C. Sean Burns
University of Missouri, Columbia, Missouri, USA

Collecting bibliographic references: A bibliometric analysis of CiteULike's collection as grounds for in-depth interviews

Abstract: CiteULike helps users manage and share bibliographic references on the Web. This study is based on a bibliometric analysis of these references. The findings will lead to email-based interviews with researchers recruited from CiteULike. The goal is to further our knowledge of the needs and practices of researchers and to explore implications for academic libraries.

Résumé : CiteULike aide les usagers à gérer et à partager des références bibliographiques sur le Web. Cette étude se base sur une analyse bibliométrique de ces références. Les résultats mèneront à des entrevues par courrier électronique avec des chercheurs, recrutés au moyen de CiteULike. L'objectif est d'accroître notre connaissance des besoins et des pratiques des chercheurs et d'en explorer l'implication pour les bibliothèques universitaires.

If we were able to examine what researchers are reading and what journal articles they are collecting, we would be able to address long-standing assumptions (Smith, 1981) within bibliometrics as well as potentially explore different dimensions of the citation network. For example, if it is true that many influences are not cited in article publications (MacRoberts & MacRoberts, 2010) and that in general most journal articles are neither cited nor read (Meho, 2007), then one possible way to discover implicit influences is by examining what scholarly communications are collected by researchers. The collections of bibliographic references built by members of CiteULike and similar applications offer one more way to examine "the social worlds of authors" (Nicolaisen, 2003, p. 18).

Inspired by these assumptions, concerns and apparent scholarly practices, this initial project began as an exploratory, bibliometric analysis of a systematic random sample of bibliographic references taken from the CiteULike website. Although exploratory, the findings were revealing and hinted at questions concerning information seeking and implications for academic libraries.

First, the CiteULike data set was downloaded on May 18, 2010 and included identification numbers to 2,419,452 bibliographic references. A total of 999 items were taken as the sample and the identification numbers were used to retrieve the bibliographic references from the CiteULike website. Four could not be located on CiteULike and eighty-nine others were discarded because they were either incomplete, in a foreign language, or they were simple bookmarks to web pages rather than bibliographic references to articles, books, etc. Each reference was then searched on Google Scholar, which was unable to locate thirty-four. For the remaining 872, the citation count for each item was noted as well as whether Google Scholar linked to a full text copy, and if so, from where.

The 872 bibliographic references were to fourteen documents types including, for example, journal articles, proceeding articles, books, book chapters, and dissertations. For all 872 items, Google Scholar located 449 (51.49%) full text copies. Full text journal articles accounted for 345 out of the 449 (76.8%). CiteSeerX (psu.edu) provided 94 full text copies. This was followed by arXiv.org with 37 full text copies and the National Institute of Health (nih.gov) with 35. Open Access publishers and various institutional repositories made up most of the rest. Other details to note: 648 out of the 872 items were references to journal articles. Out of the 648 journal articles, as noted, Google Scholar linked to 345 full text. These had a median citation count of 32 (n = 345, Mdn = 32). For those articles that did not have full copies available via Google Scholar, the median citation count was 12 (n = 303, Mdn = 12).

Due to space, not all of the findings are reported here, but these and the other results do identify or point to researcher practices and behaviors that could be explored in depth through email-based interviews (Meho, 2006). The interview protocol will be based on Meho & Tibbo's (2003) revision of Ellis's (1989; Ellis, Cox and Hall, 1993) model of information seeking behavior of social and physical scientists. There are potential benefits with email-based interviews including automatic transcripts and greater geographic coverage as well as potential costs including the loss of non-verbal communication and in person rapport. The difference between email-based interviewing and surveying should be stressed (Meho, 2006). The researcher seeks general feedback and discussion on information seeking as a topic and on the email-based interview as part of a qualitative research design.

References:

Ellis, D. (1989). A behavioural approach to information retrieval system design. *Journal of Documentation*, 45(3), 171-212.

Ellis, D., Cox, D., & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*, 49(4), 356- 369.

Emamy, K., & Cameron, R. (2007). Citeulike: A researcher's social bookmarking service. *Ariadne, 51*. Retrieved from http://www.ariadne.ac.uk/issue51/emamy-cameron/

Hull, D., Pettifer, S. R., and Kell, D. B. (2008). Defrosting the digital library: Bibliographic tools for the next generation web. *PloS Computational Biology, 4*(10): e1000204. doi: 10.1371/journal.pcbi.1000204

MacRoberts, M.H., & MacRoberts, B.R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology, 61*, 1-13. doi:10.1002/asi.21228

Meho, Lokman I. 2006. E-mail interviewing in qualitative research: A methodological discussion. *Journal of the American Society for Information Science*, 57(10), 1284-1295.

Meho, L.I. (2007). The rise and rise of citation analysis. *Physics World, 20*, 32-36.

Meho, L.I., & Tibbo, H.R. (2003). Modeling the information-seeking behavior of social scientists: Ellis's study revisited. *Journal of the American Society for Information Science and Technology*, 54(6), 570-587.

Nicolaisen, J. (2003). The social act of citing: Towards new horizons in citation theory. *Proceedings of the American Society for Information Science and Technology, 40*(1), 12-20.

Smith, L. C. (1981). Citation analysis. *Library Trends*, *30*, 83-106.