Visual Exploration Research in the Field of Library and Information Based on Co-occurrence Analysis

Hao Wang

Heilongjiang Bayi Agricultural University, Daqing, 163319, China

Abstract: Towards current modern society, our country's library and information science is not only rapid development, but it also has won the favor of many researchers in more than a decade of development. Due to new technology elements and methods are widely used in the discipline of library and information science, cause society our country's library and information science has entered a new stage with the development of information industry. This paper is based on the research theory of co-occurrence analysis, clustering co-occurrence analysis of references, social network co-occurrence analysis from researchers and research institutions of the field of library and information science, to have more precise and in depth research on the research focus of the library and information science, the authors themselves and the research institutions circumstances over the past few years. At the same time, the theory is used to analyze the effects and analysis of accurate data, to ensure that researchers can analyze the theory of library and information science for a long time.

Keywords: Co-occurrence analysis, *Field of library and information*, *Visualization*, *Exploration research*

1 Introduction

Since 1998, worldwide informatization and globalization have developed that have a huge impact on the research of variety of professions, our country's information science is also deeply affected by the development of informatization. In other words, we urgently need to understand the impact of the information science on the development of informatization, for example, changes in the research direction of the library and information science, changes in research content, research trends and current most popular research topics, etc. In addition, what changes have been brought to the information science caused by realization of sharing of information resource? These are notable research contents[1]. Co-occurrence analysis method that quantifies co-occurrence information in various information carriers, this is not only inherits the advantages of qualitative and quantitative analysis. This also overcomes low operability and strong subjective factors of the shortcomings of qualitative analysis; it can be more impressive, precisely excavate implicit or potential useful text knowledge, and discover the close relationship between the research objects, reveal the changes in the subject structure or point of thematic research of represented by the research objects. With the aid of computer technology, co-occurrence analysis with the conciseness of its method and the reliability of the analytical results, become an important approach and tool to support the research process of information content analysis.

2 Significance of research

With the rapid development of the information industry, the library and information science is constantly developing and changing in the process of information industry, more new research techniques and research methods are continuously integrated into them, the research on the library and information science has been improved in breadth and depth, let the library and information sciences embark on a new stage of historical development. Facing the fast-growing

information society, in order to quickly understand the changes of real-time hotspot research and subject structure and also research content development trends, etc. High requirements were placed on the staff of library and information, it is necessary to study the trends and structural changes of information science development from innumerable published papers published in every year[2], it is difficult to achieve hotspot content extraction if relying on traditional reading methods. It is not only because the amount is too large, but also because of the quality of the information. Therefore, the development of the collection of native network resources and the digitization of collection resources are emerged. In recent years, common word analysis and social network analysis have provided a new perspective and method, they are quantitative analysis. The application of these methods can compensate for the deviation caused by the subjective judgment of qualitative analysis, can also make up for deficiency of professional background of relevant research, through accurate data, the scientific method resist the development of the objective reflection subject.

3 Field research based on keyword co-occurrence

In the 1970s, the method of common word analysis was first proposed by a French bibliographer. Common word analysis has gone through a long process of development, confirmed and supplemented by numerous studies and experts in the process of development. After being first proposed, it has experienced three stages of development. The first is the inclusion index and the adjacent index stage, second is the common word strategic coordinate stage and last is the creation of the database stage. Propose of the inclusion index and the adjacent index, the research of dynamic changes in aquatic products is applied as the first. At last, it was applied to the common word analysis in the library and information science. The main role of the inclusion index and the adjacent index are to show the strength of the relationship between the research object or target of detection. Followed by 1991, researcher represented by CALLON has proposed the equivalent coefficient to detect the correlation value between the targets. Prior to this, CALLON also proposed two concepts of proximity maps and inclusion maps, mainly through by clustering keywords, change to the same pattern as a network map, the keywords are shown, comparison of network

maps of different types and periods, show the history of process and development of research of the information science[3]. The strategic coordinates of the common words were first proposed by researcher led by LOW in 1998, common word co-occurrence was applying in the process of development, the strategic coordinates are mainly to study the relationship and influence degree of the target in the research field[4]. Finally, Kostoff proposed the database content structure analysis method in 1995, the scope of application of database analysis is relatively broad, can be applied to research of various university courses, it is not only limited to the library and information science when using database research at the same time. Participants are not only researchers and experts of the library and information science, but also researchers and experts involved in different fields. Database analysis is mainly to study a large amount of digital text information, database analysis has many advantages, and for example, the operating notes of the database are simple, controllable, flexible, and easy to understand.

The main principle of the common word analysis method is through the same meaning or similar keywords, adding if it appears in a single document, judging the strength of connection and relationship of two keywords according to the number of occurrences. If there are more times in the same article earlier that shows the closer relationship between these two keywords, the stronger the connection between these two keywords. Therefore, when the common word analysis method is using that is based on the number of times the keyword appears in the same article, building a clustering tree figure or co-occurrence matrix or lexical matrix; intuitively understand the subject content and central idea represented by keywords based on the obtained matrix, it can also be said to be the structural theme of the research content[5]. This paper mainly uses the research method of common word co-occurrence, research on the trend of development, hotspot content and subject structure of the information science. The research method has also been used to study information retrieval, discovering changes in the subject of information retrieval, combined with multidimensional scaling analysis and research in the field of information science.

Overall, the research result of our country in the field of information science is lagging behind foreign countries. Our country research scope is relatively centralized, mainly distributed in the fields of library science and

information science, researchers are relatively few. For example, in terms of visual research, Yang Ying's research in the medical field is based on common word analysis. Li Ying has studied the development trend of domestic competitive information, the research content and development direction of our country's library and information science, real-time research hotspots, etc., the representative figures and works are done by Li Ying. Professor Qiu Jun Ping conducted a statistical analysis of the keywords of research papers in the field of library and information of our country by the method of common word analysis. Li Chang Ling is based on the high-frequency keywords of the five year master's thesis in the field of library and information, using common word analysis and cluster analysis, figures out the differences between library and information science research hotspots. In addition, the common word analysis method is also used in other fields, for example: nanotechnology, knowledge services, international biological sciences and engineering, Chinese herbal medicine, human genome, literature and other fields[6].

4 Visualization of library and information science based on co-occurrence analysis

4.1 Database selection

This study mainly selects some China journal databases as research objects, it mainly includes library and information work, library science research, information science, information theory and practice, library magazines, libraries, modern library and information technology, information journals, China books journals, national library journals, etc. these are 10 major journals for research. Statistics on all documents downloaded from these journals between 2013 and 2017, a total of 31,660 documents were retrieved, save this data in a document format as a reference for later research and analysis data[7]. The results of the receipt data are shown in Tables 1 and 2:

Table 1. Statistics of publications in the library and information science 2013 to 2017

Posting Time	2013	2014	2015	2016	2017
Number of Post	6590	6808	6523	5960	5779
Ratio of Post	20.8	21.5	20.6	18.8	18.3

Table 2. Statistics on the volume of journal and magazine of library and information science from 2013 to 2017 (partial)

Serial Number	Name of Journal/Magazine	Number of Post	Percentage %
1	Library and Information Work	5321	16.40
2	Library Science Research	2600	8.03
3	Information Science	1994	6.16
4	Information Theory and Practice	1859	5.74
5	Library Magazine	1561	4.82
6	Library	1488	4.59
7	Modern Library Information Technology	1330	4.11
8	Journal of Information Science	720	2.22
9	China Journal of Books	527	1.63
10	Journal of National Library	481	1.49

According to the data in Table 1 and 2, the amount of literature published and the amount of literature downloads in each journal are different. Specifically, it includes the highest volume of library and information work, and the lowest volume of journal of national library, through the analysis and comparison of these data, it can be found that the research situation and bias of China researchers in this field in the past five years.

4.2 Data sort out

To avoid the impact of the number of keywords on this study, this study uses manual methods to sort out keywords, keywords with the same meaning and similarity one are combined into one, or keywords that have the same meaning but different ways of expression are summarized into one. Refine keywords to improve the accuracy of literature search and the extensive of content.

4.3 Result of experimental

4.3.1 Structural analysis of the field of library and information in the past five years - co-occurrence analysis of keyword

The number of documents and publications in these journals is relatively large in every year, according to the number of publications and searches in the past five years, record the number of publications and the number of searches for each year in unit of year, in order to have much better understanding of the research results of library and information science. Then, categorizing and classifying keywords that frequently appear in the literature search to carry on research and analysis in the field of library and information science.

4.3.2 Statistics of high frequency keyword

Filter these high frequency keywords by organizing and summarizing the high frequency keywords that appear in journal searches, sort by the number of occurrences of keywords at a time. Draw into a table, then, select the top 60 keywords in the table to be considered as the research object. It is convenient to study this intuitive feeling of the research status and content of the field of library and information, as well as the research direction and theme changes in recent years.

4.3.3 Construction of lexical matrix

Through the table analysis of the occurrence of high frequency keywords every year, the visual performance in the field of library and information is not obvious, need to import these data of keywords into SPSS, analysis and clustering induction by that software, form a high frequency keyword lexical matrix. Visualize the research situation in the field of library and information through the analysis of the high frequency keyword matrix. Table 3 shows the high frequency keyword lexical matrix in the field of library and information in 2013 (partial).

Table 3. High frequency keyword lexical matrix in the field of library and information in 2013 (partial)

Number of Literature / Keyword	1	4	5	6	7	8	9	10	11	...
Library	0	0	1	0	0	0	0	1	1	...
University Library	0	0	0	0	0	0	0	0	0	...
Digital Library	0	0	0	0	0	1	0	0	0	...
Information Service	0	0	0	0	0	0	0	0	0	...
Public Library	0	0	0	0	0	0	0	0	0	...
Resource Sharing	0	0	0	0	0	0	0	0	0	...
Library Science	0	0	0	0	0	0	0	0	0	...
Knowledge Management	0	0	0	0	0	0	0	0	0	...
Competitive Information	0	0	0	0	0	0	0	0	0	...
Digital Resource	0	0	0	0	0	0	0	0	0	...
Information Resource	0	0	0	0	0	0	0	0	0	...

4.3.4 Clustering analysis

SPSS19 statistical software is selected for this paper, the method of intra group connection of the software is applied, use the ochiia coefficient as a metric, then, clustering the data from past five years into a tree structure. Through the analysis without keywords of past five years research on circumstances of research results and content changes in the field of library and information. Analysis of clustering of high frequency subject words in the field of library and information in 2013, specifically, it can be divided into three aspects, the results of analysis are as follows:

(1) Library information organization and information retrieval. Library information organization and retrieval can be summarized into two major points, respectively "information organization and retrieval under the network environment" and "information resource sharing research", etc. The specific keywords include information organization, information retrieval, digital library, etc. Library information and library information retrieval always are popular research topic in the field of library and information research and library and information researchers are also keen to study the research of information sharing. Internet technology presentations and applications, realizing resource sharing is no longer a difficult task. A joint established through the internet is resource service website, facilitate resource users to access and retrieve information.

(2) Content of the book information service. The content can be summarized into three major points, respectively "analysis of reader demand", "information digital resource service" and "research on the service mode of librarians of university library science", the content covered include service mode, shared space, database and information services, etc. Research on the content of book information services, can optimize the service system, have a better meet the information needs of resource users, transforming information services from passive acceptance to initiative start off. It is more convenient for researchers in the field of library and information research.

(3) Search engine indicators and enterprise information competition. This research content can be divided into two major points, there are "enterprise competitive information" and "search engine evaluation index system". The specific search keywords include e-commerce, knowledge management, search engine, information competition, etc. Analyze and acquire information through appropriate means, information collected through the scope permitted by law, improve the core competitiveness of the enterprise itself. A high performance search engine can bring a lot of convenience to the enterprise, and provide more opportunities and platforms in the development of the enterprise, collect more information for the enterprise, improve the overall strength of the enterprise[8].

5 Summary

With the rapid development of information technology and internet technology, the number of researcher involved in the field of library and information research is gradually increasing. This is adhering with the requirement of the progressive of era, research on the field of library and information can make it easier to achieve information resource sharing. Improve the quality of information resource and improve the quality of information service, and realize the essence of information resources. Modernize and visualize research content and processes in the field of library and information, to have better study research trends and content changes in recent years, understand the changes of resource demand, better serve the society, and meet the changing needs of the society's growing information needs and quality requirements.

References

[1] Yan DM, Yan H, Zheng XY, Dai WH. Research on Disciplinary Dynamic Knowledge Structure Based on Time-Keyword Co-occurrence Analysis - Taking Library and Information Science in Foreign Countries as an Example[J]. Library and Information Service, 2017, 61(12):21–31.

[2] Liu J, Han Y. Visualization and Comparative Analysis of Research Hotspots in Library and Information Science at Home and Abroad[J]. Information Science, 2014, 32(12):53–60.

[3] Qiu JP, Li XT, Dong K. The Development, Evolution and Innovation of Visualization Research in the Field of Pictures[J]. Library and Information Service, 2014, 58(13):125–31.

[4] Zhao RY, Wu SN. Research Progress in Information Visualization Analysis Methods in Library and Information Fields[J]. Information Studies Theory&Practice, 2014, 37(06):133–8.

[5] Chen BK. Research on Visualization of Subject Knowledge[D]. Wuhan University, 2014.

[6] Chu JW. Common Word analysis of knowledge management methods in the field of library and information in China[J]. Journal of Information, 2012, 31(11):45–50.

[7] Yang ZY. Marketing discipline map[D]. Donghua University, 2011.

[8] Yang Y. Research on Disciplinary Structure Visualization Based on Common Word Analysis[D]. China Medical University, 2010.