

Influences of digital resource acquisition on scientific research behaviors

—The statistical analysis on the full-text downloading quantity and cited times

LIU Xiaomin* & ZHANG Jianyong

National Science Library, Chinese Academy of Sciences, Beijing, 100190, China

Abstract With the emergence and further development of the digital library, the approaches of information acquisition correspondingly change a lot. This paper makes a statistical analysis on the journal downloading and citation behaviors under the digital environment conceived by the National Science Library (NSL), Chinese Academy of Sciences (CAS). It can be seen that the development of digital resources has influenced scientific research behaviors. For example, the high quantity of full-text downloading will maintain; the trend of journal downloading behaviors is basically same as the journal citation behavior; journals with high quantity of full-text downloading also boast the high cited times, and vice versa. Furthermore, authors make a linear regression analysis, with the journal downloading amount as the independent variable and journal cited times as dependent variable. Then they also prove the positive correlation between the journal downloading and citation behaviors by means of Pearson's correlation coefficient formula.

Keywords Full-text downloading, Journal citation, correlation

With the development of the Internet and transformation of information media, users more and more depend on the network to obtain the information resources, with the less and less frequency of visiting real libraries. During the process of collection development, libraries change their strategies gradually from treating the print collections as the main resource to preferring the electronic resources. Then, the new collection development mode is established finally to adopt the e-resources subscription as the main approach. According to the statistical report of the full-text downloading provided by the publishers, it can be known that the amount of full-text downloading is so enormous. And users' research behaviors are impacted more deeply and comprehensively by e-resources, compared with print collections. However, some relevant doubts exist and need to be explored. For instance, it is



CJLIS
Vol. 2 No. 3, October 2009
pp 00-00
National Science Library,
Chinese Academy of
Sciences

* Correspondence should be addressed to Liu Xiaomin (liuxm@mail.las.ac.cn).

Research Papers

difficult to figure out the reason of the so high downloading which is for some specific purposes or only for occupancy with aimlessness; it is difficult to measure or estimate the influence caused by the huge downloading to the scientific researches.

The procedure of scientific research, involving searching, analysis, practice and creation, can be summarized through careful analysis on the generating process of research achievements. In general, during the creating process, the researchers will enumerate information resources which may influence their research results, so as to elucidate the basis or background of the researches. Therefore, a selected reference list is formed accordingly. These reference literatures of the entire information collection searched during the research, usually attached to the research papers or articles, can indeed influence the progress of scientific researches. It is deemed that the impact or effect of electronic resources could be seen through analysis on full-text downloading amount and cited times. This paper just tries to make use of the comparative analysis on the electronic full-text downloading amount and cited times to illustrate the impacts of electronic resources on scientific research activities.

1 Data sources and statistical methods

Considering the availability and comparability of data, authors try to make a comparative analysis on the downloading and cited behaviors from the view of one single discipline.

15 chemical journals have been selected as the statistical objects from the full-text databases subscribed by NSL, CAS. Four selection criteria are considered: 1) The journal should be with large downloading quantity; 2) selected journals should not be from a same publisher; 3) selected journals should be published in different years; and 4) they should represent different branches of chemistry to the most extent. Authors conduct all-round analysis from the aspects of downloading quantity, publishers and core journals, etc. Such 15 journals are listed as follows:

Based on the corresponding publishers' statistical reports of full-text downloading quantities of 15 journals, authors extract the full-text downloading quantities from 2003–2005.

The citation data is derived from Chinese Science Citation Database (CSCD). CSCD has collected more than 1,000 core journals published in Chinese or English in China. 646 journals of which are in the core database of CSCD, involving subjects of natural science, medicine and engineering technology, etc. As CAS research papers are mostly published in the journals of core database, the citation data is extracted based on the 646 journals of CSCD core database. In view of the comparability between the citation data and downloading data, the reference literatures of the research papers with the authors or entities subject to CAS and published from 2002 to 2005 are selected to calculate the cited times of those 15 journals.



Table 1 The list of 15 chemical journals

Journal code	Collected database	Publication year
J1	Elsevier	2000
J2	Elsevier	1980
J3	Elsevier	1993
J4	Elsevier	1951
J5	Elsevier	1967
J6	ACS	1989
J7	RSC	1999
J8	ACS	1962
J9	Elsevier	1967
J10	ACS	1963
J11	ACS	1897
J12	ACS	1968
J13	ACS	1999
J14	ACS	1982
J15	Elsevier	1957

Note: ACS is the abbreviation of American Chemical Society.

2 Data analysis

2.1 Analysis on full-text downloading quantities

Based on the statistics of annual full-text downloading quantities listed in Table 2, the downloading quantity of 2004 is increased by 83.78% than 2003 on the average, the downloading quantity of 2005 is increased by 15.97% than 2004 on the average. The full-text downloading quantities show a trend of year-on-year rise.

In the descending order of the total downloading quantities, 4 journals of the top 5 journals are published by ACS, accounting for 76.29% of total downloading quantities of all 15 journals. The high downloading quantities of journals published by academic associations indicate that such journals comparatively drew more attention from users in related areas of specialization.

By comparing the downloading quantities and start year, it can be found that, normally, the longer publication year, the larger downloading quantities the journal with, whereas, less attention is paid to the comparatively new released journals. J11 was firstly published in 1897, and the downloading quantity is ranked the 1st among all of 15 journals; while, the downloading quantities of J1 and J7 with publication years less than 10 years, are the two lowest. But there still exist some exceptions, for example, J6 and J13 are published for less than 20 and 10 years respectively, but with relatively more downloading quantities. So, it can not be neglected that the hotspots in some subjects on these journals draw great attention from the users. And generally speaking, the degree of popularity of a journal and the information amount offered by it can affect downloading quantities to some extent.



Table 2 The downloading quantity of 15 journals from 2003–2005

Journal code	The downloading quantity of 2003 (times)	The downloading quantity of 2004 (times)	The downloading quantity of 2005 (times)	The total downloading quantity (times)
J1	557	1830	1493	3880
J2	9920	19978	20452	50350
J3	3411	9544	8670	21625
J4	–	13783	13654	27437
J5	14344	25472	25267	65083
J6	57601	74127	88298	220026
J7	–	–	2178	2178
J8	31807	42067	52826	126700
J9	17363	36678	35252	89293
J10	11806	21322	28063	61191
J11	128937	210593	306322	645852
J12	55581	62484	68250	186315
J13	30457	61427	108605	200489
J14	10965	12935	16850	40750
J15	10953	26010	30013	66976

2.2 Analysis on citation data

Based on the citation statistics of 15 journals in CSCD during 2002–2005, we find that the total cited times of 2003 is much more than 2002, with 2.3 times the number of 2002; and, there is an increasing trend in 2004 and 2005 too, however, with relatively low rising speed.

The ascending trend of the yearly total cited times is shown in Fig. 1. Furthermore, it also can be found from Fig. 1 that there is a rising trend of the cited times in each

Table 3 The cited times of 15 journals

Journal code	The cited times of 2002	The cited times of 2003	The cited times of 2004	The cited times of 2005	The total cited times
J1	–	19	40	45	104
J2	35	120	352	265	772
J3	7	79	156	232	474
J4	13	22	44	36	115
J5	165	1043	1486	1880	4574
J6	154	1042	1373	1889	4458
J7	7	75	99	155	336
J8	138	1036	1346	1461	3981
J9	362	437	1012	703	2514
J10	51	216	446	303	1016
J11	704	1223	3184	1958	7069
J12	278	701	1190	883	3052
J13	18	162	322	489	991
J14	67	90	182	189	528
J15	96	750	897	1105	2848
Total	2095	7015	12129	11593	32832



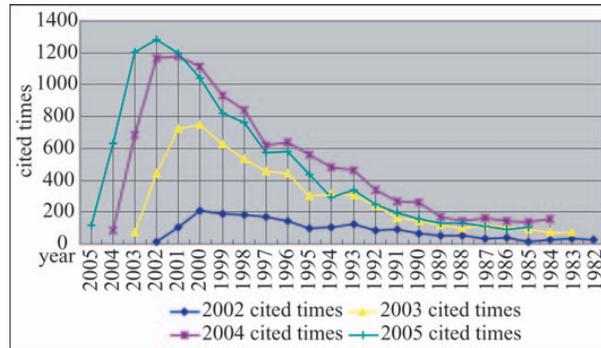


Fig. 1 The 10-publication-year distribution of cited papers in CAS respectively from cited year of 2002 to 2005

publication year of the 10-publication-year distribution respectively from cited year of 2002 to 2005. For example, in the 10-year statistical report of cited times of 2002, the top single annual cited times was only 209 times; in the 10-year statistical report of 2003, the every annual cited times all rose rapidly; in 2004 and 2005, some single annual cited times could reach more than 1,000 times. In view of this phenomenon, the ease access of the reference materials may have to be considered to explain it.

2.3 The comparative analysis on full-text downloading quantity and cited times

By comparing full-text downloading quantities with cited times, it is found that normally, the journals with high downloading quantities are cited more frequently as well.

It is shown in Fig. 2 that the distribution of downloading quantity of the journal is basically consistent with the distribution of cited times. However, the J5 and J13 are the exceptions, both of them displays a contrary trends—the downloading quantity is inversed to the cited times. Referring to the academic features, the two journals present definite disciplinary scopes, not comprehensive journals. As for the publication features, both of them can be classified into the bulletin journal with high publication frequency, which would rapidly present the research trends of relevant subject fields. Through comparative analysis on the distribution of entities downloading the full-texts and that of entities citing articles, the asymmetry entity distribution is revealed, i.e., entities with high downloading quantities are not necessarily consistent with the those citing the journals frequently. In authors' opinion, it's only a superficial phenomenon. In order to figure out the essential cause of this non-positive correlation, the disciplinary features, characteristics of research paper in these journals and users' downloading behaviors should be analyzed in an in-depth manner.



Research Papers

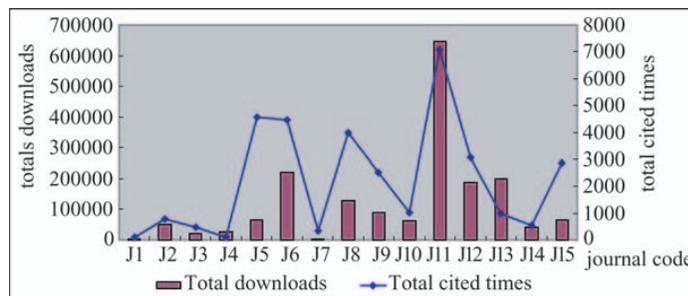


Fig. 2 The comparison between downloading quantities and cited times of 15 Journals

According to Table 3, it can be concluded that the yearly cited times of 15 journals soared rapidly after 2003. The increase of cited times has a certain relationship with users' information-receiving amount. Users wouldn't attach the citation references at the end of their research papers if they did not acquire certain information or knowledge. The production process of a research paper runs a cycle of motion, which includes several stages of conception, information retrieval, writing, modification, publication and subsequent influences spreading, etc. Fig. 3 shows the whole process from literature retrieval, downloading, extending impact, then to the production of research paper. The time difference between current paper downloading and future usage of new scientific literature is displayed as well.

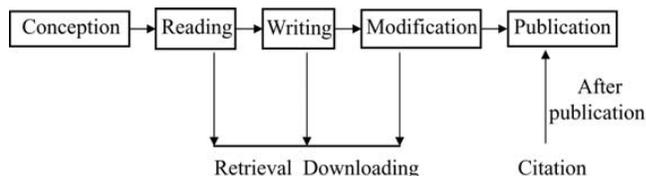


Fig. 3 Scientific research process flow chart

Before 2002, the NSL, CAS treated print collections as the main literature resources, users needed to walk into the library to acquire the information service. After 2002, NSL,CAS adopts the digital resources as the main library collections, and establish the corresponding services including access, downloading, e-file delivery, etc. so that users can acquire the information on-desk. That may explain why the cited times of above 15 journals increased so sharply since 2003. That further reveals that the digital resources impact the scientific researches deeply.

Although more downloading quantities more cited times, how about the degree of influences of the former on the latter? The Fig. 2 displays the 10 publication year distribution of cited papers in different cited years, and clearly shows the peak publication year of cited times. The peak publication years of cited times in cited year of 2002 are 1999 and 2000; the peak publication years of cited times in cited



year of 2003 are 2000 and 2001; the peak publication years of cited times in cited year of 2004 are 2001 and 2002; the peak publication years of cited times in cited year of 2005 are 2002 and 2003. That distribution of peak year is in accordance with typical regularity of citation year distribution. As the peak year reflects the author's cycle of reading, writing and publishing research papers, the inconsistency between peak year and publication year can be thought as the production cycle of scientific research.

According to the production cycle of research papers, it can be estimated that users' retrieval results in 2003 could be reflected in their research paper published in 2005. The correlation between full-text downloading quantity of 2003 and cited times of 2005 was calculated by making use of Pearson's correlation coefficient formula, with the journal downloading amount as the independent variable and journal cited times as dependent variable. The correlation coefficient was 0.66. The correlation coefficient between downloading quantity in 2003 and cited times in 2004 was 0.88. Comparatively speaking, the downloading quantity of 2003 was more relevant to cited times of 2004. Are there some journals which shorten the time length of research paper publication so that the strong correlation between downloading quantities and cited times appears within one year? Further in-depth research and analysis need to be explored. Authors tried calculating the correlation coefficient between the sum of downloading quantities from 2003 to 2004 and the sum of cited times from 2004–2005, which was 0.81. That indicates that the strong correlation between the downloading quantity and cited times did exist, and the multi-year downloading quantities can influence the multi-year cited times. That also proved the positive correlation between downloading quantity and cited times.

3 Conclusion

By comparative analysis on downloading quantities and cited times of above 15 journals, the influences of digital resource development on users' information acquisition behaviors can be clarified. Furthermore, the easier availability of large amount of information supports the development of scientific research and enhances the cited times of journals.

- Users prefer the journals published by the academic associations or organizations closely related to their research fields, compared with the comprehensive journals.
- The yearly ascending trend of full-text downloading quantity reveals the significant role of digital resources in users' information utilization.
- Large downloading quantity remarkably influences the production of users' research paper to some extent, which indirectly results in the sharp increase of cited times since the development of digital resources in CAS.



Research Papers

- There is a positive correlation between the downloading quantity and the cited times of a journal. Higher the downloading quantity, more the cited times.
- There exists relatively strong correlation between the downloading quantities and the cited times.

With the development of the Internet, users' information acquisition approaches have changed significantly. The access to digital resources is only one of the many approaches for users to obtain information. The citation data is unable to indicate the users' acquisition approaches. Therefore, more detailed and deeper researches are required for further analysis on the functions and roles of digital resource for scientific researches. However, the foregoing conclusions still illustrates that the statistical analysis on full-text downloading quantities and cited times of digital resources can contribute to the resource developments, which may aid in exploring the rational distribution of library collections in subjects and media. The relevant data analysis of which may also be used to optimize the library collection development. Finally, better information service can be offered.

References

- 1 Davis, P. M. Where to spend our e-journal money? Defining a university library's core collection through citation analysis. *Portal: Libraries and the Academy*, 2002, 2(1):155-166.
- 2 Duy, J. & Vaughan, L. Are citation data a valid measure of journal use? An empirical examination in an academic context. In Ingwersen, P. & Larsen, B. (Eds.), *Proceedings of ISSI*. Stockholm: Karolinska University Press, 2005:390-397.
- 3 Duy, J. & Vaughan, L. Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination. *Journal of Academic Librarianship*, 2006, 32(5):512-517.
- 4 Chung, H. K. An analysis model of creating a core journal collection for academic libraries. *Library Collections, Acquisitions, and Technical Services*, 2009, 33(1):17-24.
- 5 Schloegl, C. Comparison of citation and usage indicators: The case of oncology journals. Retrieved on 补检索日期 from http://www.kfunigraz.ac.at/iwiwww/publ/schloegl_2008_1.pdf.
- 6 Black, S. Impact of full text on print journal use. *Library Resources and Technical Services*, 2005, 49:19-26.
- 7 Brown, C. M. Information seeking behavior of scientists in the electronic information age: Astronomers, chemists, mathematicians, and physicists. *Journal of the American Society for Information Science*, 1999, 50:929-943.
- 8 Davis, P. M. Information seeking behavior of chemists: A transaction log analysis of referral URLs. *Journal of the American Society for Information Science*, 2004, 55:326-332.



National Science Library,
Chinese Academy of
Sciences