

Citation analysis and collection management

Suzanne Bakker, Amsterdam

Ständig steigende Kosten für Zeitschriftenabonnements zwingen Bibliothekare detailliertere, auf Untersuchungen basierende Verfahren für Abbestellungen und Bestandsmanagement zu finden. Qualitative Messungen, basierend auf Benutzungserhebungen, Benutzerbefragungen und Impactfactor wurden eingehend geprüft. Quantitative Messungen wie Benutzungsstatistiken wurden von mehreren Bibliotheken in Form von kontrollierter Rückstellung und dem Zählen von Fehlkopien im Papierkorb des Kopiergeräts angewendet. In diesem Artikel wird eine kurze Überprüfung und Zitierungsanalyse verwendet, um Zitierung von Zeitschriften, welche im Bestand geführt werden, zu messen. Die Statistik, die Kommandosprache und die Berichtsfunktion von DIMDI (besonders der Extrakt-Befehl) wurden verwendet, um die Fakultätsarbeiten, welche in ISI's Science Citation & Social Science Citation Index ermittelt wurden, zu analysieren. Von den über 200 aktuellen Titeln im Zeitschriftenbestand des Niederländischen Krebsinstitutes wiesen 10% eine extrem niedrige Zahl von Zitierungen in Fakultätsarbeiten auf und wurden deshalb als Abbestellungsvorschläge in Betracht gezogen, welche vom Bibliothekskomitee und/oder den Abteilungsspitzen diskutiert werden sollen.

Ever increasing costs of journal subscriptions force librarians to find more detailed procedures for "evidence-based" cancellations and collection management. Quality measures based on surveys, questionnaires and impact factors have been perused extensively. Quantitative measures, usage statistics, have been used in several libraries by methods of controlled reshelving and counting misprints in the waste basket of the photocopy machine. In this article a quick scan and citation analysis of faculty papers published is used to measure citations to journals held in the collection. The statistics, command language and report function of DIMDI (especially the extract command) was used to analyse faculty papers found in ISI's Science Citation & Social Science Citation Indices. Of the over 200 current titles in the journal collection of the Netherlands Cancer Institute, about 10% showed an extreme low citation score in faculty papers and were therefore considered for cancellation proposals to be discussed with the library committee and/or department heads.

The ever increasing costs of journal subscriptions with over 10% per year and the even more growing additional costs for providing electronic access to full-text have resulted in journal collection management to become equivalent or synonymous to cancellation management.

In the eighties of last century there were multiple subscriptions to one title within an institution; major emphasis in that period was aimed at merging smaller units and resulted into a merged library organisation with only one single subscription to a title/institute. In the same period shared cataloguing and computerisation of national networks for interlibrary loan systems resulted in further cancellations: extensive use was made of photocopies to provide document delivery services, resulting in more cancellations, leaving one or only a few subscriptions per country. Publishers reacted with further rises of the subscription fees. There is still no end to cancellations, as library budgets grow less than the costs of purchasing.

Over all these years library users and faculty received questionnaires about their information need; surveys about collection use resulted in data to be used to make "wise" decisions in cancellations. The trustworthiness of answers from faculty decreased over the years, knowing that their answers would be used for a policy that was opposed to faculty's interests.

The next phase of collection management consisted of using the impact factors of ISI's annual Journal Citation Reports: a high impact factor is used as a substitute for the quality and prestige of a journal. Most libraries have gone through this cycle of cancellations as well.

In research institutes faculty needs to publish papers: not only the number of papers, but also the impact factor of the journal in which the papers are published count in the annual evaluations of research departments. Publishing in and purchasing of journals converged on the most prestigious, which are supposed to be equivalent to the higher impact journals. For the library this meant that information about the titles in which faculty publishes can also be used for collection management decisions. Again a stimulus for publishers of these journals to increase volume (and prices!).

The next phase is the measuring of actual usage of the journal issues: for this purpose it is possible to use control of reshelving, but unfortunately this is often too labour-intensive, on the other hand requests to sign paper slips after consultation of issues is very reliable. Some libraries count the misprints of the photocopy machines, which has a good correlation with actual use. For faculty and staff in a research institute published articles are a measure of scientific output. The reason for having a journal collection in the

research library is to support this research and writing papers. Therefore analysis of the cited journal articles, cited by faculty, can be used for collection management: cited journal articles (journal issues/volumes) should be available in the collection and kept in the archives, according to Eldridge and Guenther (*Bull MLA 2001:89;(1):71-5*). As long as there are no major changes taking place in the focus of the ongoing research in the institute, and taking for granted that the focus of the journals does not change too much, these most-cited journals should be in the current collection. And those titles not cited, should be candidates for cancellation. Elisabeth Wood summarizes the possible use of publication analysis for journal collection management.¹

Although citation data have been available for many years, access and usage fee to the SCI database was a hindrance to extensive usage of these data for library management. However, the availability of the JCR on CD-ROM offers librarians the possibility (without extra costs!) to make use of the actual citations found in faculty's papers.

Lotte² reports about the use of cited reference data by retrieving and downloading from the SCI database on CD-ROM one by one the records of all articles published by faculty, according to the annual scientific report of the institute. Downloaded records include all the cited references. The procedure described by Lotte will result in a 100%

coverage of all data (citing and cited references). These data were uploaded in an Access database and by making use of Perl-scripts further processing for detailed comparisons. This procedure takes a substantive period of time and working hours. Although these results will be very reliable, for collection management more factors have to be taken into account when considering purchase of cancellation. Therefore I propose a more pragmatic procedure, which aims at finding the extremes, the very low cited among the current titles in the collection. For this purpose the high performance computing of DIMDI and the statistical processing and reporting of this system is well suited. The procedure presented will have the advantage that no records need to be shown or downloaded, therefore costs to perform this research are low at the present pricing policy of DIMDI. For a medium sized research institute like the Netherlands Cancer Institute (approx. 300 publications per year) the analysis can be performed within a few hours.

1. Finding the published articles:

- a) Making use of the Science Citation Index and Social Science Citation Index (DIMDI's databases IS74 and IN73)
- b) Finding the articles published by the institute, making use of information in the address field; limiting the search to the publication dates 1995-2001
- c) Finding the articles from different departments by using the author names of senior staff members and combining this set with the previous. The sets thus retrieved, consist of 28 to 272 articles per department, varying between 50 and 90% of the articles actually published by these departments (comparison was made with the database of articles published over these years in the annual scientific reports).

2. Using the "Extract Plus" command for the referenced journal name field; qualifiers for the report were set on minimum frequency 5 and output sequence to be on frequency

- a) Online (screen) output was logged; file could therefore be used for additional sorting of output reports (alphabetic order).
- b) A matrix consisting of the current journal subscriptions of the library against the 14 research departments was used to present the data of frequency of referenced (=cited) journals/department.

3. For those journals with a very low or no citations at all the search was performed the

other way round: retrieving first all articles citing the journal, combining the set with all articles published by the institute in the period 1995-2001.

- a) Using the "Extract Plus" command on the author field to analyse the citing authors with qualifiers set on minimum frequency of 2 and frequency as the output order.
- b) Names of senior staff members will be easily recognized in the (top of the) report to be able to get feed-back on the actual usage and need of the journal.

Of the ca. 200 current subscriptions to international journals in our library, 10% of the titles³ are worth tracking down for actual usage and to discuss possible cancellation and/or replacement, because these were hardly cited in the publications of our institute over the last 6 years (citations found in <10 articles)

The same procedure can be used for finding highly cited journals that are missing in the collection. This exercise was not performed for the Central Cancer Library, because over the years ILL statistics are used to identify titles worth considering for purchase.

In the near future it is expected that services of the library have to be accounted for and broken down towards the different user groups and departments in the institute. The data on citations made by faculty could help to indicate the relative benefit the departments have from the central collection. Dividing costs of purchase over the citing departments may result in usable cost/benefit or cost/citation indicators for the journal collection⁴.

Concluding remarks

Collection management and journal cancellations may provoke emotional arguments and discussions between the library and faculty and are imposed by the shrinking budgets and increasing costs. Factual data about usage, performance indicators or cost/benefit ratios are needed. Citation analysis of faculty publications can be very helpful.

¹ Wood EH. Faculty publication analysis: what journals do they need? Hypothesis: the Newsletter of the Research Section of MLA 2001;15(2):17-18.

² Lotte A. Citaatanalyse in een medische bibliotheek [Citation analysis in a medical library]. Informatie Professional 2000;4(7/8):34-7 [Dutch].

³ Acta Anaesth Scand
Am J Publ Health
Am J Respir Cell Mol Biol
Am J Respir Crit Care Med
Cancer Causes Contr
Clin Chem Lab Med
Clin J Pain
Clin Nutr
Controll Clin Trials

Crit Care Med
Curr Opin Hematol
Nutr Cancer
Radiol Diagn Imag Intervent
Scand J Environm Health
Semin Anesth
Semin Intervent Radiol
Top Magn Reson Imag
Trends Endocrinol Metab
Urol Oncol
Urol Res

⁴ Areas of interest related to research and clinical divisions in the Netherlands Cancer Institute, tentative schedule for journal costs division and cost/citation analysis:

- a) Multidisciplinary journals & institution wide usage (incl. "General oncology")
- b) Research sections (I to VIII)
- c) Radiotherapy (IX)
- d) Internal medicine (non-surgical oncology) (X)
- e) Surgery (XI)
- f) Psychosocial & Epidemiologic Research (XII)
- g) Diagnostics (XIII)
- h) Biometrics
- i) Nursing
- j) Dutch language

Suzanne Bakker M.Sc.,
Central Cancer Library, The Netherlands Cancer Institute /
Antoni van Leeuwenhoek hospital,
Plesmanlaan 121
1066 CX Amsterdam, The Netherlands
<http://www.nki.nl>
tel. +31.20.512 2597
fax: +31.20.512 2599
email: sbakker@nki.nl

Suzanne Bakker, librarian of the Netherlands Cancer Institute in Amsterdam, has been working in academic medical libraries since 1984: first the medical library of the University in Nijmegen and thereafter in the Academic Medical Centre of the University of Amsterdam.

Her professional interests are related to search & retrieval in online databases (starting with the use of DIMDI and still making use of this service, especially for cross file searching and statistical analyses (e.g. citation analysis).

Active in the library associations: biomedical interest group of the Dutch Library Association (NVB/BMI) and EAHIL, the European Association of Health Information and Libraries.

Board member of EAHIL, 1st vice president, and member of the programme committees of several EAHIL conferences (including the 8th European Conference of Medical and Health Libraries in Cologne, September 2002). Within the library associations she is most interested in planning and organizing continuing education courses.

8ECMHL Cologne 2002:

URL: <http://www.zbmed.de/cahil2002>

EAHIL:

URL: <http://www.eahil.org>

NVB/BMI:

URL: <http://www.kb.nl/infolev/bmi>