

# Exploration on Frontier Research Field of Informatics

Mingwei Sun\*

School of Computer  
Tonghua Normal University  
Tonghua, China  
E-mail: \*dbc9999@163.com

**Abstract**—In order to determine the frontier research field of information science, the literature collected in the field of information science from 2008 to 2017 was statistically analyzed by searching the (CNKI) database of Chinese academic journal network. According to the statistical results, the paper analyzes the literature of information science in the year of sampling by using hypothesis test method in statistics, and probes into the frontier research field of information science to understand the latest development and prospect of information science.

**Keywords**—Information science; Frontier domains; CNKI database; Hypothesis test

## I. INTRODUCTION

What is the frontier? According to the general understanding of the academic circles, it is the main foundation of cultivating the ability of discipline innovation that the field is forward-looking, forerunner and exploratory, and has a great influence and leading role on the future development of the discipline. The frontier field and the hot spot domain have the close relation, but it does not equal the hot spot domain. Some hot spots are only a topic of concern, or a fashion, not forward-looking, leading and exploratory, not with the essential characteristics of the frontier. Therefore, it is necessary to make full use of the knowledge and judgment of field experts, the ability of prediction, and the method of literature measurement to identify the frontier areas from many and complex topics. It is of great significance for the development of information science to grasp the frontier of information science. For this reason, this paper will make a statistical analysis of the papers in the field of information science included in the CNKI database from 2008 to 2017, in order to explore the frontier of the field of information science research. Before that, there were also many scholars keen on the research field. But the analytical methods are very different. Such as chi-square test, word frequency statistics, content analysis, ACA method and so on. This paper uses the hypothesis test in statistics to analyze the sampling papers. The author thinks that the advantage of this method is that it does not need to use specific software to analyze and avoid the trouble of operation. Second, the mathematical formula is adopted directly [1-2], the method is simple, the train of thought is clear, and the reader is easy to refer to.

## II. PRELIMINARY PLANNING

One of the most important arguments to convince college or university administrators is to contend that the library needs to better support the mission of the school. What is at the heart of any college or university? The answer is student learning. Scott Bennett<sup>6</sup> makes the case that academic library managers who are planning new buildings should utilize learning theory to design user spaces. Too often librarians know more about what does not function (for example heating/air conditioning problems, unwelcoming entrance, limited wiring, and not enough space for technology). Many projects are designed around trying to design a library with all those problems fixed. Bennett argues that library managers should begin planning way before any funding is made available and that managers should think about assessment, how the success of the building will be measured before any design is accomplished. Working with faculty and administrators, library managers need to decide how learning can take place in the library and how that learning can be measured. The early stages of this concept were the development of learning commons in schools. Even public libraries and school libraries now offer some type of space for collaborative learning with easy access to technology [3-5].

## III. BRIEF INTRODUCTION OF RESEARCH METHODS

In this study, the paper of information science in CNKI database is taken as the research object, and the time range of retrieval is set to 2008-2017. In the process of retrieval, in order to ensure the recall and precision, in the second step of standard retrieval, the scope of the key words is selected in the second step of the standard retrieval, that is, the key words appear in the title, key words and so on. The key words are determined by the following ways: retrieval through CNKI database example: input the following retrieval formula: title = ((information science) or(informatics orinformative) or (information) or (information science)) and keyword = search term, and the relevant literature obtained from it. In the third step of standard retrieval, the results of retrieval were sorted according to the time of publication, and the published year was grouped. On the basis of statistical analysis of the retrieval results, the data were analyzed by using the hypothesis test method in statistics. Used to determine the frontiers of information science [6-7].

#### IV. RETRIEVAL SUMMARY

Through the above two methods, 18 search words were obtained, and the 18 keywords were searched in CNKI database, and the results were shown in Table 1.

TABLE I CNKI DATABASE RETRIEVAL RESULTS

Search word	Number of papers per year										
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total
Information theory	7	6	1	2	1	12	15	22	29	33	128
Concept map	0	0	0	0	0	0	0	0	0	0	0
Information economics	5	2	2	1	1	2	1	0	0	2	16
Information research	92	108	88	83	101	117	123	121	127	155	1115
Information resource management	4	6	4	6	7	4	2	8	4	1	46
Knowledge management	35	37	29	26	21	12	13	17	13	7	210
Research on Network user behavior	0	0	0	0	0	0	0	0	0	0	0
Digital Library	20	22	26	28	15	18	17	15	12	6	179
Information policy and law	0	0	0	0	0	0	0	0	0	0	0
Information retrieval	12	15	9	8	7	5	17	1	4	2	80
Information visualization	4	4	5	3	2	2	5	7	8	10	50
Knowledge organization	3	4	2	3	1	2	4	3	4	3	29
Knowledge map	3	2	5	1	4	2	0	4	0	1	22
Intelligent agent	0	1	1	1	0	2	3	3	3	4	18
Information construction	3	5	3	1	1	3	0	0	1	0	17
Metadata	3	3	3	4	3	2	0	4	6	4	32
Ontology	2	5	1	1	0	3	3	2	1	0	18
Intelligent information processing	1	0	1	1	1	1	2	2	1	1	11

(Note:As a result of Concept map,Research on Network user behavior And Information policy and law The three key words are CNKI No hit papers in the database,Therefore, this is not discussed in the following analysis)

#### V. DATA ANALYSIS

How to determine which of the above key words correspond to the fields that represent the pre-information field in the number of papers, There should be two features at the same time.: first, There should be a slight increase in the number of papers. First, Since frontier issues are forward-looking, leading questions, that, The amount of information science research in this field should be gradually increasing. Second, The upward trend should not be significant., If there is a significant increase, Then the problem has become a relatively mature field. Second, The relative proportion of the number of studies should be small. If the amount of research is huge.

The problem has become a hot area. Below, The author will analyze the CNKI Quantitative characteristics of Information Science papers retrieved in Database, On this basis, the frontier problems of information science are judged.

CNKI An Analysis of the trend of changes in the number of Information Science papers in the Database the author puts forward the first five years of the above ten years(2008in that year,2009in that year,2010in that year,2011in that

year,2012Year)Total number of papers and the latter five years(2013in that year,2014 in that year, 2015 in that year, 2016 in that year,2017Year)The total number of papers is compared. If the "total number of papers in the last five years" is significantly larger than the "total number of papers in the previous five years" in a certain field of research, there is a tendency to increase the number of studies in that field.; Vice versa. If the "total number of papers in the last five years" in a given field is equal to "the total number of papers in the previous five years", Further analysis is needed on the total number of papers in this field and the trend in the past ten years.

##### A. The basic steps of hypothesis testing

The total number of papers in the last five years and Comparison of "Total number of papers in the first five years", The author95%At the confidence level of, Matching table1The key words in this paper are tested in order of hypothesis. The basic steps for hypothesis testing are as follows:

Make assumptions: Original hypothesis $H_0: \mu_1 = \mu_2$

Alternative hypothesis $H_1: \mu_1 < \mu_2$

Statistics and distribution:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2} S^2}} \sim t_{(n_1 + n_2 - 2)}$$

$$S^2 = \frac{1}{n_1 + n_2 - 2} ((n_1 - 1)s_{n_1}^2 + (n_2 - 1)s_{n_2}^2)$$

Significant level  $\alpha$ , Check t Critical value obtained by distribution table, From  $t_{\alpha}(n_1 + n_2 - 2)$  and  $(\mu_1 - \mu_2)$  The confidence of the 100  $(1 - \alpha)\%$  The upper confidence bound of the:

$$U_{\alpha} = \left( \bar{X} - \bar{Y} \right) + t_{\alpha} \sqrt{\frac{(n_1 + n_2)S^2}{n_1 n_2}}$$

- Statistical decision: If  $T > t_{\alpha}$ , Refusal  $H_0$ ,

Accepted  $H_1$ ; If  $T < t_{\alpha}$  Accepted  $H_0$

Here only the table 1 The key word in the Information theory as an example. Five years after its establishment (2004 in that year, 2005 in that year, 2006 in that year, 2007 in that year, 2008 Year) Related "Information theory" the mean and variance of the number of papers in the  $\bar{X}_1, S_1^2$  Five years before its establishment (1999 in that year, 2000 in that year, 2001 in that year, 2002 in that year, 2003 Year) The mean value of the number of papers is  $\bar{Y}_1, S_1^2$  The joint variance is  $S^2$ , Table 1 Middle data calculation,  $\bar{X}_1 = 9, S_1^2 = 18.5, \bar{Y}_1 = 2.4, S_1^2 = 5.3, S^2 = 11.9, T = 3.0252$ . Check distribution table available, In 95% Under the confidence level of  $(\alpha = 0.05)$ , When the degree of freedom is  $t_{0.05}(5+5-2)$  Time,  $t_{0.05} = 1.8596$ .

As a result of  $T > t_{0.05}$ , So reject the original hypothesis, Acceptance of alternative assumptions, That is CNKI Database about the Information theory", The latter five years (2008 in that year, 2009 in that year, 2010 in that year, 2011 in that year, 2012 Year) The number of papers in the previous five years was significantly larger than that in the previous five years, The significant level is 95%. Visible, The key to comparison lies in the TIs the value greater than  $t_{0.05}$ , That is, the TIs the value greater than 1.8596.

**B. Information science search word T Value table.**

CNKI of the various key words of information science in a database T Values such as tables II As shown. Observation table II, You can find it. These keywords T Values can be roughly divided into three categories: The first is  $T > 1.8596$  (significant increase trend); Category II  $T > 0$ , But  $T \leq 1.8596$  (Not significant Increasing trend); Category III  $T < 0$  Negative growth trend

TABLE II EACH KEY WORD T VALUE TABLE

Search word	T value
Information theory	3.0252
Information economics	-0.975
Information research	2.9792
Information resource management	-0.8074
Knowledge management	-1.5189
Digital Library	-1.285
Information retrieval	-0.8687
Information visualization	1.4141
Knowledge organization	0.6086
Knowledge map	-1.132
Intelligent agent	1.6325
Information construction	-1.316
Metadata	0
Ontology	0
Intelligent information processing	1.6093

From this, We can draw the following conclusions: at the significant level 95% In the case of Category I: CNKI Five years after Information Science papers in Database (2004 in that year, 2005 in that year, 2006 in that year, 2007 in that year, 2008 Year) The number of (1999 in that year, 2000 in that year, 2001 in that year, 2002 in that year, 2003 Year) Number of papers, That is, 2008~2017 Year, A significant increase in the number of information science papers in the first category; Category II: The number of papers in the last five years of information science is larger than that in the previous five years, But significant. Sexual insufficiency, That is, within the year of sampling, No significant increase in the number of information science papers of the second category; Category III: The number of papers in the latter five years of information science is not significantly larger than the number of papers in the previous five years, And the sample data also showed a negative growth trend, That is, 2008~2017 Year, The third category of information science papers has at least a negative growth trend. ( $T < 0$ ). In the fourth category, the number of papers in the last five years of an information science paper, although equal to the number of papers in the previous five years, However, the search results for each year are 0 It is not within the year of sampling and is negligible. Finally, according to the classification of the above four categories, To bring the 2008~2017 Years CNKI The key words of information science in the database are summarized in the table III.

TABLE III CNKI CLASSIFICATION TABLES OF KEY WORDS IN INFORMATION SCIENCE IN DATABASE

Category	Search word
The first type of information science search words(significant increase)	Information theory, Information research,
The second category of information science key words(Not significantIncrease)	Information Visualization, knowledge Organization, Intelligent Agent, Intelligent Information processing
The third category of information science key words(significant decrease)	Information Economics, Information Resource Management, Information Retrieval, Knowledge management, Digital Library, Information Construction, Knowledge map
The fourth category of intelligence search terms (not retrieved as 0)	Metadata, Ontology,

## VI. CONCLUSION

According to the CNKI the retrieval results of database are analyzed, and the frontier research fields of information science are judged.

Table III You can get the second category Not significant Additional intelligence keywords 4 Where the above table is combined with the intermediate computation process in the hypothesis test method 2 In the fields of intelligent agent and intelligent information processing T The value of the test is close to that of the But not greater than It can still be seen as a frontier area. In summary, we can see that the frontier field of information science is a common one. They are: Information Visualization, knowledge Organization, Intelligent Agent, Intelligent Information processing.

## ACKNOWLEDGMENT

This research was financially supported by Research Foundation of Education Bureau of Jilin Province (Grant No.2016238 and Grant No.JJKH20180860KJ) and by the Project of Education Department of Jilin Province College students venture project (Grant No.201708301E).

Information visualization In the field of scientific and technological research, the term "information visualization" is generally applied to large-scale non-digital types. Information resources Visual representation of. The ultimate purpose of information science research is to realize the understanding of information.. The method of information visualization has a powerful understanding of "one picture is better than one word", so information visualization has penetrated into all fields of information science.

Knowledge organization is the process or behavior of revealing knowledge unit, mining knowledge association, and providing effective knowledge or information to users most quickly. These activities are usually carried out by Picture file Profession the personage finished.

Intelligent agent Is a program that collects information or performs services on a regular basis, without human intervention, with a high degree of intelligence and autonomous learning, and can be based on user-defined criteria, It is a prospective and advanced method of network

information retrieval, which is applied to the specific field of network information retrieval.

Intelligent information processing. The technology of automatic recognition of objects, images, speech, characters and so on by computer. Its general process includes: sample collection, digitization of information, preprocessing, extraction of data features, comparison with standard mode. The establishment of intelligent processing mechanism for classifying and identifying information must be based on the basic principle of information science, otherwise it will make the information work fall into the quagmire of numerous and complicated. Based on the system of intelligent processing and the basic principles of information science, the intelligent information processing system can be divided into three layers: information / data layer, logical business layer, and function layer. The basic principles of information science have different levels of application and embodiment in these three levels.

## REFERENCES

- [1] Lai Mao-sheng, King Lin. Investigation and Analysis on the Frontier Field of Information Science. Library and information work, vol.52, pp.6-10, March, 2008.
- [2] Lai Mao-sheng, King Lin. Investigation and Analysis on the Frontier Field of Information Science. Library and information work, vol.52, pp.6-10, March, 2008.
- [3] Qian Feng. Comparison of domestic and foreign knowledge management research hotspots based on chi-square test. Intelligence magazine, pp.56-58, September, 2012
- [4] Yang Wenxin. Explore the frontier research field of information science 2000-2005 Year LIS Statistical analysis of database documents. Intelligence magazine, pp.137-140, June, 2013.
- [5] Lai Mao-sheng, Wang Lin, Yang Wenxin, Li Yunning. The determination and discussion of the Frontier Field of Information Science. Library and information work, pp.15-18, March, 2015.
- [6] Leaf eagle. Comments on the Frontier Research Field of Library and Information Science. Journal of Chinese Library, pp.63-70, July, 2018.
- [7] Guo Yuehua. Probability Theory and Mathematical Statistics. Beijing: Science Publishing House, 2007.