

## **The Use of Transaction Logs to Model User Searching Behaviours**

**Mary C. Schlembach<sup>1</sup>, William H. Mischo<sup>1</sup>, and Joshua Bishoff<sup>1</sup>**

<sup>1</sup>Grainger Engineering Library, University of Illinois at Urbana-Champaign, USA

**Abstract.** To optimize search and discovery services, it is important to develop evidence-based models of user information seeking behaviours within distributed retrieval environments. The University of Illinois at Urbana-Champaign Library has been collecting custom transaction log data from their main gateway interface and its underlying Easy Search (ES) federated search system since 2007. The Illinois team performed a detailed analysis over 1.4 million user searches and 1.5 million user target clickthroughs performed in 2010 – 2011. This analysis has revealed rich information on user search characteristics, search assistance usage, and clickthrough actions.

**Keywords.** Transaction log analysis, user searching behaviour, University of Illinois

### **1. Introduction**

In order to optimize search and discovery services, it is important to develop evidence-based models of user information seeking behaviors within distributed retrieval environments. While a large number of user information seeking studies have been performed, our knowledge of user searching patterns, particularly in online catalogs (OPACs), is incomplete and often contradictory Markey (2007a, 2007b). The University of Illinois at Urbana-Champaign Library has been collecting custom transaction log data from their main gateway interface and its underlying Easy Search (ES) federated search system since 2007. ES provides contextual and adaptive search assistance mechanisms that present the user with search modification and reformulation suggestions and perform additional target searches in the background. For example, for multiword user search arguments, the system performs additional phrase and title searches to assist the user in selecting the most appropriate results. The Illinois team performed a detailed analysis of the project's custom transaction logs collected over the Fall 2010 and Spring 2011 semesters. This analysis looked at approximately 1.4 million user searches and over 1.5 million user

target clickthroughs. This analysis has revealed rich information on user search characteristics, search assistance usage, and clickthrough actions.

## **2. Methodology**

Transaction log analysis is a commonly-used methodology for examining the characteristics of user-system interactions and patterns. As a tool, it can inform system design by providing details regarding user action (Jansen 2006a; Peters, 1993). It is, however, not as useful for understanding user motives and determining user satisfaction. In addition, as Peters (1993) has noted, to glean some types of detailed information it is often necessary to go beyond algorithmic software analysis techniques and “hand-inspect” and re-do searches from samples of the log entries.

Rather than relying on an analysis of standard system web logs, the Easy Search team designed and constructed custom transaction logs with all search and clickthrough actions written into a dynamic relational database. This database contains two tables, one containing search information and the other the target clickthrough information. The search table contains separate entries for: the web site referring page; the user’s IP address; the system generated session ID; the user’s previous search; the user search argument or the search generated by the user execution of a specific search suggestion; the type of search being performed; the system generated search suggestions; the user agent which records web browser and operating system used by the client; a link key to the clickthrough table; and the date and time. The clickthrough table contains the date and time; a 3-letter code for the target that was clicked; and the link key with the search table.

The log analysis is conducted by executing various SQL (Structured Query Language) commands against the transaction log database and collecting these search results.

This log analysis will focus on gateway transaction logs gathered during the Fall 2010 and Spring 2011 semesters of the academic year at Illinois. The analysis was conducted over a set of 1,394,838 gateway searches, eliminating all searches performed by developers and testers. During this time period users also selected and clicked on specific search targets 1,531,344 times. Targets that the user then selected and went to (clickthroughs) are recorded in the transaction logs and the user is then subsequently redirected to the specific content target.

## **3. Findings**

Among the findings: users of the Illinois gateway enter an average of 4.33 terms per search query – much higher than previous studies;

A little more than 48% of the search sessions contain more than one search term or a combination of search terms and search assistance actions – also higher than other studies; and while 66% of all searches originate as default keyword searches, the percentage of known-item or specific title/author searches exceeds 51% of the search queries. Known-item searches are performed in almost 55% of the search sessions.

Number of Words in User Search Query	Frequency of Occurrence
1	102,028
2	212,072
3	219,466
4	127,875
5	77,126
6	47,315
7	32,038
8	24,464
>8 words	99,643
Average	4.33 words per query

**Table 1.** Terms per Query. There was an average 4.33 words per query.

	2007 Study	2011 Study
Total Number of Searches Examined	3,100	8,474
Overall % of Known-Item Searches	49.4%	51.2% of Searches 54.8% of Sessions
Author/ Title Searches	7.4%	6.04%
Author Searches	28.9%	16.9%
Book/ Monographic Title	40.5%	33.6%
Index/ Abstract Title	6.8%	4.2%
Specific Journal Article Search	5.7%	26.9%
Specific Journal Title Search	11.8%	12.4%

**Table 2.** Known Item Searching

In addition, the ES search assistance suggestions and custom links are well-accepted by users; in 32.45% of all search sessions and 58% of the sessions with more than a single search query, users employed one or more search assistance operations.

Search Assistance Suggestions made by System	Percentage of all sessions
Reduce results to exact phrase/title words	36.43%
Matching Journal Title found	17.79%
Spelling change suggested	12.61%
Redo as Author search suggested	7.69%
Direct Link/frequent search result found	6.09%
Citation entered – go to linker module	4.11%

**Table 3.** Search Assistance suggestions made by Easy Search within search sessions.

Search Suggestions Employed by Searchers	Assistance Employed by Searchers	Percentage of times each SA type was used when suggested	Percentage of Multi-Query sessions where SA was used
Spell change suggestions clicked		29.04%	7.86%
Direct link clicked		57.56%	6.58%
Journal Title match click		21.41%	6.95%
Limit to Phrase/Title or clicked on title results		28.38%	15.76%
Author Redo Offer clicked		7.95%	1.36%
Ask-a-Librarian clicked		0.14%	0.29%
Complete citation passed to linker module		2.4%	0.25%

**Table 4..** Utilization of search assistance suggestions by users.

The logs also revealed that users are entering complete or partial journal titles and then clicking through at a high frequency into an A-to-Z e-journal list link and that the exact phrase/title words added links shown in selected results displays are heavily used. Users click on the presented journal title link 21.41% of the time that they are suggested and in over 6.86% of all search sessions. In addition, the journal title search option tab constitutes over 12% of the searches within the gateway. The use of publisher e-book matches is also high -- with clickthroughs into all the e-book content targets totaling 9.31% of all result target clicks and taking place in 11.36% of all search sessions.

In addition, In terms of clickthroughs to results targets, journal article clickthroughs (via A&I services) account for over 47% of the user target selections with OPAC (26%) and all monographic access clicks (33%) also being heavily utilized as targets by users.

User Category	Clickthroughs by Percentages
Article databases	47.25%
All books	33.9% (OPACs 26%, E-Books 7.9%)
Journal/Database titles	17.41%
Web search engines	0.71%
Newspapers and news sites	0.83%
Reference titles	0.35%

**Table 5.** User Clickthroughs into Resource Types.

In our study, the journal article literature and monographic target clickthroughs together account for 80.66% of the user clickthroughs. However, 19.35% of all user clickthroughs are directed to resources outside of the OPAC and journal article targets, including journal/database title or other resources.

#### **4. Conclusion**

Our study shows that academic library search and discovery systems must accommodate known-item searching, search assistance mechanisms, and limiting operations. The results reported in this paper can inform development efforts in the nascent web-scale discovery systems.

#### **References**

- Jansen, B. J. (2006a). Search log analysis: What it is, what's been done, how to do it. *Library and Information Science Research*, 28, 407-432.
- Markey, K. (2007a). Twenty-five years of end-user searching, part 1: Research findings. *Journal of the American Society for Information Science and Technology*, 58(8), 1071-1081.
- Markey, K. (2007b). Twenty-five years of end-user searching, part 2: Future research directions. *Journal of the American Society for Information Science and Technology*, 58(8), 1123-1130.
- Peters, T. A. (1993). The history and development of transaction log analysis. *Library Hi Tech*, 11(2), 41-66.