A Bibliometric Study of Search Engine Literature in the SSCI Database

Wen-Jen Yu
Department of Management Information Systems
National ChengChi University, Taipei, Taiwan
Email: yuwenjen@gmail.com

Shrane Koung Chou
Department of Management Information Systems
National ChengChi University, Taipei, Taiwan
Email: kchou@nccu.edu.tw

***Abstract*—This paper investigates the publishing trends of search engine literature as catalogued in the social sciences citation index (SSCI) database in the period 1995-2009. Our findings indicate that (1) the quantity of recent research on search engine study is expanding remarkably; (2) the frequency indices of author productivity appear to abide by Lotka's Law; (3) most research papers on search engine study are generated by multiple authors; and (4) applications of search engine study are most frequent in research areas such as information science, information systems of computer science, and interdisciplinary applications of computer science. Finally, future directions of research on search engine study are considered. According to Bradford's Law, the three zone ratio comparisons are almost equal at $1:7:7^2$, indicating that the data appear to match Bradford's Law. Six core journals in the search engine study are identified and analyzed.**

***Index Terms*—Lotka's Law, Search Engine, Search Engine Studies**

I. Introduction

The SSCI database contains more than 2,000 social science journals and nearly 3,300 kinds of scientific and technical journals. to select relevant information covering about 50 kinds of topics. This study aims to apply bibliometrics to SSCI data library research literature related to the search engine over a period of 15 years (1995~2009), to try to understand this new information technology in social sciences, analyze its characteristics, including growth, and the distribution of literature, and author productivity, and determine whether the situation is in line with Loca's Law, Price's Law, and the square root of the 80/20 rule, and Bradford's Law, to understand the research literature on search engines for social sciences in terms of its growth in literature, creative way, major research institutions and major journals.

II. Introduction of the Search Engine

The Origin of the Search Engines

As the World Wide Web began in 1990, the search tools available for checking information scattered in various computer files included Archie and Gopher. With the rapid development of the Internet, as well as HTTP, using the web. The rapid diffusion of technology, such search tools, has been unable to meet user needs. In January 1994, the first tool for searching classifications could also browse the directory of EINet Galaxy (Trade wave Galaxy) on-line and provided support for Gopher and Telnet searches. In April 1994, Yahoo Web was born, with the number of visits and included links, providing support for a simple database query. This is a directory of early navigation systems, Yahoo's shortcoming was that to collect and update information, it required manual maintenance. During a rapid surge in data message traffic, this was not very effective.

In July 1994, Lycos launched robot-based data mining techniques, and provided search results sorted by relevance. Lycos was the first to provide automatic company summaries in the search page results. Infoseek was another important representative of the period, and provided another important step forward in search engines.

In 1995, a new search engine tools emerged—the one-bit search engine. The first one-bit search engine was developed by students at the University of Washington and called the Metacrawler. Users only needed to send a search request, the search engine was responsible for conversion by the bit after treatment sent to multiple pre-selected independent search engines; it separated the query results from the search engine centralized the processing, then returned the result to the user.

The December 1995 debut of AltaVista demonstrated a number of innovative features, making it quickly reach the pinnacle of search engines. It was the first to support the search in natural language, and was equipped with web-based content analysis, intelligent processing capacity, and the first realization of an advanced search syntax (such as AND, OR, NOT, etc.). AltaVista searched news groups and supported groups, searched for pictures, and provided other landmark features. In the same period, Inktomi, HotBot, and other search engines also appeared.

In August 1997, the Northern Light Co. Officially launched its search engine, which was the first to support

doi:10.4304/jsw.5.12.1317-1322

search results in a simple automatic classification, and offered the largest database of all search engines.

In October 1998, Google was born, and became one of the most popular search engines, with many unique and outstanding features, and a revolutionary new interface.

THE DEFINITION OF SEARCH ENGINE

The term search engine refers to a process—the search for files using specified key words. Key words found in the results are returned and collated into user information. While the search engine is a general-purpose application, the term is often used for specific systems, such as Google, Alta Vista, and Yahoo. A user can use any of these to search through all the information available on the Internet.

Typically, the search engine works by sending web spiders to obtain the largest web site files. Another program, called the index, reads all the recovered files and indexes them, on a word basis. Each search engine uses proprietary algorithms to build the information, and return meaningful information to users.

III. CHARACTERISTICS OF A LITERATURE SEARCH ENGINE

This study covers nearly 15 years (1995 ~ 2009) of the ISI's Web of Science of the Social Sciences Citation Index (SSCI) search and search engine for the analysis of the literature relevant to the subject as the right. The search method sets Topic = ("search engine *") to limit the subject (topic) and retrieved a total of 1,340 articles in the field, made from 1995 to 2009. The amount of published literature and trends for the past 15 years can be found in Table 1 and Figure 1.

From the volume published in the growth trend, from Figure 1 to observe publications from the past 15 years documents the growth of research literature on search engines. While there was a peak in 2009 of 175 (13.06%) over the past four years Jieyou 100 or more. Figure 2, which shows the citations from the literature, shows annual growth similar to that in Fig. 1, and suggests that this is an emerging discipline, and still in its growth stage.

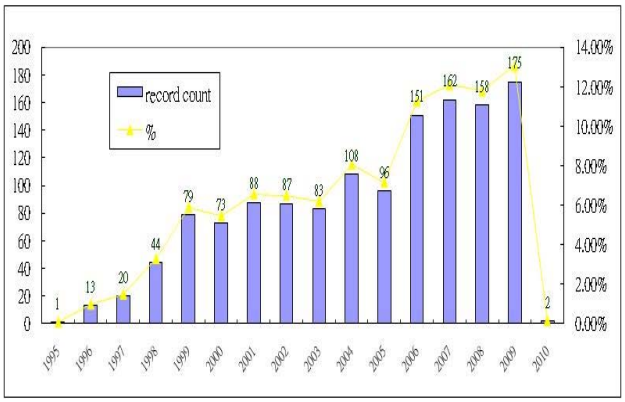

Figure 1. The growth of research literature on search engines over the past 15 years (Source: SSCI database).

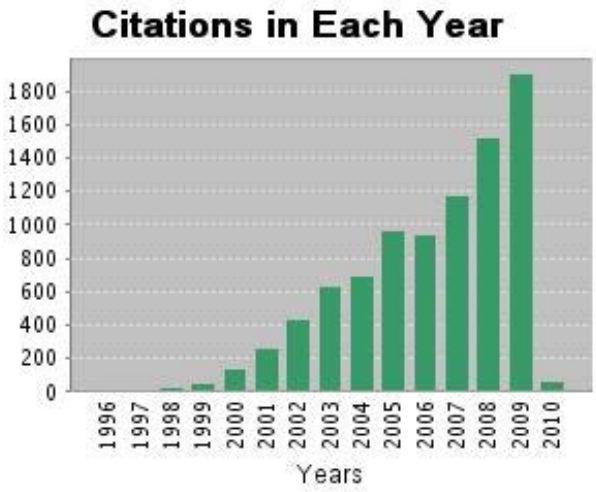

Figure 2. Cited statistics from search engine research literature over the past 15 years (Source: SSCI database).

Table 1 shows the number of documents published from 1995 through 2009, by country. The United States led the way with a total of 559 (41.69%), followed by England with 166 (12.38%), and Canada with 91 (6.79%). Table 2 represents the data in terms of the top ten institutions (in terms of publications). The United States undoubtedly has had the largest research output over recent years, followed by England and Canada.

Table 3 shows the publications by subject area. The top three were information sciences, with 881 (65.75%), computer science information systems, with 659 (49.18%), and the cross discipline application of computer science accounting for 64 (4.78%). The rest offer a very wide distribution of subjects, such as public categories, medical information, communications, business, psychology, education, and legal subjects. This shows the areas affected by the subject and its influence on various aspects of human life.

TABLE I.
THE TOP 10 COUNTRIES IN TERMS OF SEARCH ENGINE LITERATURE DISTRIBUTION FROM 1995 TO 2009

Rank	Country	Publicaton Number	%
1	USA	559	41.69%
2	ENGLAND	166	12.38%
3	CANADA	91	6.79%
4	GERMANY	63	4.70%
5	AUSTRALIA	56	4.18%
6	PEOPLES R CHINA	39	2.91%
7	ISRAEL	36	2.68%
8	TAIWAN	35	2.61%
9	SPAIN	34	2.54%
10	TURKEY	29	2.16%

TABLE II.
THE TOP TEN SOURCES OF SEARCH ENGINE LITERATURE FROM 1995- 2009

Rank	Institute	Publication Number	%	Country
1	WOLVERHAMPTON UNIV	49	3.65%	England
2	PENN STATE UNIV	40	2.98%	USA
3	UNIV WISCONSIN	31	2.31%	USA
4	UNIV WESTERN ONTARIO	25	1.86%	Canada
5	UNIV PITTSBURGH	23	1.72%	USA
6	UNIV TENNESSEE	20	1.49%	USA
7	UNIV SHEFFIELD	19	1.42%	England
8	NANYANG TECHNOL UNIV	18	1.34%	Singapore
9	HEBREW UNIV JERUSALEM	16	1.19%	Israel
9	ULUDAG UNIV	16	1.19%	Turkey
9	UNIV MICHIGAN	16	1.19%	USA
10	MCGILL UNIV	15	1.12%	Canada
10	UNIV ARIZONA	15	1.12%	USA
10	VICTORIA UNIV WELLINGTON	15	1.12%	New Zealand

TABLE III.
SEARCH ENGINE RESEARCH LITERATURE TOPICS FOR 1995-2009. NIAN TYPE OF DISTRIBUTION OF SUBJECTS BEFORE THE TABLE 10

Rank	Subject Type	Amount	%
1	INFORMATION SCIENCE & LIBRARY SCIENCE	881	65.75%
2	COMPUTER SCIENCE, INFORMATION SYSTEMS	659	49.18%
3	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	64	4.78%
4	PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH	48	3.58%
5	MEDICAL INFORMATICS	44	3.28%
6	PSYCHIATRY	41	3.06%
7	COMMUNICATION	37	2.76%
8	BUSINESS	31	2.31%
9	PSYCHOLOGY, MULTIDISCIPLINARY	28	2.09%
10	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	26	1.94%
11	EDUCATION & EDUCATIONAL RESEARCH	25	1.87%
12	NURSING	24	1.79%
13	HEALTH CARE SCIENCES & SERVICES	21	1.57%
14	MANAGEMENT	20	1.49%
15	LAW	18	1.34%
16	PSYCHOLOGY, CLINICAL	16	1.19%
17	SOCIAL SCIENCES, INTERDISCIPLINARY	16	1.19%
18	TELECOMMUNICATIONS	16	1.19%
19	COMPUTER SCIENCE, CYBERNETICS	15	1.12%
20	ERGONOMICS	14	1.04%

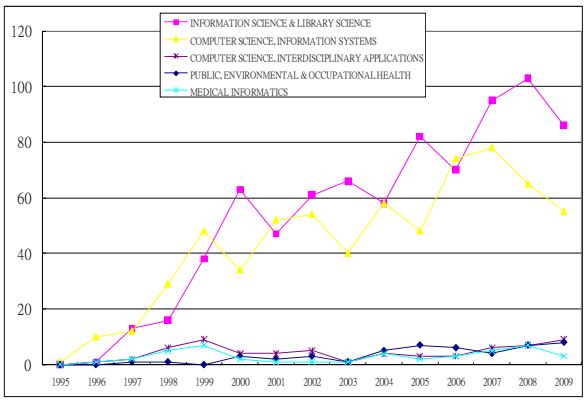

Figure 3. The top five trends in research literature for search engines over the past 15 years.

Based on this information, it seems that the current research literature on search engines is still in its infancy, with the majority of the output in the United States, Canada, and England.

IV. THE AUTHORS OF RESEARCH LITERATURE ON SEARCH ENGINES—A PRODUCTIVITY ANALYSIS

This section investigates the distribution of the authors, apart from Xian analysis of author pages, Loka's Law will be verified.

Table 4 analyzes the data from 1,340 publications in the literature, and calculates by an equal number of collective creation that is every author's contribution to the same degree as the respective statistics, can be drawn for a total of 607 authors; the average person will have written 2.20. The maximum was 47 (1), more than six authors wrote more than 14 people, and there were 280 individual published authors, accounting for 42.13%.

TABLE IV.
AUTHOR CONTRIBUTIONS TO SEARCH ENGINE RESEARCH LITERATURE PUBLISHED IN 1995-2009

Number of published literature	Number of author	Subtotal Number of literature	Cumulative number of literature	% of Cumulative number of literature	Cumulative number of author	% of Cumulative number of author
47	1	47	33	2.46%	1	0.16%
34	1	34	67	5.00%	2	0.33%
25	1	25	92	6.87%	3	0.49%
22	1	22	114	8.51%	4	0.66%
18	1	18	132	9.85%	5	0.82%
14	1	14	146	10.90%	6	0.99%
12	3	36	182	13.58%	9	1.48%
11	3	33	215	16.04%	12	1.98%
10	1	10	225	16.79%	13	2.14%
9	1	9	234	17.46%	14	2.31%
8	6	48	282	21.04%	20	3.29%
7	2	14	296	22.09%	22	3.62%
6	5	30	326	24.33%	27	4.45%
5	14	70	396	29.55%	41	6.75%
4	17	68	464	34.63%	58	9.56%
3	58	174	638	47.61%	116	19.11%
2	211	422	1060	79.10%	327	53.87%
1	280	280	1340	100.00%	607	100.00%

Table 5 lists authors who published more than 18 papers. The first author is Thelwall, M. (England), with 47 articles. Examining the list of the top five authors, England has one and the United States accounts for two. If a careful examination of several literature accounts for the country's ratio, can be found in relation to England, the author's ratio of membership is small, the author's literature in England nationals accounted for the significant rate in the country, apparently a U.S. citizen by a few should come more than England. These authors of information science and computer science in areas such as more.

TABLE V.
THE FIVE MOST PROLIFIC AUTHORS OF SEARCH ENGINE RESEARCH LITERATURE PUBLISHED IN 1995-2009

Rank	Name	Number of literature	% of Total	% of author country	Country	Institute	Subject areas
1	THELWALL, M	47	3.51%	25.27%	England	Wolverhampton Univ	Information Science & Library Science
2	SPINK, A	34	2.54%	38.64%	Australia	Queensland Univ Technol	Information Science & Library Science
3	JANSEN, BJ	25	1.87%	3.07%	USA	Penn State Univ	Information Science & Library Science
4	BAR-ILAN, J	22	1.64%	52.38%	Israel	Bar Ilan Univ	Computer Science, Information Systems
5	NOTESS, GR	18	1.34%	2.21%	USA	Montana State Univ,	Computer Science, Information Systems

Disciplines in the distribution and productivity of the authors of research, you can use the Loca Law (Lotka's Law) to explore. Lotka's Law, also known as "scientific productivity of the inverted square law," means: on x is the number of papers published author of a published literature by dividing the total x2. Analysis using the Lotka Law to verify the applicability of the literature search engine is required to calculate the value of the slope n, constant c value, and the Ke Shi (KS) test determines whether the distribution of (Tsai Ming-yue, 2003). From basic data, nearly 15 years of search engines for only one book documents the author of 20.9%, with the original loca c-value of 60.79% law incompatible, we can see with the original sub-locka law. Using the least squares method to calculate the n and c values, you can further test Loca Law compliance.

TABLE VI.
YOUNG WRITERS AND THEIR PRODUCTION OF TRAFFIC ANALYSIS LITERATURE I (1995-2009)

Number of published literature	Number of author	Subtotal of literture	Cumulative number of literture	% of Cumulative number of literture	Cumulative number of author	% of Cumulative number of author	Number of published literature of Lokta	% of Cumulative number of Lokta
47	1	47	47	3.47%	1	0.16%	1.6721	-2.7832
34	1	34	81	5.98%	2	0.33%	1.5315	-2.4822
25	1	25	106	7.83%	3	0.49%	1.3979	-2.3061
22	1	22	128	9.45%	4	0.66%	1.3424	-2.1811
18	1	18	146	10.78%	5	0.82%	1.2553	-2.0842
14	1	14	160	11.82%	6	0.99%	1.1461	-2.0050
12	3	36	196	14.48%	9	1.48%	1.0792	-1.8289
11	3	33	229	16.91%	12	1.98%	1.0414	-1.7040
10	1	10	239	17.65%	13	2.14%	1.0000	-1.6692
9	1	9	248	18.32%	14	2.31%	0.9542	-1.6371
8	6	48	296	21.86%	20	3.29%	0.9031	-1.4822
7	2	14	310	22.90%	22	3.62%	0.8451	-1.4408
6	5	30	340	25.11%	27	4.45%	0.7782	-1.3518
5	14	70	410	30.28%	41	6.75%	0.6990	-1.1704
4	17	68	478	35.30%	58	9.56%	0.6021	-1.0198
3	58	174	652	48.15%	116	19.11%	0.4771	-0.7187
2	211	422	1074	79.32%	327	53.87%	0.3010	-0.2686
1	280	280	1354	100.00%	607	100.00%	0.0000	0.0000

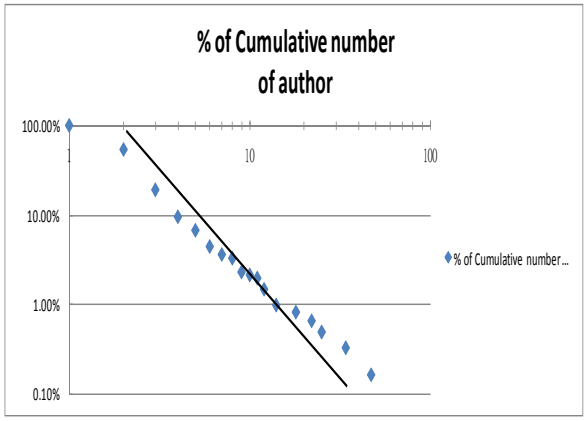

Figure 2. Distribution of authors producing literature.

Using results calculated from Table 6 use the formula with the available slope n value -1.72:

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - \left(\sum X\right)^2}$$

Find the n value, and then use the following formula to obtain c:

$$c = \frac{1}{\sum_{1}^{P-1} \frac{1}{x^n} + \frac{1}{(n-1)\left(p^{n-1}\right)} + \frac{1}{2p^n} + \frac{1}{24(p-1)^{n+1}}}$$

p=18, x=1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17

A C value of 0.4967 can be derived from n = - 1.72; c = 0.4967 can be deduced from f (x) = 0.4967/x1.72 as shown in Fig. 4. The figure under the slash from left to right Naishi theory (expected) line; the blue boxes represent the actual (observed values) distribution.

For the n and c values, the original Loca Law n is about -2, c is 0.6079, showing the distribution of search engine literature by the law fully complies with the original loca. Figure 4 shows that the data distributions did not differ significantly. To test whether the theoretical value is consistent with the observed value, according to the calculated values of n and c, use the push by the formula to calculate the expected value and the accumulated value of the testing history for Ke (KS test).

By the Ke Shi test (KS test), as Table 7 shows the Dmax = 0.0354, and since the number of samples is greater than 35, the critical value of 1.63 / (607) 1 / 2 = 0.066. As Dmax is less than the critical value, it can be deduced that the distribution of productive authors of this study is consistent with the Loca Law—which is the Loca law applied to nearly 15 years of search engine research literature by the distribution of productivity data. Powell suggested a value of Loka's Law n between 1.2 and 3.8 to form a broad Loca Law (Tsai Ming-yue, 2003); this study in line with such a claim, but the search engine field of scientific productivity should be the inverse square of the original law, amended to the inverted square law, meaning that the authors of x number of published papers is published by the number of documents divided by x^3.

TABLE VII.
YOUNG WRITERS AND THEIR PRODUCTION OF TRAFFIC ANALYSIS LITERATURE II (1995-2009)

Literature Number	Observations of author	Observation of the accumulated value of author	Expectations of author	Expected total value of author	Absolute difference
		Sn(X)		F0(X)	|F0(X)-Sn(X)|
1	0.4613	0.4613	0.4967	0.4967	**0.0354**
2	0.3476	0.8089	0.1506	0.6473	0.1616
3	0.0956	0.9045	0.0750	0.7223	0.1822
4	0.0280	0.9325	0.0457	0.7680	0.1645
5	0.0231	0.9555	0.0311	0.7991	0.1565
6	0.0082	0.9638	0.0227	0.8218	0.1420
7	0.0033	0.9671	0.0174	0.8392	0.1278
8	0.0099	0.9770	0.0139	0.8531	0.1239
9	0.0016	0.9786	0.0113	0.8644	0.1142
10	0.0016	0.9802	0.0094	0.8738	0.1064
11	0.0049	0.9852	0.0080	0.8818	0.1034
12	0.0049	0.9901	0.0069	0.8887	0.1014
14	0.0016	0.9918	0.0053	0.8940	0.0978
18	0.0016	0.9934	0.0034	0.8974	0.0960
22	0.0016	0.9951	0.0024	0.8999	0.0952
25	0.0016	0.9967	0.0019	0.9018	0.0949
34	0.0016	0.9984	0.0011	0.9030	0.0954
47	0.0016	1.0000	0.0007	0.9036	0.0964

V. Search engine research literature in the social science disciplines or subject areas of the distribution

This study collected 1340 T 380 kinds of literature found in journals. Following the principle from Bradford's law of the partition, statistics for each journal (the number of articles published in the literature), and the amount of descending order according to the number of articles on the number of journals, literature number, the cumulative number of journals, the cumulative number and cumulative literature journal the number of commonly used pairs of values, the distribution of tables made of ABM journal literature. Furthermore, the first 31 journals published more than half of the literature (785, accounting for 57%) and the remaining 43% of the literature are located in 349 kinds of journals, of which only 243 kinds of journals published in a literature, obviously, ABM journal literature is highly fragmented.

Table 8 shows that the total number of each district, the ratio of 6:45:330 journals, equivalent to about 2:15:110, can be seen as 1∶7∶72, which is broadly in line with Bradford's Law text description.

TABLE VIII.
DISTRIBUTION OF STATISTICS JOURNALS

	(1) No. of journal	(2) No. of articles	(3) Range of No. of articles	(4) Average articles
A	6	441	40~122	74
B	45	453	5-30	10
C	330	466	1~5	1

As can be seen in Table 9, out of all the publications, the first six journals account for all the literature about the three into which, JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY There 122 (9.1%) than the first two of the ONLINE INFORMATION REVIEW (92 articles, 6.86%).

TABLE IX.
CORE ZONE JOURNAL

Journal Title	Count	%	Acc. %
JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	122	9.10%	9.10%
ONLINE INFORMATION REVIEW	92	6.86%	15.96%
INFORMATION PROCESSING & MANAGEMENT	91	6.79%	22.74%
ONLINE	55	4.10%	26.85%
ELECTRONIC LIBRARY	41	3.06%	29.90%
JOURNAL OF INFORMATION SCIENCE	40	2.98%	32.89%

VI. THE MAJOR SEARCH ENGINES ON THE INTERNET TOP LEVEL

The popularity of the major search engines on the Internet is growing and the research literature is growing proportionally. In this study, "Google Insights for Search" for all the world to perform in the Google domain sampling of the Google web search and analysis, calculation in the last period of time, as opposed to the Google search performed on the total amount of major search engines the words the number of times the word is queried. Figure 5 shows a “changes in search volume” chart zoomed in the results of the data points 0-100shows the search terms in the time period (Greenwich Mean Time) in terms of popularity.

Figure 3. Google Insights for Search (Source: Google).

Google and Yahoo search engines were the main representatives of research from 2004 until 2009. Google and Yahoo found that search popularity showed a steady increase and, in 2010, the forecast has also shown up on the forecast trend line. The results of a Google Insights for Search are shown in Fig. 1 for "Over the last fifteen years the research literature search engine growth chart" and, in Fig. 2 for "Over the last fifteen years the research literature search engine reference statistics."

The literature from this search about search engine growth and the major Internet search engines correlate positively correlated and are in line with the discussions in this study.

VII. CONCLUSION

This study aimed to explore the 1995-2009 SSCI for research literature, related to search engine features, and

determine author productivity. SSCI database using literature data collection to analysis, the results are summarized as: 1. The search engine research literature in SSCI continued to grow, with the major research institutions and the largest output of the United States, England, and Canada, outside of national organizations is very focused on the distribution of productive forces. 2. The range of subjects of the research literature mainly focuses on information science, computer science, and information systems and computer science based on cross-disciplinary applications. 3. Author productivity distribution in is line with Loca's Law, the law can be applied to estimates by the loca percentage. 4. The search engine research literature is broadly in line with Bradford's Law of the text description, with the core area, the relevant areas, and the distribution of marginal statistical proportion of the literature, roughly 1:7: 72. There are six kinds of core journals. 5. The major search engines search for popularity with the search engines to grow into the literature a positive correlation.

REFERENCES

[1] Tsai Ming-yue (China 92). Metrology characteristics of information and literature. Taipei: HwaTai.

[2] Tsai Ming-yue (China 94). International Conference on Taiwan's life sciences research productivity literature. Library and Information Journal 53, 1-17

[3] Tsay, M. Y., Jou, S.J., Ma, S.S. (2000). A Bibliometric Study of semiconductor Literature, 1978-1997. Scientometrics. 49(3), 491—509

[4] Aydm Erar (2002). Bibliometrics or Informatics : Displaying Regulating in Scientific Patterns by using statistical Distributions, Hacettepe Journal of Mathematics and Statistics Volume(31), 113-125

[5] Almind, T. C. and Ingwersen, P. Informatic analysis on the World Wide Web, methodological approaches to Webometrics", Journal of Documentation 53 (4), 404-426, 1997.

[6] BAR-ILAN (1999), Search engine results over time: A case study on search engine stability, Cybermetrics, 2/3, paper 1. ISSN: 1137-5019 (http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html; visited 08.11.2000).

[7] Stock,W.G. (2006): On relevance distributions. Journal of the American Society of Information Science and Technology, 57, 1126-1129.

[8] Blachman, N. 2005. Google guide: Making searching even easier. URL: http://www.googleguide.com/ [viewed September 20, 2005].

[9] Jacso, P. 2004. Google Scholar Beta. Peter's Digital Reference Shelf. URL: http://www.galegroup.com/reference/archive/200412/googlescholar.html [viewed September 20, 2005].

[10] Mayr, P. & Tosques, F. (2005). Webometrische Analysen mit Hilfe der Google Web APIs. Information Wissenschaft Praxis, 56, 41-48.

[11] Zuccala, A. (2006) 'Author Cocitation Analysis Is to Intellectual Structure A Web Colink Analysis is to...?', Journal of the American Society for Information Science and Technology, 57(11), pp 1487-1502.

[12] Vaughan, L., Tang, J. and Du, J. (2009) 'Examining the robustness of Web co-link analysis', Online Information Review, 33(5), pp. 956-972.

[13] Vaughan, L., Kipp, M. and Gao, Y. (2007) 'Are co-linked business web sites really related? A link classification study', Online Information Review, 31(4), pp 440–450.

[14] Vaughan, L., and You, J. (2009), 'Keyword enhanced Web structure mining for business intelligence', Lecture Notes in Computer Science, Vol. 4879, pp. 161–168.

[15] S. Robinson et al., The role of networking in research activities (NetReAct D4.1) (Empirica Gesellschaft für Kommunikations- und Technologieforschung mbH, Bonn, Germany, 2006)

[16] Thelwall, M., & Wilkinson, D. (2003). Three target document range metrics for university Web sites. Journal of the American Society for Information Science and Technology, 54(6), 490 – 497.

[17] M. Thelwall, A Web crawler design for data mining, Journal of Information Science, 27 (2001) 319–325.

[18] Lawrence, S., & Giles, C.L. (1999). Accessibility of Information on the Web. Nature, 400, 107–109