. (2021). *Chinese language resources in STEM fields.* https://guides.lib.umich.edu/ChineseSTEM

Gordin, M. D. 2015). Absolute English. *Aeon*. https://aeon.co/essays/how-did-science-come-to-speak-only-english

Jiang, H. (2015). Review on the quality of 3 Chinese full-text journal databases. *Xiandai qingbao = Journal of Modern Information, 35*(9), 84-88, 170

SciELO. (2021). *Scientific electronic library online.* https://scielo.org

Sherrington, C. S. (1934, October). Language distribution of scientific periodicals. *Nature, 134*, 625.

Superstar (2021). *Chaoxing.* https://qikan.chaoxing.com

Superstar (2021a). *Duxiu.* http://www.duxiu.com/bottom/about.html

Tong, M. (2016). Comparative study on the included core journals of the 3 Chinese journals full-text databases. *Nong Ye Wang Luo Xin Xi = Agriculture Network Information, 2016-8*, 78-83.

UlrichsWeb. (2021). *UlrichsWeb global serials database.* https://www.ulrichsweb.com

U.S. National Science Board. (2020). *The state of U.S. science and engineering 2020.* https://ncses.nsf.gov/pubs/nsb20201/global-science-and-technology-capabilities

Vera-Baceta, M. A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics,* 121, 1803–1813 https://doi.org/10.1007/s11192-019-03264-z

Wanfang. (2021). *Wanfang data.* https://www.wanfangdata.com

Wang, J. (2006). Major Chinese full-text electronic information resources for researchers and scholars. *Serials Review, 32*(3), 164-171. https://doi.org/10.1016/j.serrev.2006.06.006

Wang, Y., Fang, Q., & Peng, W. (2018). China's recently launched English-language science and technology journals, 2012–16. *Journal of Scholarly Publishing, 50*(1), 37-47. https://doi.org/10.3138/jsp.50.1.07

Wood, D. N. (1967). The foreign language problem facing scientists and technologies in the U.K. - Report of a recent survey. *Journal of Documentation, 23*(3), 117–30.

Yu, Y. (2016). Comparative analysis of three domestic journal databases. *Hebei ke ji tu yuan = Hebei Library Journal of Science and Technology, 29*(4), 54-58.

Zhou, X. (2019). Comparative analysis of SinoMed, Weipu, Wanfang and CNKI Chinese literature network retrieval platforms. *Yi xue tu shu qing bao za zhi = Chinese Journal of Medical Library and Information Science, 28*(10), 63-69. https://doi.org/10.3969/j.issn.1671-3982.2019.10.009

Zhu, Q. (2012). *A guide to the core journals of China.* Beijing University Press