PROJECT MUSE®

Transaction Logs and Search Patterns on a Children's Portal
/ Journaux de transaction et modes de recherche sur un
portail web destiné aux enfants

Jamshid Beheshti, Andrew Large, Marni Tam

Canadian Journal of Information and Library Science, Volume 34,
Number 4, December / décembre 2010, pp. 391-402 (Article)

Published by University of Toronto Press

DOI: <https://doi.org/10.1353/ils.2010.0011>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/404854>

Transaction Logs and Search Patterns on a Children's Portal

Journaux de transaction et modes de recherche sur un portail web destiné aux enfants

Jamshid Beheshti, Andrew Large, and Marni Tam
School of Information Studies
McGill University
3661 Peel, Montreal, QC, H3A 1X1
jamshid.beheshti@mcgill.ca

Résumé : Les journaux de transactions d'un portail web destiné aux enfants ont été analysés dans le but de dégager les modes de recherche lorsque les utilisateurs du système disposent de quatre choix de recherche. Les résultats montrent que la taxinomie des sujets et les options de recherche alphabétique comptent pour 83 pour cent de toutes les recherches, ce qui indique une préférence des utilisateurs pour la navigation plutôt que la recherche par mots-clés. Les implications pour la conception de portails pour enfants sont discutées.

Mots-clés : journaux de transaction, enfants, voyage d'aventure dans l'histoire, navigation web, recherche par mots-clés

Abstract: Transaction logs for a children's portal are analyzed to investigate search patterns when the system users are presented with four search options. The results show that the subject taxonomy and alphabetic search options account for 83 per cent of all searches, indicating users' preference for browsing rather than keyword searching. The implications for designing children's portals are discussed.

Keywords: transaction logs; children; history trek; browsing; keyword searching

Introduction

Children, like adults, frequently use the Web to find information for a variety of purposes including leisure, as well as assigned tasks such as class projects. Like adults, they use Google, MSN, and Yahoo! rather than search engines and portals specifically designed for young users (Large et al. 2004). Children's searching behaviour, however, may differ from that of adults as they can encounter a number of problems in using

keyword searching and applying Boolean operators. Children may be competent computer users, but they are not always able to retrieve information efficiently and effectively. Many of the problems that children typically encounter when searching include determining the appropriate concepts, translating these concepts to keywords that the information retrieval system can understand, finding appropriate synonyms, and distinguishing homonyms (Large, Nessel, and Beheshti 2008). Perhaps the most formidable obstacle that children can face is the ability to recover and modify failed strategies (Bilal and Kirby 2002).

The difficulties in keyword searching that young users encounter prompted us to probe and discuss the inclusion of several other search options in a portal, History Trek (<http://www.historytrek.ca>), designed by a team of adult researchers and elementary school students and intended specifically for use by young people. In this paper, we report on the information-searching patterns on this portal as recorded in its transaction logs. The objective of the research is to investigate the search patterns on a portal when young users are presented with four search options: keyword, topic taxonomy, alphabetical, and advanced.

Background

Since the 1990s, researchers have investigated children's information-seeking behaviour in digital environments (Beheshti et al. 2006; Bilal 2000; Bilal, Sarangthem, and Bachir 2008; Borgman et al. 1995; Chelton and Cool 2007; Dresang 2005; Druin 2005; Hirsh 1999; Kuhlthau 1991; Large and Beheshti 2000; Large et al. 2006; Large, Nessel, and Beheshti 2008; Shenton and Dixon 2003). While many of these studies utilized both quantitative and qualitative methodologies, they generally relied on relatively small numbers of users, and most were conducted under experimental conditions.

Transaction logs have been used extensively in the past decade to investigate search patterns and behaviours of users of Web-based search engines (see, for example, Asunka et al. 2009; Bleicic et al. 1998; Jansen and Spink 2005). The majority of these studies utilized transaction logs because large quantities of data can be collected unobtrusively in a short time, and data collection is inexpensive (Jansen 2006). While History Trek has been positively evaluated by other elementary school students

under both experimental and operational conditions (Large et al. 2006), the scope of testing in terms of sample sizes has been relatively modest. Transaction log analysis (TLA) allows us to investigate search patterns at a macro level, using a large data set representing thousands of users. Moreover, TLA provides the opportunity to study real-life search patterns, both imposed and self-directed, rather than testing in an artificial experimental environment.

Transaction logs analysis, however, has its own limitations. The IP address may or may not represent a single user, and the identity of users, and therefore their age, socio-economic background, demographics, personality traits, and other characteristics are unknown. Finally, the user's intention, motivation, and goal cannot be determined (Jansen 2006). Despite these shortcomings, we believe that TLA of History Trek will provide insight into the search behaviour of users when they are provided with several retrieval options. We are not aware of any other study of a children's portal that utilizes a large number of transaction logs to investigate search patterns.

Methodology

A children's portal, History Trek, was created using the Bonded Design methodology, which employed an intergenerational team comprising adult designers and young students collaborating over thirteen one-hour sessions. The students were an integral and essential part of the team as they provided the perspective of the target audience—children (Large et al. 2007). History Trek provides access to more than 2,300 webpages in English, French, or both that are appropriate in content and language for elementary students and relate to some aspect of Canadian history. The portal is bilingual: English-language webpages can be found when searching within the English-language interface, and French-language webpages from within the French-language interface. Users can switch from one interface to the other by clicking on a button on the home page of each version. The "En français" button to switch from the English to the French version can be seen in figure 1.

The portal includes the several retrieval and browsing mechanisms: keyword searching, topic taxonomy (a hierarchical subject directory), an alphabetical word search, and an advanced search that allows a search to be restricted to words in the title of a website, to subject descriptors that

Figure 1: Four main search options

have been added to all websites (each record includes a number of hyperlinked subject index terms assigned by professional librarians), or to adjacent words (phrases). A scrollable timeline was also included to facilitate browsing for major events in Canadian history. The portal includes help screens and access to several Web-based Canadian history quizzes (figure 1).

Web traffic transaction data have been stored since History Trek became publicly available in September 2007. The portal was originally designed as an experimental site and, as such, a standard format for recording transactions as defined by the World Wide Web Consortium (<http://www.w3.org/TR/WD-logfile.html>) was not employed. The logs are further limited in the amount of data that they contain, since tracking cookies were not used to trace the IP addresses. We felt that a children's portal should be as free as possible from potential technological barriers that may hinder its use in public schools and libraries: many of these institutions try to protect their young users by forbidding access to intrusive websites, many of which utilize tracking cookies.

Between September 2007 and January 2010, more than 800,000 transactions were recorded from around the world. Naturally, a portal on Canadian history is popular in Canada, with 34.1 per cent of transaction logs for searching generated in that country. The United States accounts for 55.9 per cent of transactions, while the remaining 10 per cent are

Table 1: Sample of History Trek transactions

Country	2007 (Sept.)	2008	2009	2010 (Jan.)	Total	%
United States	12,099	74,912	143,944	10,937	24,1892	55.9
Canada	27,699	60,107	52,284	7,601	14,7691	34.1
China	28	899	9,288	1,519	11,734	2.7
Spain	8	1,174	2,527	207	3,916	0.9
Sweden	130	2,251	1,250	17	3,648	0.8
Netherlands	20	384	2,859	265	3,528	0.8
United Kingdom	104	1,317	748	152	2,321	0.5
France	405	1,122	410	177	2,114	0.5
Italy	369	1,494	122	78	2,063	0.5
Germany	31	956	778	70	1,835	0.4
Korea, Republic of	32	574	616	1	1,223	0.3
Russian Federation	5	487	460	241	1,193	0.3
Japan	12	111	789	121	1,033	0.2
Israel	7	419	92	404	922	0.2
Poland	4	809	62	0	875	0.2
Brazil	4	76	702	5	787	0.2

distributed among eighty-three countries, with the bulk generating from China (table 1).

Figure 2 shows the steps taken to derive the 92,226 transactions containing data on information searching. The selection of the transactions was based on unique IP addresses, indicating unique access points. While each IP may not represent one user (for example, in a school the IP may be shared by many users), we can assume that the transactions reflect the information searching patterns of young users on History Trek.

Results and discussion

The transaction log analysis shows that, overall, the Topic taxonomy is the most popular search option at 42 per cent of transactions, followed by the Alphabetic searching at 41 per cent. The Keyword search option

was used in only 15 per cent of cases, with the Advanced Search the least popular among the four search options at only 3 per cent (table 2). While we do not know the exact motivation or the intention of system users for choosing one (or more) options over others, we can speculate that they choose the topic and alphabetic options to explore and browse for assigned or self-generated tasks.

The search data can be further divided into English-language and French-language searches. As the data indicate, there is a significant

Table 2: Search options

Options	Total use	Total (%)	English	English (%)	French	French (%)
Topic	38,633	42	30,368	44	8,265	35
Alphabetic	37,416	41	23,981	35	13,435	56
Keyword	13,727	15	12,568	18	1,159	5
Advanced	2,450	3	1,515	2	935	4
Total	92,226	100	68,432	100	23,794	100

difference in the search options chosen between the two languages ($\chi^2 = 4744.25$, $df = 3$, $p = .000$). While 44 per cent of the users chose the Topic search in English, only 35 per cent chose this option in French, but 56 per cent utilized the Alphabetic search in French as opposed to 35 per cent in English. Significant differences can also be observed in using the Keyword and Advanced options; 18 per cent in English versus only 5 per cent in French, and 2 per cent in English versus 4 per cent in French, respectively. We can only speculate about the explanations for these differences, which we did not find when we observed elementary school students from English-language and French-language schools searching for information under experimental conditions (Large et al. 2006). Judging by the log data's indication of the users' country of origin, it seems likely that the majority of the users on the French-language interface were not in fact francophone (less than 15 per cent of all the transactions were generated from a francophone country or province). In such cases the users may well have found it difficult to derive keywords in French or even to select French terminology from the topic structure; alphabetical searches based on clicking the first letter of a term is the easiest option for a user unfamiliar with the language.

Topic taxonomy

The topic search comprises a taxonomy of 2,674 terms (1,348 English, 1,326 French) relating to Canadian history, created by the original History Trek research team, with eight main topics and up to four levels of depth within these topics. The topic levels are not created equally; for some of the topics the taxonomy does not go beyond level 3 (table 3). Furthermore, since the terms in the taxonomy were developed specifi-

Table 3: Taxonomy levels

Taxonomy levels	English terms	French terms
Level 1	8	8
Level 2	129	188
Level 3	671	873
Level 4	540	257
Total	1,348	1,326

Table 4: Usage frequency of the topic taxonomy

Taxonomy level	Access frequency	%
Level 2	44,207	38.8
Level 3	60,988	53.5
Level 4	8,664	7.6

cally to reflect the subject content of the webpages included on the portal, and these pages were different in French and English (although in many cases the same content was available from bilingual sites), the number of terms employed in the English and in French taxonomies is different.

Table 4 shows the usage frequency of the second, third, and the fourth levels of the taxonomy. The fact that some of the topics in the taxonomy do not go beyond level 3 may explain the much smaller usage frequency for level 4, at only 7.6 per cent. Although browsing the Topic taxonomy may appear to be child-friendly, it poses its own problems for effective information retrieval.

Ideally, the Topic taxonomy should allow children to recognize concepts that are relevant to their information needs, potentially involving less cognitive effort than retrieving from memory terms to be used for a keyword search. Taxonomies in children's Web portals such as Yahoooligans! (<http://www.squirrelnet.com/search/yk/yahoooligans.asp>), KidsClick (<http://www.kidsclick.org/>), and History Trek may, however, impose heavier cognitive loads than keyword searching. The child must recognize a suitable entry point into the taxonomy, and that should be a straightforward so long as the terms displayed in the taxonomy are relevant to the child's information needs. It would be challenging for a child if a relevant term is not displayed at the first level of the taxonomy; the greater the depth of the taxonomy, the greater the potential for failure (Large et al. 2009). The challenges facing the designers of the taxonomy's categories, which should try to reflect a child's categorization of any particular subject (Bilal and Wang 2005) are complicated. Nevertheless, the TLA does demonstrate users' willingness to explore the Topic taxonomy. Its usage is significantly higher than keyword searching, and level 2 and level 3 of the taxonomy are visited frequently.

Alphabetical searching

In the Alphabetical search the first letter of a word is chosen, displaying a list of all subject descriptors beginning with that letter. Spelling is a major obstacle for many young users (Bilal 2000). The Alphabetical search option allows children to scan through letters of the alphabet as if they were flipping through pages of a book. Shenton and Dixon (2003) found that when children were searching for precise information in printed sources, they often flipped through the pages of the entire book, attempting to sequentially access the required information. The TLA shows that 41 per cent of the total number of requests was for the Alphabetical option, demonstrating the system users' preference for this option over keyword searching.

Keyword searching

History Trek Keyword searching incorporates spellchecking (but only with English-language words), right-hand stem truncation (using Porter's stemming algorithm) and a limited amount of control for synonyms (for example, the keyword *automobile* will also find *cars*). As Google is the most widely used search engine by adults and younger users, it is somewhat surprising that History Trek's Keyword search option was utilized relatively infrequently. Druin (2005), however, found similar results for the children's digital library project—visitors used the graphical tools for searching about 90 per cent of time versus using text keyword searching at approximately 10 per cent of time.

Advanced searching

The Advanced search consists of three restrictive options: Title words, Subject descriptors (assigned to each record by professional librarians), and Words in a phrase. Given that Advanced searching is a more sophisticated version of keyword searching, the result of the TLA showing very little use of Advanced searching is not surprising. It also conforms to the search behaviour observed by students when using History Trek in either experimental or operational environments (Large et al. 2006).

Conclusion

History Trek is perhaps the only example of a children's Web portal that is designed by an intergenerational team actually including children, and

fully functioning on the Web. This paper presents analysis of the search patterns based on the transaction logs from a very large number of users from around the world. TLA shows that 83 per cent of all users tend to prefer browsing through a topic taxonomy or alphabetical lists than keyword searching (and especially advanced keyword searching). Transaction logs cannot tell us whether the users were children or adults, but the interface design as well as the portal content make it highly probable that in practice the overwhelming majority of users attracted to this site would be children (with perhaps their teachers as the most prevalent adult users community).

These findings confirm that young people, in principle, seem to opt for browsing just because it is less demanding cognitively, since it imposes an information structure that immediately restricts choice to a limited number of options (Borgman et al. 1995). The overwhelming majority of information retrieval systems such as Google, however, are designed to accommodate keyword searching, which require users to recall from memory appropriate keywords, placing a heavier cognitive load upon the young user than browsing. System developers may be well advised to incorporate search options such as subject taxonomies and alphabetical searching for portals designed for younger audiences. The problem, here, of course, is that not only must appropriate taxonomies be created, but webpages must then be assigned terms from the taxonomy. In the absence of effective automated indexing capabilities, this imposes heavy demands upon human indexers. The question, then, is whether the gains in retrieval effectiveness compensate for the increased development and maintenance costs of the portal—as with so many facets of life, a question of economics!

References

- Aula, A., R.M. Khan, and Z. Guan. 2010. How does search behavior change as search becomes more difficult? *CHI 2010: Proceedings of the 28th International Conference on Human Factors in Computing Systems*, Atlanta, GA, 35–44. New York: ACM.
- Asunka, S., H.S. Chae, B. Hughes, and G. Natriello. 2009. Understanding academic information seeking habits through analysis of Web server log files: The case of the teachers college library website. *Journal of Academic Librarianship* 35 (1): 33–45.
- Beheshti, J., L. Bowler, A. Large, and V. Nessel. 2006. Towards an alternative information retrieval system for children. In *New directions in cognitive information*

- retrieval, ed. Amanda Spink and Charles Cole, 139–65. The Information Retrieval Series, 19. Amsterdam: Springer.
- Bilal, D. 2000. Children's use of the Yahoooligans! Web search engine: 1. Cognitive, physical and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science* 51 (7): 646–65.
- Bilal, D., and J. Kirby. 2002. Differences and similarities in information seeking: Children and adults as Web users. *Information Processing & Management* 38 (5): 649–70.
- Bilal, D., S. Sarangthem, and I. Bachir. 2008. Toward a model of children's information-seeking behavior in using digital libraries. ACM International Conference Proceeding Series. *Proceedings of the Second International Symposium on Information Interaction in Context*, London, UK, 145–51. New York: ACM.
- Bilal, D., and P. Wang. 2005. Children's conceptual structures of science categories and the design of Web directories. *Journal of the American Society for Information Science and Technology* 56 (12): 1303–13.
- Blecic, D., N.S. Bangalore, J.L. Dorsch, C.L. Henderson, M.H. Koenig, and A.C. Weller. 1998. Using transaction log analysis to improve OPAC retrieval results. *College and Research Libraries* 59 (1): 39–50.
- Borgman, C., S. Hirsh, V. Walter, and A.L. Gallagher. 1995. Children's searching behavior on browsing and keyword online catalogs: The Science Library Catalog project. *Journal of the American Society for Information Science* 46: 663–84.
- Chelton, M.K., and C. Cool, eds. 2007. *Youth information-seeking behavior II: Context, theories, models, and issues*. Lanham, MD: Scarecrow.
- Dresang, E.T. 2005. The information-seeking behavior of youth in the digital environment. *Library Trends* 54 (2): 178–96.
- Druin, A. 2005. What children can teach us: Developing digital libraries for children with children. *Library Quarterly* 75: 20–41.
- Hirsh, S.G. 1999. Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science* 50 (14): 1265–83.
- Jansen, B.J. 2006. Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research* 28: 407–32.
- Jansen, B.J., and A. Spink. 2005. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management* 42: 248–63.
- Kuhlthau, C.C. 1991. Inside the search process: Information seeking from the users' perspective. *Journal of the American Society for Information Science* 42 (5): 361–71.
- Large, A., and J. Beheshti. 2000. The Web as a classroom resource: Reactions from the users. *Journal of the American Society for Information Science* 51: 1069–80.
- Large, A., J. Beheshti, V. Nasset, and L. Bowler. 2004. Designing Web portals in intergenerational teams: Two prototype portals for elementary school students. *Journal of the American Society for Information Science and Technology* 55 (13): 1150–4.

———. 2006. Web portal design guidelines as identified by children through the processes of design and evaluation. *Proceedings of the American Society for Information Science and Technology Annual Meeting, Austin, Texas, November 2006*. Silver Spring, MD: ASIST.

———. 2007. Children's Web portals: Can an intergenerational design team deliver the goods? In *Youth information-seeking behavior*, ed. Mary K. Chelton and Colleen Cool. Lanham: Scarecrow, 2: 279–311.

Large, A., V. Nasset, and J. Beheshti. 2008. Children as information seekers: What researchers tell us. *New Review of Children's Literature and Librarianship* 14: 121–40.

Large, A., V. Nasset, N. Tabatabaei, and J. Beheshti. 2009. Bonded design revisited: Involving children in information visualization design. *Canadian Journal of Information and Library Science* 32: 107–40.

Shenton, A.K., and P. Dixon. 2003. A comparison of youngsters' use of CD-ROM and the Internet as information resources. *Journal of the American Society for Information Science and Technology* 54 (11): 1029–49.