AN ANALYSIS OF COMMON SEARCH IN CHINESE OF GOOGLE THE META - LIBRARY

Likang Zhang*

Library of Chinese Academy of Social Sciences, Beijing 100732
**Email:*zhanglikang@sohu.com

ABSTRACT

Focusing on common searches in Chinese, this paper analyzes the web search retrieval mechanism for Chinese words concerning library-related core periodicals and meta-library vocabulary. Going into further details by counting and sorting the search words appearing at given periods of time on the web, the impact of web search on words related to library and information sciences as well as its development trend are intended to be revealed.

Keywords: Library science, Information science, Google search, Core periodicals, Relevant words

1   THE CONCEPT OF "CORE" IN WEB SEARCHING

1.1 The success of monopolizing information

It is an undeniable fact that the Google search engine, which ranks number one among web search platforms, has captured almost all available information. Its common search service in Chinese is distinctly different from the academic ones such as "GOOGLE SCHOLAR."

First, its common service exhausts to the extent possible all information related to the searched word. The information is then sorted according to time and relevance in a descending order. Although such information also includes so-called junk information that is not really relevant and obsolete information that has not phased out in a timely manner, the service still gives the user great satisfaction in terms of the time spent and the speed with which it renders such service. Further, those who do use this service are precisely a readership that is far greater in number than those who physically visit libraries.

Second, the common search service under the monopoly of the flagship of Google outperforms other web-based services in terms of its accuracy. This is based on its quantity-based operational policy, the mechanism of highly interrelated web-pages, and well-designed link structure. Such a search platform provided by Google facilitates for the users quick identification of what they want and guarantees a super performance that enables the user of common search in Chinese to download information without having to register or log in. On the other hand, "GOOGLE SCHOLAR," although highly professional and academic, requires registration and payment of fees, thus denying access to the many outlying user classes.

1.2 The "catfish effect" and the Google business principle

We may regard searched words with high click rates as the "catfish" on the web-based platform. For the purpose of this paper we will call them the "core words," and the huge amount of information relating to them is similar to sardines. It is only because the high frequency words can bring about the related information that the supporters or sponsors are willing to financially support these "catfish," so that these high frequency words become more active and more "nutritious" in an academic sense, accelerating the process of concentrating a core group of terms in a given profession. Meanwhile, terms that have low click rates will be dropped out over time. According to this demand, therefore, a well organized data bank is formed that offers dynamic and progressively accurate information on leading players in given fields as reported in core and integrated publications.

The "Catfish Effect," which was originally applied to talent management, is equally effective for the Google search platform because the rule that "the active few would stimulate the whole group" is also valid here, resulting in a constant increase in logins and gaining greater popularity among the users. It is the policy of information completeness that has paved the way for digital development of all trades including that of the libraries.

2 ASSESSING WEB SEARCH VS. CORE PERIODICALS

2.1 The immobilized core periodicals and the flexible web-search

At a time when the academic circle was busy defining "core periodicals" as a concept in bibliometrics, the Google search engine actually ignored it or immaterialized its connotation. Instead, Google offered meta-library resources on the Internet. This "counter measure," although seemingly by-passing the "core" concept, offers in fact a more impartial definition by means of inexhaustible sources of information. The reason being that, whatever definition the experts may offer, it is undeniable that terms or periodicals with higher click rates have higher popularity. And this is all that matters. The conventional "core" concept is turned into a type of bunching or cluster effect in the context of web searching. Google is quick in successfully dumping meta-library resources on the web, attracting funds from supporters or sponsors, and creating a win-win situation in which books, periodicals, and relevant technical terms that have high popularity can further flourish, while hinting at the development trend of the library-related technical terms or periodicals as of today or in the future.

2.2 Current status of core periodicals available on web search platform

The once-every-three-year assessment of "core periodicals" determined that 19 publications are the core periodicals for library science in China for the current three-year period, among them are the *Journal of the Library Science in China*, *Library and Information Science*, *Journal of Academic Libraries,* and *Library Theory and Practice*. Based on data from *A Guide to the Humanities & Social Science Core Journals in China 2004 Edition*, this author conducted a statistical analysis of the home pages of the core periodicals, the total search entries, the number of relations found, and sponsored links at a given point of time. Arranged in a left to right and top to bottom order, core periodicals are assessed and analyzed to discover their referral rankings in searches as an indication of their degree of influence on this disciplinary domain as shown in Table 1 below (Note: numbers of search entries as indicated in this Table are based on 10 entries per page):

TABLE 1.
Analysis of common searches in Google Chinese for core periodicals on library and information sciences

Positions in the number sequences under each heading (Library, Library Tribune etc) stand for the following respectively: Number of home pages → Total entries → number of relations founds → number of sponsored links)

1	2	3	4
1	Library 1 →940→6→1	Library Tribune 1→800→10→1	Information and Documentation Services 1→830→5→2
2	Information Science 1→840→5→1	Library Work and Study 1→790→10→0	Library and Information 1→700→10→2
3	Information and Documentation Services→ 1→830→5→2	Library Development 1→780→10→0	Library 1 →940→6→1
4	Journal of Information 1→830→0→1	Archives Science Study 1→770→10→0	Information Science 1→840→5→1
5	Library Tribune→ 1→800→10→1	Library Theory and Practice→1→710→10→0	Journal of Information 1→830→0→1
6	Library Journal 1→800→0→0	Journal of Academic libraries 1→710→10→0	Library Tribune 1→800→10→1
7	Library Work and Study 1→790→10→0	Archives Science Bulletin →700→10→10	The Journal of Library Science in China 1→630→10→1
8	Library Development 1→780→10→0	Library and information 1→700→10→2	Journal of the China Society for Scientific and Technical Information 1→630→9→1
9	Archives Science Study 1→770→10→0	The journal of the library science in China 1→630→10→1	Information Studies: Theory & Application 1→540→10→1
10	Knowledge of Library and Information Science 1→770→0→0	New Technology of Library and Information Service 1→630→10→0	Library Journal 1→800→0→0
11	Library and Information Science 1→760→2→0	Information Studies: Theory & Application 1→540→10→1	Library Work and Study 1→790→10→0
12	Journal of Academic Libraries 1→710→10→0	Journal of the China Society for Scientific and Technical Information→ 1→630→9→1	Library Development 1→780→10→0
13	Library Theory and Practice 1→710→10→0	Library 1 →940→6→1	Archives Science Study 1→770→10→0
14	Archives Science Bulletin 1→700→10→0	Information Science 1→840→5→1	Knowledge of Library and Information Science 1→770→0→0
15	Library and information 1→700→10→2	Information and Documentation Services 1→1830→5→2	Library Theory and Practice 1→710→10→0

16	The Journal of Library Science in China 1→630→10→1	Library and Information Science→ 1→760→2→0	Library and Information Science→ 1→760→2→0
17	Journal of the China Society for Scientific and Technical Information 1→630→9→1	Knowledge of Library and Information Science 1→770→0→0	Journal of Academic Libraries 1→710→0→0
18	New Technology of Library and Information Service 1→630→10→0	Journal of Information 1→830→0→1	New Technology of Library and Information Service 1→630→10→0
19	Information Studies: Theory & Application 1→540→10→1	Library Journal 1→800→0→0	Archives Science Bulletin 1→700→10→0

2.3 Analysis of web search for library core periodicals

Table 1 supports the following four conclusions: (1) in addition to the fact that all of the 19 periodicals have their home pages on the web, all have developed fairly strong link reference abilities as well. The title of the periodicals and the latest research developments are all placed at the front of the search page, with the related contents being given in the following pages in descending order according to time and relevance; (2) home pages of the core periodicals are well designed, keep to certain standards, and facilitate easy access to the required information with clearly written instructions to contributors and strong secondary linking ability; (3) some of the periodicals have high popularity as reflected by the great total number of entries but are weak in other aspects such as linking ability; (4) some periodicals provide searches for highly time-related research achievements at the most conspicuous locations on the first page for keyword searching and guide the user to further readings in "GOOGLE SCHOLAR." However, if the full text is required when using this service, one has to register for access. By doing so, the librarian provides himself an opportunity to expand his range of service.

3 THE IMPACT OF HIGH FREQUENCY WEB WORDS ON THE TECHNICAL TERMS OF TRADITIONAL LIBRARY SCIENCE

3.1 Do libraries need a new definition under the impact of the web?

However precisely and accurately the definitions of library are given in conventional dictionaries of library and information sciences, they seem to become increasingly less so given the current upsurge of Internet technology. Many users of library websites simply go straight towards a target without fussing about the so-called definitions. For instance, traditional dictionaries could define libraries as "social devices," "containers," "venues," "institutions," or "social agencies" on various grounds. The more generally accepted definition might be: "Libraries are places to keep data on human thoughts expressed in different languages, whether in written form or in some other ways" or "social agencies that store accumulated knowledge in written form, which is passed on to groups or individuals with the help of the librarian," etc.

Libraries are known to be institutions whose development has kept pace with that of the times. The current upsurge of the Internet provides them with immense opportunity and space for further development. This development will always be understood in a dynamic way regardless of the form in which it exists. It is also necessary to define it in terms of a distinct time factor. This is similar to the fact that anyone who is knowledgeable

about a subject will be regarded as an expert regardless of capability of writing in the traditional way with a brush-pen, as in the past, or using a fountain pen at all, a skill which will probably disappear in the Internet age. In other words, it is time that the traditional definition of library was revamped.

3.2   Analysis of the Rate of Web Appearance of Traditional Dictionary Entries

It is of far-reaching significance to explore the rate of appearance of traditional dictionary entries in search engines by applying the traditional classification method typical of the frequently used *Library Science Dictionary* compiled by Wu Xuezhen and Zhang Nianhong (1989). Limited by space, this paper will omit the section containing the names of China's bibliophiles and bibliographers. At the same time, of all the library terms, only the first 5 that appear most frequently in the search engine are chosen. Based on the analysis approach applied to Tables 2.1 to 2.11, identical terms found by the search engine are analyzed in terms of the "number of lexical units (entries)," "number of relations found," and "number of supporters" respectively in units of 10 entries per page; entries whose appearance is less than this are omitted. The results are given in Tables 2.1 – 2.11 as below:

Num	Related words	Total lexical unit	Related words	Total relations found	Related Words	Total Supporters
				Table 2.1 **For term "Library Science"**		
1	Library history	810	Library history	10	Library history	1
2	Library futurology	810	Library futurology	6	Library futurology	0
3	Library science	790	Library network	0	Librarianship study	0
4	Library network	780	Librarianship study	0	Library futurology	0
5	Librarianship study	770	Library futurology	0	Library network	0

Num	Relation words	Total lexical unit	Relation words	Total relations found	Relation Words	Total Supporters
				Table 2.2 **For term "Library Books Collection Building"**		
1	Information	980	Journal	6	Journal	8
2	Magazine	910	Information	5	Purchase	8
3	Purchase	890	Datum	5	Information	1
4	Datum	860	Purchase	5	Datum	1

| 5 | Journal | 850 | Magazine | 4 | Magazine | 1 |

Table 2.3
For term "Library Management of Reader Services"

Num	Relation words	Total lexical unit	Relation words	Total relations found	Relation Words	Total Supporters
1	Loan	890	Loan	7	Free newsroom	2
2	Reading public	860	Free newsroom	7	Reading public	1
3	Free newsroom	850	Copy	6	Copy	0
4	Copy	840	Reading public	0	Loan	0
5	Entry	630	Entry	0	Entry	0

Table 2.4
For term "Library Catalog"

Num	Relation words	Total lexical unit	Relation words	Total relations found	Relation Words	Total Supporters
1	Writer	840	Catalog	10	Make catalog	2
2	Catalog	810	Make catalog	6	Catalog	0
3	Entry	760	Entry	5	Writer	0
4	Make catalog	640	Writer	0	Translator	0
5	Translator	540	Translator	0	Entry	0

Table 2.5
For term "Book Classification"

Num	Relation words	Total lexical unit	Relation words	Total relations found	Relation Words	Total Supporters
1	Systematize	920	China book methodology	10	Systematize	1
2	Kind	800	Dewey Classification	6	Cataloging system	1
3	Dewey Classification	630	Systematize	5	Taxon	1
4	China book methodology	770	Taxon	2	Dewey Classification	0

| 5 | Cataloging system | 630 | Cataloging system | 0 | China book methodology | 0 |

Table 2.6
For term "Bibliographic Science"

Num	Relation words	Total lexical unit	Relation words	Total relations found	Relation Words	Total Supporters
1	Subject	890	Subject	9	Bibliographic	1
2	Pemmican	800	Private collection	8	Subject	1
3	Books and periodicals index	730	Bibliographic	7	Pemmican	0
4	Bibliographic	720	Pemmican	6	Private collection	0
5	Private collection	610	Books and periodicals index	1	Books and periodicals index	0

Table 2.7
For term "Chinese Reference Books"

Num	Relation words	Total lexical unit	Relation words	Total relations found	Relation Words	Total Supporters
1	Handbook	880	Apparatus	10	Instrument	5
2	Yearbook	870	Dictionary	9	Yearbook	2
3	Apparatus	860	Instrument	9	Handbook	1
4	Dictionary	850	Yearbook	7	Dictionary	0
5	Instrument	810	Handbook	5	Apparatus	0

Num	Relation words	Total lexical unit	Relation words	Total relations found	Relation Words	Total Supporters
				Table 2.8 **For term "Document Study"**		
1	Document	870	Collected papers	10	Collected papers	5
2	Publications	850	Icon	9	Publications	2
3	Collected papers	830	Publications	5	Document	1
4	Document retrieval	810	Document	4	Document retrieval	1
5	Icon	800	Document retrieval	0	Icon	1

Num	Relation words	Total lexical unit	Relation words	Total relations found	Relation Words	Total Supporters
				Table 2.9 **For term "Informatology"**		
1	Poop	920	Find language	10	Data	4
2	Data	900	Poop	8	Informatology	2
3	Informatology	790	Informatology	8	Poop	0
4	Find language	640	Data	5	Working	0
5	Working	540	Working	0	Find language	0

Table 2.10
For term "Chinese Book Stock and Cataloging Acquisitions"

Num	Relation words	Total lexical unit	Relation words	Total relations found	Relation Words	Total Supporters
1	Memoir	880	Old book	6	Memoir	
2	Old book	640	Twenty four mo.	5	Old book	
3	Twenty four months	610	Memoir	5	Twenty four months	
4	Biographical sketch	560	Rare binding		Biographical sketch	
5	Rare binding	340	Biographical sketch	0	Rare binding	

Table 2.11
For term "Chinese book and library history"

Num	Relation words	Total lexical unit	Relation words	Total relations found	Relation Words	Total Supporters
1	Book	900	Book	7	Cover	1
2	Collected papers.	890	Collected papers	0	Book	0
3	Edit	880	Edit	0	Collected papers.	0
4	Cover	860	Cover	0	Edit	0
5	Copy	760	Copy	0	Copy	0

Notes:

(1) **Number of entries:** With the most frequently searched term being "information", amounting to 980 entries, the top ranking 11 terms include "Information," "Systematize," "Poop," "Book", Subject", "Memoir," "Handbook," "Document," "Author," "Library history," and "Entry" in descending order.

(2) **Number of relations found:** The entry with the greatest number of relations found (10) is "Library history," to

be followed by "catalogue" and "China Book Classification" etc. The top 11 in this category include: "Library history," "Catalogue," "China Book Classification," "Reference index," "Collected papers," "Retrieval language," "Subject," "Book lending," "Books," "Periodicals," and "Classical documents".

(3) **Number of Supports:** The entry with the greatest number of supporters (8 in all) is "Periodical," and the top ranking entries in this category include: "Periodical," "Instrument," "Collected papers," "Data," "Free newsroom," "Cataloguing," "Library history," "Classification," "Bibliography," and "Cover".

The analysis also indicates that dictionary terms frequently appearing in web-searches, such as "library network" in the library science, "information" in library collection building, "copying" relevant to library's reader services, "images" in philology, and "retrieval language" in the information science enjoy high click rates on the web search platform.

4   ANALYSIS ON WEB SEARCHES FOR META-LIBRARY TERMS

4.1 Comparison between vocabularies of traditional dictionaries and the meta-library

This analysis gives the conclusion that currently popular terms rank high on the list, among them "network," "search," "copyright," "printing," "media," "edition," "systems engineering," "coding," "news coverage," "CD ROM," etc. At the same time, non-library science terms such as "network" and "computer" are also frequently searched, indicating that sciences on information, computer, and library tend to overlap in certain aspects. In relations found, terms like "information pollution," "subject librarian," "globalization," "mass communication," "information bank," and "copyright" have become high frequency words, while in dictionaries on libraries published a decade ago these terms were not to be found. The high click rates and high frequencies of appearance of these terms as revealed by the current statistics indicate that it is imperative that the traditional dictionaries in libraries be expanded for greater coverage.

It is worth mentioning that "CD ROM" topped the supporter/sponsor linking, and "printing," "network," "systems engineering," and "digital library" also attracted more financing from supporters. Businesses represented by their respective keywords in search links for investment highlights tend to be economically promising. For instance, although taking the last (the 20th) position in the entry listing, the term "mass communication" received the third greatest number of financial sponsors, becoming one of the most promising search terms. It may be safe to say that these highly popular terms receiving strong technical support will constitute an essential part of the revised library dictionaries and digital library platforms in the days to come.

4.2   Summary of the analysis of library-related terms

On the virtual information expressway, the originally non-mainstream library terms are gaining popularity due to frequent visits by readers. Readers' behavior patterns in using the search engines accelerate the convergence between non-library and library terms, which will inevitably give rise to derived terms. Based on the above-mentioned high frequency words in web searches, sorting and analyzing the meta-library terms offer a clearer clue to the search pattern. An analysis on the frequency at which these "hyped up" words are used will play an active and realistic role in assessing the trend of development and in vitalizing the library services.

It is suggested that libraries should take advantage of these prevailing terms in adjusting and establishing their own digital platforms and in revising their dictionaries on library science. Such terms should be selected and collected on the basis of standardization to provide the readers with greater access to library resources, thus creating new opportunities for library development.

5   REFERENCES

Chinese Academy of Social Sciences Library (2004) *A Guide to the Humanities & Social Science Core Journals in China*, Beijing. China: Bibliometrics Research Office, Chinese Academy of Social Sciences & Peking University Press

Li, Z. (2003) *Skills in Using Search Engines*. Retrieved April 20, 2007 from the World Wide Web: Http://lib.sdut.edu.cn

Wu, X. & Zhang, N. (1989) *Library Sciences Dictionary*. Beijing, China: Beijing Haitian Publishing House