Chinese-language literature about Wikipedia: a meta-analysis of academic search engine result pages

Han-Teng Liao
Oxford Internet Institute
University of Oxford
Oxford, United Kingdom
hanteng@gmail.com

Bin Zhang
Center for the Studies of Information Resources
Wuhan University
Wuhan 430072, P. R. China
zb0205@126.com

ABSTRACT

This paper presents a webometric analysis of the academic search engine result pages (SERPs) of the Chinese-language term of "Wikipedia" across major Chinese-speaking regions of mainland China, Hong Kong and Taiwan. Because of the academic outcome, the findings can also be interpreted for further meta-analysis, or "research about research", of the Wikipedia research in Chinese-language literatures. The findings cover the results from four major search platforms: CNKI Scholar, Google Scholar China, Google Scholar Hong Kong and Google Scholar Taiwan. Cross tabulation of the results shows the major institutions (journals and academic departments) and scholarly archives for Chinese-language Wikipedia research. The findings suggest that there exists a divide between mainland Chinese academic sources/search results on one hand, and Hong Kong/Taiwanese ones on the other. Meta-analysis based on academic SERPs have implications for identifying the gaps and potentials in internationalization of Wikipedia research.

Categories and Subject Descriptors

 [**World Wide Web**]: Web searching and information discovery – *Web search engines*, *Page and site ranking*

Keywords

Chinese Internet, Chinese research on Wikipedia, academic search engines, academic databases.

1. INTRODUCTION

The ideal of using human knowledge to engage citizens has often been subject to parochial and national concerns despite the universal ideal of enlightenment encyclopedias. During the European enlightenment, geographic and linguistic barriers were among the major challenges for knowledge collection and diffusion [2, 5, 12]. Facing similar challenges, the Wikimedia Foundation, the hosting organization for all Wikipedia projects, has targeted the "Global South" regions of Brazil, India, and the Arabic language countries for engagement [17]. Nonetheless,

some research has suggested that cultural and linguistic factors have prevented wider acceptance of Wikipedia, especially the Chinese and Korean versions [14]. Thus, internationalization of Wikipedia research is needed for better understanding of the Wikipedia research around the world beyond English-language literatures. As academic databases and big data were highlighted as some of the frontier topics of information science[11], it is generally important to examine the role of Chinese-language academic databases and search platforms.

This paper aims to contribute to such internationalization efforts by looking at the current published Chinese-language literature as reported by major academic search engine results.

2. Methods

The exploratory method mimics the information seeking actions likely to be executed by Chinese-language Wikipedia researchers. The query of the Chinese-language term "Wikipedia" is submitted to major academic search engine platforms, and then the search engine result pages (SERPs) are scraped and mined for further analysis.

Because of the academic outcome, the expected findings can be interpreted for further meta-analysis, or "research about research" [1], of the Wikipedia research in Chinese-language literatures. It should be noted, however, the meta-analysis applied to social science research often involves a hypothesis that are being examined by a body of work. The work to be presented in this paper, on the other hand, is more of a descriptive meta-analysis of search engine result pages, or "search about research". The aim of the exploratory study is to identify search result patterns based on the results provided by Chinese-language academic search platforms, not to test a hypothesis normally seen in meta-analysis.

2.1 Data selection and data sets

This study includes CNKI scholar and three localized versions of Google Scholar (China, Hong Kong and Taiwan). CNKI refers to Chinse National Knowledge Infrastructure that was established in 1999 by Tsinghua University and Tsinghua Tongfang, with the aim to "achieve full social sharing and dissemination of knowledge resources"[4, 6, 16].

For data collection, the Chinese term "Wikipedia" was submitted to the four academic search engines in April 2014, with the number of results shown in Table 1. The queries of "Wikipedia" were submitted to each platform differently because mainland China uses simplified Chinese script and Hong Kong and Taiwan use traditional Chinese script[9]. The query of "维基百科" is used for mainland Chinese platforms; the query of "維基百科" is used for the other. Each platform reported to have different numbers of results: ranging from the highest one reported by Google Scholar China to the lowest one reported by CNKI scholar.

Table 1. Number of search engine result items

	CNKI Scholar	Google Scholar China (CN)	Google Scholar Hong Kong (HK)	Google Scholar Taiwan (TW)
Reported	464	27,300	3,950	3,950
Scrapped and studied	464	1,000	1,000	1,000

However, these reported numbers do not correspond to the numbers of results that users can actually view. For instance, CNKI scholar has a limit at 500 and the Google Scholar has a cap at 1000, thereby imposing a limit on the number of samples for this research. The second row of Table 1 shows the number of results actually scraped and studied in this paper. The dataset examined in this paper thus contains 3464 data points in total, each of which contains, if available, meta information of title, authors, institutions, links and short descriptions of the result.

2.2 Cross-tabulation analysis

By extracting the meta information from the search results, several cross-tabulation is made to examine the patterns regarding the major publishing institutions, archive platforms, and publication years.

3. Results

The findings of crosstabs are presented as follows. As the results for Google Scholar Hong Kong and Google Scholar Taiwan are almost identical (only 13 different out of 1000 results), only the results for Google Scholar Taiwan will be listed if the crosstabs results are identical for both Google Scholar Taiwan and Google Scholar Hong Kong.

3.1 Major (publishing) institutions

Table 2 lists the most frequently-appearing source institutions, which include academic journals, degree-granting institutions and popular science magazines. Only those receive 10 or more hits in total are listed.

Table 2. Most frequently appearing source institutions

Institutions (in Chinese)	Institutions	CNKI Scholar	Google Scholar CN	Google Scholar TW	Grand Total
情报学报	Journal of the China Society for Scientific and Technical Information	0	65	0	65
国家图书馆学刊	Journal of the National Library of China	1	41	0	42
臺灣大學企業管理碩士專班學位論文	MBA dissertations, National Taiwan University	0	0	29	29
臺灣師範大學工業科技教育學系學位論文	Department of Industrial Technology Education Dissertations, Taiwan Normal University	0	0	28	28
圖文傳播藝術學報	NTUA Department of Graphic Communication Arts	0	0	23	23
现代图书情报技术	Modern Technology of Library and	5	12	0	17
互联网周刊	China Internet Weekly	6	10	0	16
图书情报工作	Library and Information Service	5	9	0	14
青年记者	"Youth Reporter" magazine	3	11	0	14
中文信息学报	Journal of Chinese Information Processing	6	7	0	13
互联网天地	China Internet Magazine	7	6	0	13
上海交通大学	Shanghai Jiao Tong University	13	0	0	13
情报资料工作	Information and Documentation Services	3	9	0	12
PAR表演藝術雜誌	Performance Arts Review	0	0	12	12
图书馆理论与实践	Library Theory and Practice	5	6	0	11
情报理论与实践	Information Studies: Theory & Application	5	6	0	11
哈尔滨工业大学	Harbin Institute of Technology	10	0	0	10
情报杂志	Journal of Intelligence	7	3	0	10
现代情报	Modern Information	2	8	0	10

The academic journals contain mostly mainland Chinese information and library science journals, including *Journal of the China Society for Scientific and Technical Information*, *Journal of the National Library of China*, *Modern Technology of Library and Information*, *Library and Information Service*, *Journal of Chinese Information Processing*, *Information and Documentation Services*, *Library Theory and Practice*, *Information Studies: Theory & Application*, *Journal of Intelligence*, and *Modern Information*.

Taiwanese journals such as *NTUA Department of Graphic Communication Arts,* and *Performance Arts Review* are also listed in Table 2. The listed degree-granting academic bodies include both Taiwanese (MBA and education dissertations) and mainland Chinese (Shanghai Jiao Tong University, and Harbin Institute of Technology) ones.

The remaining listed institutions include three popular science magazines: *China Internet Weekly*, "*Youth Reporter*"(author's translation) magazine, and *China Internet Magazine*. A clear division is shown in Table 2 between the mainland Chinese and Taiwanese results. (The Hong Kong results are identical to the Taiwanese ones.) Table 2 shows no overlapping results.

Between CNKI Scholar and Google Scholar China, substantial overlapping results exists for journals and magazines. Google Scholar China does not seem to include as many degree dissertations as CNKI Scholar.

3.2 Major (archive) platforms

Table 3 lists the most frequently appearing source platforms for each search engine results. (The Hong Kong results are identical to the Taiwanese ones.) As these data points indicate potential clicks to the websites that provide either the metadata or full texts of the citation, they provide important insights into the major academic search platforms across the Chinese-speaking regions.

Table 3. Most frequently appearing source platforms

	CNKI Scholar	Google Scholar China	Google Scholar Taiwan	Grand Total
CNKI platfrom	464	219	0	683
Education Taiwan	0	2	605	607
Miscellaneous	0	346	248	594
CQVIP platform	0	378	0	378
Airitilibrary platform	0	0	147	147
Wanfang Data platform	0	55	0	55
Grand Total	464	1000	1000	2464

It is clear that CKNI Scholar is the academic database itself with all the links pointing to its own sites. In contrast, Google Scholar China links outwards to three mainland Chinese academic databases: CNKI platform, CQVIP platform, Wangfang Data platform. CQVIP refers to Chongqing VIP Information Co., Ltd., a company has its historical roots under the Chinse Ministry of Science and Technology and is a strategic partner of Google search since 2005. Wangfang Data is another joint enterprise by various Chinese research institutes and publishers [4, 6, 16].

Google Scholar Taiwan links to mostly Taiwanese education websites with the domain names of "edu.tw" and a Taiwanese academic database called Airitilibrary platform.

Again, judging from the archive academic database results, there is little overlapping among the search results. The only exceptions are first the links to CNKI platform by CNKI Scholar and Google Scholar China, and second the two links to Taiwanese educational websites by Google Scholar China and Google Scholar Taiwan.

When compared with CNKI Scholar, Google Scholar China is more neutral in terms of platform choices given to the users. Table 4 shows the same data points listed in Table 3 with the details of the subdomain names. CNKI Scholar and Google Scholar China links to different web interfaces hosted by the CNKI, with heavy concentration of links in certain subdomain names. Google Scholar Taiwan's links to Taiwanese educational websites are much less concentrated.

Table 4. Most frequently appearing source platforms and their subdomain names

	CNKI Scholar	Google Scholar China	Google Scholar Taiwan	Grand Total
CNKI platfrom				
www.cnki.net	447	0	0	447
www.cnki.com.cn	0	103	0	103
cdmd.cnki.com.cn	0	103	0	103
cpfd.cnki.com.cn	0	13	0	13
dbpub.cnki.net	13	0	0	13
d.scholar.cnki.net	4	0	0	4
Subtotal	464	219	0	683
Education Taiwan*				
thesis.lib.ncu.edu.tw	0	0	75	75
pc01.lib.ntust.edu.tw	0	1	52	53
other*	0	1	478	479
Subtotal	**0**	**2**	**605**	**607**
Other websites	**0**	**346**	**248**	**594**
CQVIP platform				
www.cqvip.com	0	377	0	377
2010.cqvip.com	0	1	0	1
Subtotal	**0**	**378**	**0**	**378**
Airitilibrary platform				
www.airitilibrary.com	**0**	**0**	**147**	**147**
Wanfang Data platform				
d.wanfangdata.com.cn	**0**	**55**	**0**	**55**
Grand Total	**464**	**1000**	**1000**	**2464**

Note*: Taiwanese educational websites include 89 instances, with 9 subdomain names having more than 30 links to them.

3.3 Growth and freshness

The reported publication years can provide some insights into the growth and freshness of the found Chinese-language results on Wikipedia. Figure 1 shows the results from 2002 to 2014, indicating a general trend of growth from 2002 to 2011. (The Hong Kong results are identical to the Taiwanese ones.)
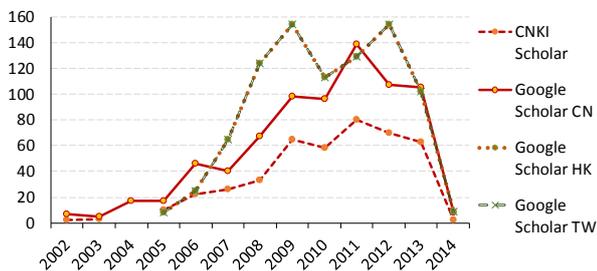

Figure 1. The trendlines for number of publications each year

One should not jump into the conclusion by interpreting the results of the years of 2012-2014 as indication of decline. The new publications may still be in the process of archiving or indexing. In early 2012, based on the doctoral and master theses from Taiwanese and mainland Chinese databases, we observed a growing trend from 2003 to 2009 (for Taiwan) or 2010 (for mainland China), and then a decrease in numbers from 2010 onwards to 2012 [8]. In light of the new findings shown in Figure 1, the 2011 results indicate a continuing growth for the Google Scholar China and CNKI scholar results, and a rebound for the Google Scholar Taiwan results. Hence, it is more likely that there exists a two- to three-year window where fresh publications can be archived and indexed by major academic databases.

4. Discussion

The status of the Chinese-language literature regarding Wikipedia, based on the SERPs from major academic search platforms across the regions of mainland China, Taiwan and Hong Kong, provide interesting findings as follows.

First, the presence of major institutions of academic journals, degree-granting bodies, and popular science magazines indicate substantial interest of these regions in Wikipedia. The perspectives include information and library science, management studies, and popular science. Little overlapping exists though between mainland Chinese and Hong Kong/Taiwanese results.

Second, the linking patterns show a distinct choice of academic databases used by these academic search platforms. CNKI Scholar seems to keep the links to its own websites. Google Scholar China links more evenly among three mainland Chinese academic databases: CNKI, CQVIP and Wangfang Data. In addition to several Taiwanese educational websites, Google Scholar Taiwan links to a Taiwanese academic database Airitilibrary platform.

Third, the trend graph based on the reported publication years indicate a growing number of publications from 2002 to 2011. In conjunction with the previous observation made in 2012[8], the currently shown decreasing results for 2012-2014 are likely the outcomes of the possible two- to three-year window for a Chinese-language publication to be archived and indexed.

The findings show a worrying situation where there exists little overlapping results between the mainland Chinese and Hong Kong/Taiwanese results. It is worrying because Chinese Wikipedia project is actually an open collaboration project contributed and managed by Chinese Wikipedians across these Chinese-speaking regions and beyond[7, 10]. Wikipedia researchers who intend to conduct a literature review on Chinese-language literatures must, at the current situation, acknowledge the division of the found academic search results and find literature across Chinese-language academic search platforms. Otherwise, the coverage of the literature review is likely to be bounded and thus limited to a certain regional perspective.

After trying entering different Chinese scripts into different search engines, we found that the Google Scholar results appear to return different results if the keywords were typed in different scripts. In contrast, the CNKI Scholar results appear to return similar and even smaller number of results. It means that users of Google Scholar can simply switch the keywords between traditional Chinese and simplified Chinese scripts for different scopes of results. It also presents a search engine design challenge in integrating sources that contain some degree of geolinguistic variations [9]. This research has captured the majority of users choice of Chinese scripts for each of the regions of mainland China, Hong Kong and Taiwan, future research can be conducted to include both scripts.

Another direction for future research is to consider the results of other academic search engines in China such as CQVIP platform and Wangfang Data. It is also noted that Baidu, the biggest search engine company in China, only launched its academic search service Baidu Xueshu on 13 June 2014 (after the research is conducted), which should be included by future research[18].

5. Conclusion

Internet internationalization has been an important aspect of its development [15] and the internationalization of media studies has been slowly catching up to examine the implications[3]. Wikipedia research faces the similar situation where internationalization of methods and findings is needed for better research and practices. This paper contributes to such an effort with a meta-analysis of academic SERPs for Chinese-language literatures.

Indeed, digital support for the main East Asian languages (Chinese, Japanese and Korean) becomes an important milestone for the Internet's internationalization. As put by a major East Asian Science Technology and Society (STS) scholar, Nakayama Shigeru [13]:

> *East Asians are accustomed to dealing with a multibyte system, in contrast to Western monobyte reductionist culture. It may be that in the future our multibyte culture will prove advantageous for dealing with complex systems.*

It is in this technical and linguistic context that Chinese Wikipedia contributes to the global Wikimedia movement with their internationalization innovations in multi-script editing platforms and automatic conversions that are now part of the MediaWiki codes [7].

In contrast to Chinese Wikipedian contributors, the Chinese-language Wikipedia research seems to be lagging if not stagnant in promoting exchanges across Chinese-speaking regions. Future research is needed to examine whether the citations of Chinese-language Wikipedia research are also divided between mainland China on one side and Hong Kong/Taiwan on the other. The findings here clearly indicate such a divide, in terms of academic search engine platforms, publication institutions and academic archive databases.

Similar research efforts can be conducted for other languages which have contributed to the Wikipedia research, for instance, Spanish, Arabic, Korean and Japanese. Although the meta analysis of academic SERPs cannot replace in-depth literature review, future work is needed to foster the knowledge exchange beyond just English language. Nevertheless, the meta analysis based on major academic SERPs conducted here presents a set of reproducible procedures to monitor the status of Wikipedia research through the different information windows that are used daily by academics and students, thereby pointing researchers to the gaps and potentials in internationalization of Wikipedia research.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Card, N.A. 2012. *Applied Meta-analysis for Social Science Research*. Guilford Press. New York.

[2] Delanty, G. 2009. *The Cosmopolitan Imagination: The Renewal of Critical Social Theory*. Cambridge University Press. Cambridge, UK.

[3] Goggin, G. and McLelland, M. 2009. The internationalization of the internet and its implications for media studies. *Internationalizing Media Studies*. D.K. Thussu, ed. Routledge. 294–307.

[4] Hsu, C.-H.[徐嘉晧] 2008. *A Comparative Study on Chinese Electronic Journal Systems between Taiwan and China [海峽兩岸中文電子期刊系統之比較研究]*. National Chengchi University.

[5] Israel, J.I. 2001. *Radical Enlightenment: Philosophy and the Making of Modernity 1650-1750*. Oxford University Press, USA. New York.

[6] Li, J.L.[李金兰] 2011. Comparative Analysis on Resources of CNKI, Wanfang Data and Weipu Information. [CNKI, 万方, 维普资源比较与分析]. *Information Research [情报探索]*. 4, (2011), 59–61.

[7] Liao, H.-T. 2009. Conflict and Consensus in the Chinese version of Wikipedia. *IEEE Technology and Society Magazine*. 28, 2 (2009), 49–56.

[8] Liao, H.-T. 2012. Growth of Academic Interest in Wikipedia From Major Chinese-speaking regions: Has it peaked? *Oxford Internet Institute Blog*.

[9] Liao, H.-T. 2013. How does localization influence online visibility of user-generated encyclopedias? A case study on Chinese-language Search Engine Result Pages (SERPs). *Proceedings of the 9th International Symposium on Open Collaboration* (Hong Kong, Aug. 2013).

[10] Liao, H.-T. 2014. *The Cultural Politics of User-Generated Encyclopedias: Comparing Chinese Wikipedia and Baidu*. University of Oxford.

[11] Ma, F. 2013. Historical Review of the Development of Information Science with Proposing Frontier Topics (情报学发展的历史回顾及前沿课题). *Document, Information & Knowledge (图书情报知识)*. 2 (2013), 4–12.

[12] Roche, D. 2006. Encyclopedias and the diffusion of knowledge. *The Cambridge History of Eighteenth-Century Political Thought*. M. Goldie and R. Wokler, eds. Cambridge University Press. 172–194.

[13] Shigeru, N.(中山茂) 2001. The Digital Revolution and East Asian Science. *Historical Perspectives on East Asian Science, Technology, and Medicine*. A.K.L. Chan et al., eds. World Scientific.

[14] Shim, J. and Yang, J. 2009. Why is Wikipedia not more widely accepted in Korea and China? Factors affecting knowledge-sharing adoption. *Decision Line*. 40, 2 (2009), 12–15.

[15] Stone, A. 2003. Internationalizing the internet. *IEEE Internet Computing*. 7, 3 (May 2003), 11–12.

[16] Tan, J.[谭捷] et al. 2010. A Comparative Study of Chinese Academic Journal Databases [中文学术期刊数据库的比较研究]. *Document, Information & Knowledge [图书情报知识]*. 4 (2010), 4–13.

[17] Wikimedia Meta 2012. Global Development. *Wikimedia Meta*.

[18] wokeji 2014. Baidu Xueshu build academic platform with hundred billion Chinese and English items indexed for national market (百度学术搜索打造国内亿级中英索引量学术平台). *huanqiu.com(环球网)*.