

Mining information-seeking behaviour data to enhance library services

Jenny Walker

Sitting on a Gold Mine

'Google Wants Your Links, Not Your Content' is the title of a recent posting to the Society for Scholarly Publishing (SSP) blog, which draws our attention beyond the content to the use of that content.¹

Users indicate their preferences and selections as they navigate between sets of resources and this valuable data—user 'clickstreams'—can be collected and analysed. User clickstreams, indicating the actual selection of users, are being used by service providers such as Google² and Amazon³ to help users discover items of interest. To date, user information-seeking behaviour data has largely been overlooked for enhancing library services, but steps are being taken in this direction, such as the development of recommender services and new metrics for scholarly evaluation.

If you looked at this item, you might also be interested in ...

Recommender services are now appearing in library applications. LibraryThing⁴, BibTip⁵ and bX⁶ are such examples. These applications record the information resources users access and the order in which they access them, in order to make suggestions for other users. LibraryThing and BibTip record access to the 'catalog' and recommendations are offered at the title level, typically the book. bX can potentially record access to the whole library collection, including remotely-hosted resources; recommendations are typically offered at the article level.

BibTip and bX, deriving from research projects at Karlsruhe University, Germany and the Los Alamos National Laboratory respectively, use a statistical analysis of user information-seeking behaviour to generate recommendations. Recommendations are offered based on co-retrieval of items within a user's session. BibTip uses data from local OPAC usage, while bX aggregates link resolver usage logs from multiple institutions around the world.

BibTip and bX are good examples of harnessing collective intelligence from library users to serve the needs of the library. Recommender services proactively help the user find

information without requiring explicit user queries; interesting items find the user instead of the user explicitly searching for them.

Metrics for scholarly evaluation

The harnessing of collective intelligence from library users is also being explored for the provision of new metrics for scholarly evaluation. Although in the last decade the scope of scholarly communication has broadened well beyond the print environment, the evaluation of research is still largely based on citation and authorship data and has its genesis in the print domain.

User-driven evaluation offers an interesting alternative to citation-based evaluation; shifting the focus from authorship to readership, this alternative offers more immediacy in reflecting the importance of articles for users, and could be especially helpful for journals with high undergraduate or practitioner use. Further, it has the potential to cover new materials and new types of material not currently covered by the Journal Impact Factor⁷. Metrics based on usage are unlikely to replace the well-established impact factor, but could be an important complement.

There are a number of current initiatives towards the determination of usage-based metrics for scholarly evaluation, including the United Kingdom Serials Group (UKSG)¹⁰ Usage Factors project⁸ and project MESUR⁹.

UKSG Usage Factors

In 2006 UKSG commissioned a project to investigate the potential for usage data as a way of generating metrics for scholarly evaluation. The starting point was the vast collection of COUNTER¹¹-compliant usage data.

Positive indications emerged from the results of surveys conducted with librarians and publishers (2006-2007); and from the subsequent testing and modelling with real usage data (2008). The next steps to be undertaken aim at identifying potential candidate usage metrics for longer-term testing on a grand scale. These involve data analysis and modelling using data from a number of content providers.

Project MESUR

Project MESUR, lead by Johan Bollen and Herbert Van de Sompel from the Los Alamos National Laboratory, USA, and supported by the Andrew W Mellon foundation, has earlier this year reported on the outcome of their investigations into usage-based metrics.¹²

The MESUR team collected more than a billion transactions from OpenURL link resolvers and significant scientific publishers and aggregators. These transactions reflect user behaviour across a wide and diverse set of scholarly resources, and represent electronic data searches in which users moved from one journal to another, thus establishing associations between them.

Project MESUR has surveyed a number of different citation- and usage-based metrics (nearly 40) that each represent a unique perspective on scientific impact. Some key dimensions emerge along which scientific impact can vary: most particularly the speed with which a metric indicates changes in scientific interests over time, and also the popularity of a journal versus its prestige or influence. As such each metric expresses a mix of these aspects of scientific impact, and can be selected to favor one or the other.

Map of Science

In addition to proposing usage-based metrics for scholarly evaluation, the MESUR team used their large set of usage data to create a detailed and contemporary view of scientific activity (fig.1). Each dot on the map represents a journal, and the journals are colour-coded for easy subject recognition. The interconnecting lines reflect the probability that a reader will move from one journal to another on the computer screen, each time clicking on articles of interest.

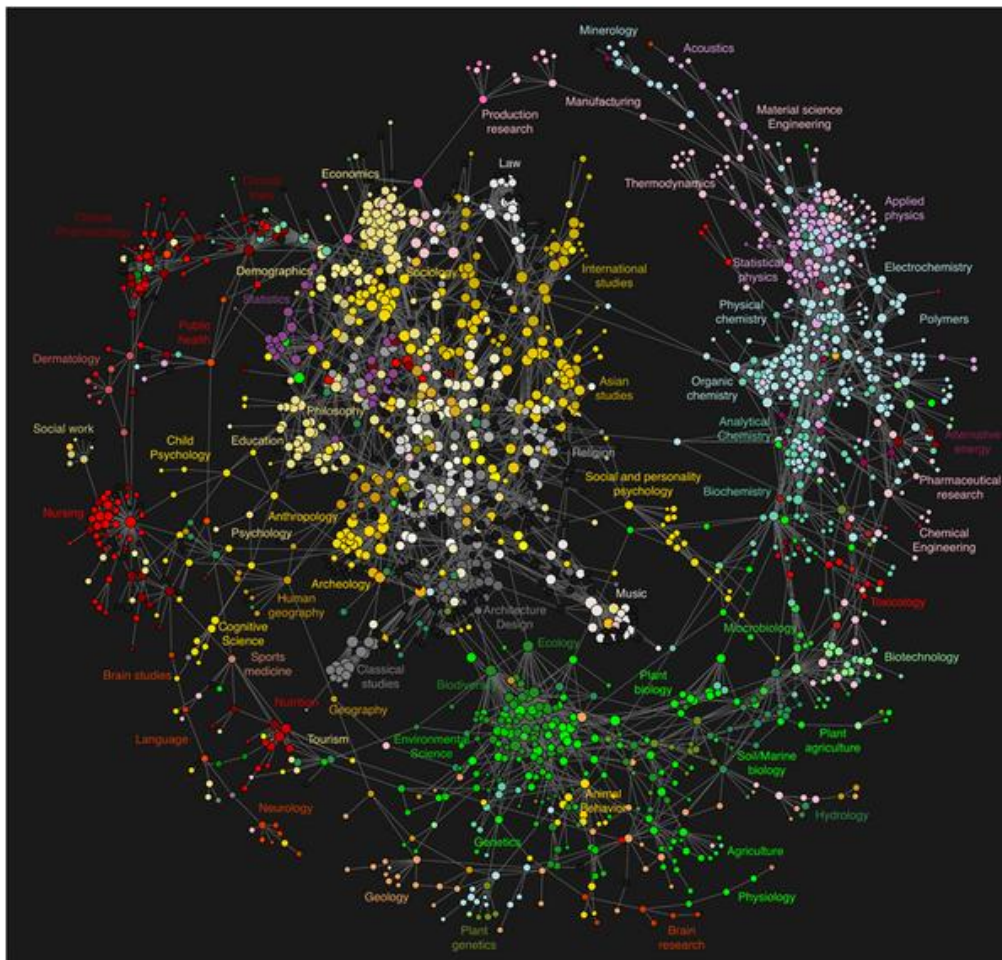


Figure 1: Map of science

This map differs significantly from similar maps constructed on the basis of citations rather than usage, and corrects the under-representation of the social sciences and humanities that is commonly found in citation data. According to Dr Bollen, clickstream maps offer an immediate perspective on what users are doing, and can therefore assist in the detection of emerging trends, inform funding agencies and aid researchers in exploring interdisciplinary relationships.¹³ Further, such maps can help researchers to identify important journals in any particular domain of interest.

Next steps

These first steps in mining user behaviour data to enhance library services are important ones and set libraries on the road to appreciating the value locked up in the data they hold. Information-seeking behavior patterns can serve for a better understanding of the links between items that make up a library, enable better guidance in the use of library resources and can help assess the value of scholarly materials. With our society increasingly focused on measuring research outputs and research quality, serious consideration must surely be given

to new usage-based metrics. In the future, combining user clickstreams with user-profile information has the potential to make this data even more valuable.

References

1. Anderson, K, Google wants your links, not your content. April 14, 2009. Available at: <http://scholarlykitchen.sspnet.org/2009/04/14/links-matter-more-than-content-folks/>
2. Google
<http://www.google.com>
3. Amazon
<http://www.amazon.com>
4. LibraryThing
<http://www.librarything.com/>
5. BibTip
<http://www.bibtip.org/>
6. bX Recommender Service
<http://www.exlibrisgroup.com/category/bXOverview>
7. Thomson Scientific Journal Impact Factor
http://thomsonreuters.com/business_units/scientific/free/essays/impactfactor/
8. UKSG Usage Factors
<http://www.uksg.org/usagefactors>
9. Project MESUR
www.mesur.org
10. UKSG
<http://www.uksg.org/>
11. COUNTER
<http://www.projectcounter.org>
12. Bollen J, Van de Sompel H, Hagberg A, Chute R. 2009. A principal component of 39 scientific impact measures
Available at: <http://arxiv.org/abs/0902.2183>
13. Bollen J, Van de Sompel H, Hagberg A, Bettencourt L, Chute R, et al. 2009 Click-stream Data Yields High-Resolution Maps of Science. *PLoS ONE* 4(3): e4803. doi:10.1371/journal.pone.0004803. Available at <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0004803>

Note: this article first appeared in Serials, the UKSG newsletter, Volume 22, Issue 2, pp182-184.

Jenny Walker is an Information Industry Consultant. Prior to starting her consultancy practice in 2008 Jenny held a number of senior marketing roles with technology and content providers, including executive vice-president marketing at Credo Reference, vice-president marketing at Ex Libris, and director of technology product management at SilverPlatter. Jenny has a keen interest in the development and deployment of interoperability standards and currently serves on the NISO architecture committee.