UKSG

Augmenting discovery data and analytics to enhance library services

Discovery services have become a norm in academic libraries. Usage data from these services has historically emphasized quantitative data such as full-text downloads, numbers of searches performed and title usage. Okanagan College Library incorporated Google Analytics into their discovery service in order to go beyond vendor-supplied quantitative data and obtain valuable qualitative information about users. This information can now be used to shape library services such as collection development, public services and interface design. The user-behaviour data findings are presented, along with suggestions for using this information to enhance library services.

Introduction

In recent years, unified article-level discovery services have become a mainstay in academic libraries. Federated search was plagued by slow response times and only a satisfactory user experience[1]. Discovery products have surpassed the sophistication of the previous generation of federated search with the development of unified indexes, better relevancy ranking algorithms and features that resemble mainstream web search engines[2]. Usage data taken from discovery services frequently demonstrates an increase in number of searches and full-text downloads[3]. However, in isolation of richer analytics, this data can tell us very little about the actual search behaviour and information literacy of users. While significant attention has been devoted to cost-usage metrics, often due to library budget pressures, less attention has been paid to data that provides insight regarding user experience, despite the fact that this is what ultimately determines the success of discovery services. Libraries that can gather and analyze user behaviour data from discovery services have the ability to create both a quantitative and qualitative evidence base to inform and shape library services[4].

ROËN JANYK
Web Services Librarian
Okanagan College, British Columbia

Developing discovery services at Okanagan College

Okanagan College is a community college resting in the heart of British Columbia's Okanagan Valley. The research needs of users varies widely, as students come to the Library from a variety of programs, ranging from Bachelor of Science in Nursing to Associate of Arts Diploma. Following a 2009 usability study conducted by Okanagan College Library, it became apparent library users needed a single search box to find resources successfully. At the time, users typed all library searches into the one search box on the page, the catalogue search box. Users had the expectation they should be able to fulfill their information needs using a single search.

In 2010, the Library further examined the needs of both library staff and users, and found a web-scale discovery product was necessary in order to remain relevant. Okanagan College provides a unique challenge for vendors, as the Library does not use a proprietary link resolver or e-journal knowledge base. As a supporter of open source software and locally grown systems, the Library is an active user of the reSearcher suite of software, including the CUFTS journal knowledge base, electronic resource management system (ERM) and GODOT link resolver. Vendors were challenged with the delivery of a

> "… a quantitative and qualitative evidence base to inform and shape library services"

discovery system that could easily be integrated with non-proprietary systems; yet still provide users with a superior search experience. The product needed to be flexible, able to go beyond default interface design, and also provide options for deep modifications to system functionality and settings. EBSCO Discovery Service (EDS) was chosen to provide that flexibility using a variety of customization options and back-end system settings. In September 2011, Okanagan College went live with EDS.

Choosing a discovery system that supported deep integration meant external tools could be added to existing vendor-provided features. As an existing Google Analytics user, the Library discovered the flexibility of EDS when it was able to insert tracking code to record data on the search behaviour of users. What and how individuals search, the amount of time a user spends after a search, the average number of result pages users will view, the demographic make-up of users, and countless other reports and fine tuning can be accomplished. The data from Google Analytics was able to provide far more qualitative information about user's search behaviour than was previously available to Okanagan College.

Devising a user-focused methodology to gather discovery analytics

In 2013, the Library began the process of creating a strategic plan and discussed how students, faculty and community members use the Library and its services. Physical measures such as library gate counts or circulation of print items tell a smaller part of the story than before, as the usage of libraries has drastically shifted to an online environment. Libraries are being consistently asked for usage data, such as the usage of certain e-journals, but it is also important to ask how and why they found those e-journals in the first place.

The Library integrated Google Analytics into EDS as a means of collecting both quantitative and qualitative information about searchers. The HTML tracking code was inserted into the footer of EDS to ensure each visited page was tracked, including both search and result pages. Data was obtained from Google Analytics over an eight-month period and included information from the busiest months of each academic semester. Reports were generated in Google Analytics and exported as Microsoft Excel spreadsheets. The Excel spreadsheets provided a variety of assessment options, and the ability to sort searches and transactions a number of ways allowed for detailed data analysis.

To navigate the excess amounts of data generated by Google Analytics, some key points of interest were identified. These included the specific searches made, data on search terms, search strings and search revisions, and whether a user included search parameters such as quotation marks for phrase searching, Boolean search terms and truncation symbols.

> "The search system data revealed valuable information about users"

Analysis and findings

The search system data revealed valuable information about users, specifically about their search strategies, behaviours and competencies, as well as the impact of library instruction. In terms of instruction, the number of page views and associated date and time stamps informed librarians and faculty about the most effective times to schedule research instruction sessions. Examining time frames in relation to numbers of page views demonstrates times of heavy usage and gives insight into key points in the semester for research.

The demographic and location information retrieved from Google Analytics indicated the majority of users were using the library search tool from off-campus (63%). Knowing where users are accessing information informs whether to buy in electronic or print formats, and contributes to licensing negotiations with database vendors. In terms of types of devices used to conduct research, figures showed very few searches on mobile devices, 1,000

searches from tablets, and 46,000 on computers. This information gives good insight into the research practices of users, particularly on mobile devices.

The EDS basic search box is embedded in the Library's home page as well as on the native platform. Data showed few searchers opting to use the advanced search feature even though librarians actively encourage use of the feature. Interestingly enough, data showed advanced search users left the system less and were more successful at retrieving the final information source. It is unknown whether the advanced search feature brought success, or if those who easily found their information were more advanced searchers already familiar with retrieving sources. The basic and advanced search page designs could also be influencing their tendencies to choose one search method over another. All library links point to the basic search page, and the advanced search link is not prominent. Unless a user is actively looking for the advanced search option, there is a high likelihood it goes unnoticed[5].

> "… individuals treated the discovery service as they would … Google"

Following the examination of search strings and user behaviours, the theme emerged that more time needs to be devoted to the preparation of research strategies. Understandably, individuals treated the discovery service as they would a large search engine, such as Google[6]. However, search term and strategy data revealed a high number of unproductive searches. Natural language searches were frequent, phrase searching was not well modelled, and entire sentences or research questions were often typed into the search box. This led to the assumption that little time and effort goes into the development of keywords or search strategies. The recurrent use of stop words and long search strings far exceeding search box character limits indicated that users lack an understanding of underlying discovery service functionality. Researchers need to become more efficient and effective by putting more time into learning proper search strategies and developing quality keywords.

While analyzing search strings in detail, some tendencies became evident. The search box was regularly used to search for databases, denoting that users do not have a good understanding of how the search system functions, the content that is included, or even the concept of a research database. Some of the most common database names searched were 'Academic Search Premier', 'Hoovers', and 'CPI.Q'. Providing linked database records within discovery service result lists would support users regardless of their knowledge of the system's functionality, as well as using information literacy and research instruction sessions to explain how the research process works, and the hierarchy of information resources in the Library. Presenting a search interface that clearly identifies what users are searching may also help with their understanding.

The high number of tracked searches containing acronyms, particularly those related to citation styles, meant the Library needed to better assist users looking for citation information. Modifications were made to EDS and depending on a user's search, a dynamic widget now prominently appears at the top of result lists when a user searches APA, MLA or Chicago, and provides links to access citation manuals and help pages.

> "… the Library needed to better assist users looking for citation information."

According to the collected data, searchers rarely used Boolean terms to expand or narrow their search. Individuals were more inclined to begin a new search with different search terms rather than rerunning the same search and incorporating Boolean terms. This behaviour is consistent with other findings that have shown when students are unable to find a relevant item on the first page of search results, they reword the search terms and rerun the search[7,8]. If librarians are still focusing their instruction on Boolean search practices, this data may provide evidence that it should no longer be the main focus of library instruction[9]. Instead, librarians may want to focus on keyword formation, search strategies, and navigating search interfaces and functionalities.

Search string analysis provided evidence that assignments greatly influenced usage of the discovery tool, as well as search terms. This illustrates the importance of communication

between faculty and librarians to ensure the anticipated resource needs of students are addressed prior to the start of terms[10]. By faculty choosing to share their assignments, librarians can be better equipped to meet the needs of students.

The average number of result pages viewed, as well as the average number of records accessed, is also important information to monitor. Google Analytics data revealed the number of result pages viewed during a discovery search session was 1.83, implying that after a search is conducted, many users do not move to the second page of search results before exiting the search system or modifying their search[11]. Other studies have confirmed that search revisions are normally made by rewording a search, and students are not inclined to use or even notice many search limiters and facets until prompted[12,13]. This could be partly related to interface design and how intuitive it is for users to find and use these functions. That being said, individuals still seemed disinclined to move to additional result pages, holding the expectation that required information should be on the first page of search results[14].

> "… students are not spending a lot of time in the information retrieval stage of their assignments."

In conjunction with individuals avoiding additional result pages, they also do not access many detailed records. The average number of accessed pages per visit was just over six, indicating that on average a user will click on six different detailed records in a result list before leaving the search service or modifying the search. As the number of results that appear on a page is 50, relevant resources are likely left out of consideration. With approximately 37% of searches being refined, it brings into question what happens to the other 63% of searches. Are individuals able to meet their information need after the first search, or do they simply exit the discovery service altogether?

The average time after a search was approximately three minutes and 46 seconds: evidently not a long time to spend evaluating results. With the tendency to avoid the second page of search results, and the small number of detailed records accessed, it seems students are not spending a lot of time in the information retrieval stage of their assignments. The data shows researchers do not tend to search efficiently or effectively, bringing into question how users are finding the reliable information they need in such a short period of time[15]. The explanation could be tied to two theories, Rational Choice Theory (in this case 'Satisficing') and User Gratification Theory.

Rational Choice Theory

Rational Choice Theory considers people as distinctive in the world because of their ability to set preferences and have intentions[16]. Individuals do not rely solely on instincts and instead use reason to also provide guidance[17]. In this day and age, however, students are often short on time, working with tight assignment deadlines and other pressures, and therefore do not always make the most rational decision. Rather than facing a challenge, individuals will often choose the option that brought acceptable outcomes in the past[18]. Students may understand it is better research practice to review more sources and find higher quality articles, but they may choose instead to use the articles that appeared on the first page of search results and were convenient to find, due to other influences such as time constraints or other commitments[19]. If the student receives a respectable mark on an assignment after using the easy-to-find articles that appeared on the first page of search results, they may be more inclined to use the same research strategy for future assignments.

Uses and Gratification Theory

Uses and Gratification Theory has its basis in the field of communications, and is built on the premise that an individual's motivations for selecting or acquiring media are based on their ability to fulfill their need and further lead to gratification[20]. Gratification Theory identifies motivations for the acquisition of information through certain media[21]. In a discovery service context, if a user accesses only a small number of detailed records or a single page of results, yet is quickly able to find a relevant information source that meets all the necessary criteria, they will continue to use that tool because it gave them success in fulfilling their

information need[22]. Gratification is a good predictor of future media use[23]. Therefore, a student's success or failure at finding relevant information can have repercussions on their future use of a search tool and behaviour within the tool.

Shaping library services

Libraries are now able to seek a better understanding of users and their tendencies, information that can be used to enhance and inform library services. As Rational Choice and the Uses and Gratification Theories have demonstrated, students naturally seek immediate gratification and therefore rarely spend a lot of time examining search results or choosing information sources based solely on rationality[24,25]. To minimize convenience as a deciding factor in the research process, information literacy plays an important role in educating researchers and students who use discovery search services[26]. Qualitative user behaviour data creates opportunities for library instruction programs to focus on highlighted areas of need and address ineffective search habits. Exposing students to information literacy essentials early creates an opportunity to build upon previous knowledge and move beyond information literacy basics in a progressive manner. User behaviour data will change over time, and information literacy programs should also change to maintain relevancy.

Analyzing search terms over each academic term can ensure physical and online library collections continue to meet the needs of users. Data that demonstrates popular or common search strings may be attributed to assignments from multiple different departments. Identifying search topics from interrelated disciplines is useful for making collections decisions such as journal renewals, e-book package purchases and understanding the need for print copies.

Gathering qualitative information also provides insight into the success of a discovery service's interface design and usability. The continuity of use depends on whether a system is intuitive to navigate, whether modification of searches can be made easily, and whether users can view and evaluate information sources quickly and accurately[27]. With the skill level of students presented here, the search interface must be intuitive, and effective labelling must be used for limiters and filters in order to provide guidance during the research process[28]. From a back-end design standpoint, discovery services must allow for collaboration with external tools and systems, to ensure long-term flexibility for libraries.

Conclusion

The library users we see today cannot be compared to those from even ten years ago. Individuals now have a wealth of knowledge at their fingertips. The data collected using Google Analytics at Okanagan College demonstrates that effective and efficient discovery services are required to sustain the demands of current researchers. The data revealing individuals' search practices, such as few detailed record views and search service exits after very short periods of time, leads to the assumption that students do not take time to carefully plan research strategies prior to beginning research. Discovery services must be designed to accommodate for the imperfections of novice searchers and use their behaviours to enhance the search experience[29].

The onus cannot only be placed on discovery services alone; the Okanagan College user data indicates individuals need to gain an understanding of how the underlying search systems function to avoid inefficiencies such as searching long sentences that exceed character limits of search boxes, using improper keywords, searching acronyms, and searching database names. Information literacy programs at academic libraries are more important than ever to give students the skills they need to be successful researchers and consumers of information[30].

Libraries are capable of using newly available qualitative user behaviour data to inform services. Data retrieved from Google Analytics could provide other libraries with the new understanding Okanagan College now has about information users are looking for, the

knowledge gaps that exist, and additional support the Library could provide to assist users. With the flexible discovery services on the market, and streamlined library search tools, it is finally possible for libraries to augment discovery analytics to better shape and customize library services for the current generation.

References

1. Korah, A, and Cassidy, E, Students and Federated Searching: A Survey of Use and Satisfaction, *Reference & User Services Quarterly*, 2010, 49(4), 325 – 332.

2. Wisniewski, J, Web Scale Discovery: The Future's So Bright, I Gotta Wear Shades, *Online*, 2010, 34(4), 55–57.

3. Vaughan, J, Investigations into Library Web-Scale Discovery Services, *Information Technology & Libraries*, 2012, 31(1), 32–82.

4. Bates, M, Current State of Advanced Search: Progressing or Receding? *Online Searcher*, 2013, 37(6), 80.

5. Bates, M, ref 4.

6. Vaughan, J, ref. 3.

7. Cassidy, E D, Jones, G, McMain, L, Shen, L and Vieira, S, Student Searching with EBSCO Discovery: A Usability Study, *Journal of Electronic Resources Librarianship*, 2014, 26(1), 17-35.

8. Holman, L, Millennial Students' Mental Models of Search: Implications for Academic Librarians and Database Developers, *Journal of Academic Librarianship*, 2011, *37*(1), 19–27.

9. Holman, L, ref. 8.

10. Fry, A and Rich, L, Usability Testing for e-Resource Discovery: How Students Find and Choose e-Resources Using Library Web Sites, *The Journal of Academic Librarianship*, 2011, 37(5), 386–401.

11. Holman, L, ref. 8.

12. Cassidy, E D, Jones, G, McMain, L, Shen, L and Vieira, S, ref. 4.

13. Holman, L, ref. 8.

14. Williams, S C and Foster, A K, Promise Fulfilled? An EBSCO Discovery Service Usability Study, *Journal of Web Librarianship*, 2011, 5(3), 179–198.

15. Research Information Network, *E-Journals: Their Use, Value and Impact*, 2009, London, Research Information Network.

16. Hardin, R, Rational Choice Theory. In: *Encyclopedia of Sociology* (2nd ed.), 2001, New York: Macmillan Reference USA.

17. Simon, H A, A Behavioral Model of Rational Choice, *The Quarterly Journal of Economics*, 1955, 69(1), 99–118.

18. Hardin, R, ref. 16.

19. Connaway, L S, Dickey, T J and Radford, M L, "If it is Too Inconvenient I'm Not Going After it:" Convenience as a Critical Factor in Information-Seeking Behaviors, *Library & Information Science Research*, 2011, 33(3), 179–190.

20. Whiting, A and Williams, D, Why people use social media: a uses and gratifications approach, *Qualitative Market Research: An International Journal,* 2013, 16(4), 362–369.

21. Lariscy, R W, Tinkham, S F and Sweetser, K D, Kids These Days: Examining Differences in Political Uses and Gratifications, Internet Political Participation, Political Information Efficacy and Cynicism on the Basis of Age, *American Behavioral Scientist*, 2011, 55, 749.

22. Fry, A and Rich, L, ref. 10.

23. Whiting, A and Williams, D, ref. 20.

24. Connaway, L S, Dickey,T J and Radford, M L, ref. 19.

25. Research Information Network and Consortium of University Research Libraries, *Researchers' Use of Academic Libraries and their Services*, A Report Commissioned by the Research Information Network and the Consortium of University Research Libraries, 2007, London, Research Information Network and the Consortium of University Research Libraries: **http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/researchers-use-academic-libraries-and-their-serv** (accessed 24 October 2014).

26. Richardson, H, Revelations from the Literature: How Web-Scale Discovery Has Already Changed Us, *Computers in Libraries*, 2013, 33(4), 12–17.

27. Connaway, L S, Dickey,T J and Radford, M L, ref. 19.

28. Gallaway, T and Hines, M, Competitive Usability and the Catalogue: A Process for Justification and Selection of a Next-Generation Catalogue or Web-Scale Discovery System*, Library Trends*, 2012, 61(1), 173–185.

29. Breeding, M, The State of the Art in Library Discovery, *Computers in Libraries*, 2010, 30(1), 31–34.

30. Richardson, H, ref. 26.

Roën Janyk
Web Services Librarian
Okanagan College, British Columbia
E-mail: rjanyk@okanagan.bc.ca

ORCID iD: http://orcid.org/0000-0001-9665-986X

To cite this article: