An Interface for Opportunistic Discovery of Information for Young People

Jamshid Beheshti and Andrew Large

School of Information Studies, McGill University, 3661 Peel
Montreal, Quebec H3A 1X1, Canada
{jamshid.beheshti,andrew.large}@mcgill.ca

Abstract. The exploratory study investigates a virtual reality interface for opportunistic discovery for young people. We recruited ten volunteer students, who performed a pre-assigned information search task in a VR library which consists of about 1500 web sites depicted as books organized on the shelves. We calculated a metric for opportunistic discovery of information (ODI) based on the distance of books chosen by the participants from those initially found by a conventional search. The VR interface supported ODI more than do conventional search engines.

1 Introduction

The objective of this exploratory study was to observe the information behavior of young people in a virtual reality environment designed mainly for browsing and opportunistic discovery of information. Research suggests that browsing may be a viable alternative to keyword searching for children and young adults, who might otherwise have problems in seeking information from the web (see for example, [8], [13]). Browsing, however, is not a simple act of scanning for information. Wilson devised a typology of browsing consisting of: passive attention, passive search, active search, and ongoing search [25]. Toms discussed serendipitous information retrieval in the context of browsing: "People immerse themselves in the items that interest them, meandering from topic to topic while concurrently recognizing interesting and informative information en route." They browse, scan, and read new information, the purpose of which was not their original intent [22, p.3]. De Bruijn and Spence define opportunistic browsing as: "the continuous but largely unconscious monitoring and filtering of information with the potential to trigger more purposeful behaviour [7, p. 363]. André et al [1, p. 2033] define serendipity as "the act of unexpectedly encountering something fortunate", which they regard as "valuable". They use a matrix to describe various types of serendipity, including finding information that is not relevant to the initial information need. Here, we focus on opportunistic discovery of information (ODI)[1], (or serendipitous discovery of information) which may be defined as 'information encountering' [9] that happens when one is looking for

[1] Sanda Erdelez, Guilherme DeSouza, Chi-Ren Shyu, Antonie Stam and Kevin Wise. International Workshop on Opportunistic Discovery of Information, October 21-22, 2010, Columbia, Missouri.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 541–548, 2011.

information on one topic and on the way finds information on another topic, which may or may not be relevant to the information need.

Browsing is primarily a visual activity [16], and visual-based exploratory interfaces support search activities for learning and investigating [17]. In this project, we have chosen virtual reality[2] technology as a means to facilitate visual ODI. VR environments have been used successfully in the entertainment and gaming worlds, with which our targeted audience is likely to be familiar and enjoys interacting. While young people are increasingly using electronic resources and the Internet they are still well acquainted with traditional libraries [23]. This familiarity with the library environment leads us to believe that the VR environment should represent a physical library. We chose the library metaphor to capitalize on the navigational affordances of recognized artifacts [21] over reconstructing a new model in memory [24]. The library metaphor has been used in experimental projects for more than a decade, with different degrees of success [2] [19] [20]. Zavesky et al [27, p. 619] suggest: "To fully utilize a user's inspection ability, a system must be engaging … Traditional page-based navigation is time-consuming and can be boring." For young people in particular the information environment must be engaging to ensure continual scanning and discovery of information. Engagement is also a key concept in education and in enquiry-based learning. Opportunistic discovery of information may enhance and augment the learning experience of young users of information systems.

Our research sought to answer the question: Is a virtual reality library representing the real world a more effective environment for opportunistic discovery of information for young people than a conventional information retrieval system, such as Google?

2 VRLibrary

We have developed a VR library, (henceforward called "VRLibrary") using the metaphor of a physical library with rooms, bookcases and books. The user, just as in a physical library, can walk around the library, move among the bookcases, scan the titles of books that are arranged on the bookshelves, select individual books, and open them. The difference is that the library is virtual and the books actually represent websites; when a book is "opened" it displays the contents of the web site in a window (for more details see [3]). In this environment, users can utilize *search stations* (Figure 1) located in different locations of the library to conduct conventional keyword and term searches, the results of which are displayed as red dots on a plan of the library. The user can then 'walk' through the library to spot red arrows and dots pointing to the "books" found by the search (Figure 1 and Figure 2). These features were added to the system based on the observations and results from our previous experiments [4] [5].

VRLibrary contains about 1500 links to English-language websites on Canadian history deemed to be appropriate in content and language for elementary students. The database of links was created for History Trek (www.historytrek.ca), a portal designed for children developed by an intergenerational team of researchers and children [14]. The Dewey Decimal Classification (DDC) system has been used to classify all the websites to provide a structure and organization for VRLibrary similar that used in a typical public or school library (Figure 2).

[2] Throughout this paper, virtual reality (VR) refers to desktop-based applications, where the environment is projected on a monitor screen and is non-immersive.

Fig. 1. VRLibrary with a *Search Station*

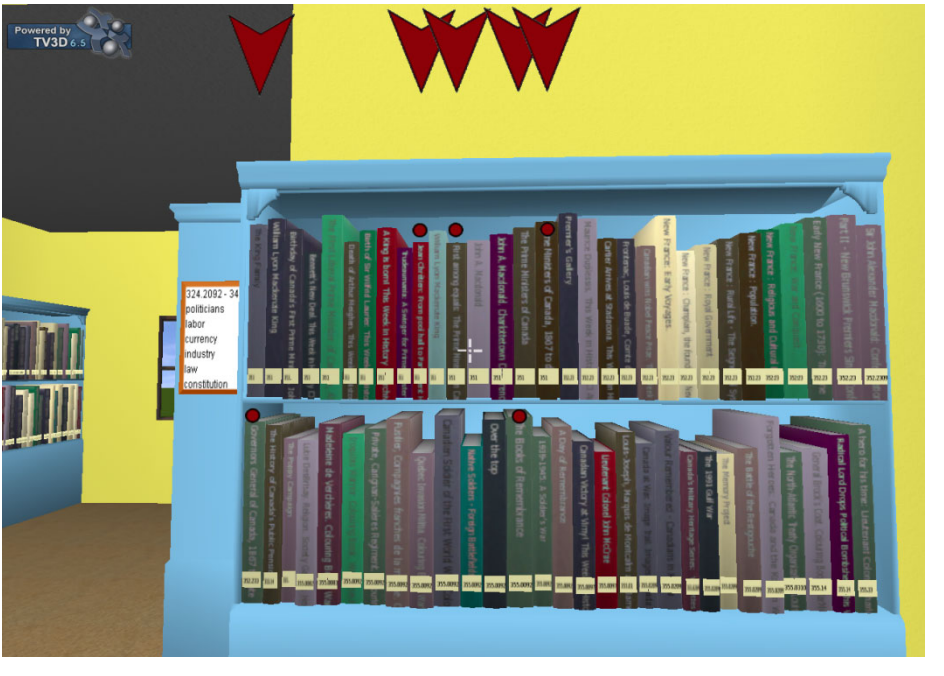

Fig. 2. Search results (websites) indicated by arrows

3 Conceptual Framework

Lehmann et al [15] used the degree of interest (DOI) to identify articles of potential interest in Wikipedia in a browsing context. We use a modified version of DOI to postulate a metric for the ODI itself:

$$ODI_{(v)} = \Sigma |V - D_{(v)}| \quad (1)$$

where V is the target website, and D is the distance of a website of interest from the target website measured in terms of information units (in this case number of websites). Although the distance $D_{(v)}$ is a function of the classification system used to organize the websites, it may be viewed as a relative measure of the browsing activity and ODI. Given the relatively small size of the database of website links in our study, we can determine the websites that are relevant for any specific information task. We can then measure the distance $D_{(v)}$ from the target website V located by the initial search to the location where opportunistic browsing or ODI has taken a user, and where other websites have been discovered. The average $ODI_{(v)}$ over all users will yield a relative measure of the browsability and opportunistic discovery of the system.

To contextualize and benchmark the metric, we calculated the highest theoretical value for $ODI_{(v)}$ for Google. Assuming that the first link to a website on the first page (screen) of the results from a Google search is a potential target website V, the maximum $ODI_{(v)}$ for the default first page containing 10 links is calculated as 45. If we include the second page of results containing 10 additional links, then the maximum $ODI_{(v)}$ is 190. However, research shows that users often do not venture beyond the first few hits, let alone the first page [11] [12]. Therefore, in practice, for Google the maximum likelihood $ODI_{(v)}$, assuming 10 links per result page is 45.

4 Methodology

4.1 Participants

Ten volunteer students, equal numbers of boys and girls ranging between 11 and 13 years old, participated in the study. A pre-test questionnaire was designed to solicit from them demographic information as well as frequencies of internet and video game usage, library visits, and information-seeking habits. All participants but one were in grades 7 and 8 (the youngest participant was in grade 6). Although they had a variety of ethno-cultural backgrounds, they were all educated in Quebec and were familiar with Canadian history. Half of the participants used the internet every day, while the other half used it at least a few time per week. They all used Google to find information for their homework and for leisure, and employed both searching and browsing strategies. The majority (7 students) used a library at least once a week, while the remainder used it at least once a month. All but two played video games, ranging in frequency from everyday to less than once per month. All were familiar with the experimental subjects chosen for the retrieval tasks.

4.2 Tasks

The experimental tasks consisted of asking participants to find information (or websites) on a topic related to Canadian history, in the context of a school project. The first task involved finding information on *holidays in Canada*. The database contains five links to websites deemed to be directly relevant to the task. For this particular topic, most of the neighbouring websites may not be directly relevant to the topic (target websites), and therefore participants should choose to ignore them. The second task was to find websites about a *parliament building*, with three relevant links to websites in the database. These tasks were chosen to represent typical topics covered in the history curriculum, and with which students were likely to be familiar. Using the DDC classification for a small collection of books resulted in dispersed clusters of subjects, which may not have been inductive to browsing. Nevertheless, it is a classification system used most often in school and public libraries.

4.3 The Procedure

Each student undertook one task independently. At the outset the purpose of the research was briefly explained to the student by the research assistant, following a script to ensure consistency across students/assistants. It was emphasized that the interfaces and not the student was being evaluated. Each student was introduced to the project with a brief demonstration of VRLibrary, showing the search and browsing features, which took about five minutes. The student was then given one of the two tasks and was asked to retrieve relevant websites in the VRLibrary. They were also asked to limit their searching and browsing to the first page of the website and not to follow the internal hyperlinks on the site. Participants' information seeking process was observed unobtrusively by a research assistant. Once students were satisfied with the information (websites) they had selected, a post-test interview was conducted to determine their impressions of the VRLibrary, and the reasons for their selections of websites other than those found in the initial search.

Four students were asked individually to locate information on the first task, *holidays in Canada*, while six were given the second task, *parliament building*. The research assistants noted the student's affective behaviour as well as any obvious problems, comments or questions raised by the student. The on-screen activity of the students as well as any spoken commentary by either student or research assistants was recorded.

4 Results

All students began with keyword or term searching on one of the search stations in VRLibrary. In practice it is difficult, if not impossible, to begin browsing immediately before identifying a starting point in the collection. On Task 1, finding information on *holidays*, three of the four students managed to locate (by clicking on the book spines) the five target websites with relative ease. The first student continued browsing the shelves after finding the target websites and clicked on three additional books (and opened the websites) before going back to the search station. The distances between these additional books (websites) and the closest target websites

were calculated based on the order of the books arranged in DDC in the database. We assumed that any book that was chosen directly as a result of a keyword/term search on the search stations should not be counted as an ODI act, and therefore did not include them in our calculations. The next two participants, while browsing the shelves, did not click on any other books but the target websites, and were satisfied with their search results. The last participant to perform Task 1, also browsed and clicked on several books that were relatively far from the target websites. The ODI for Task 1 was calculated using the formula (1), and over the four participants averaged 156 books (links to websites).

Six students performed Task 2, finding information on *parliament building*. The ODI activity for each participant varied significantly, ranging from 0 to 589 books. The average ODI $_{(v)}$ for Task 2 was 168 books.

Participants' subjective evaluation of the system, captured through the post-test interviews, showed a very high level of satisfaction with VRLibrary. The young people in the study used terms such as 'cool', and 'easy to use' to describe VRLibrary, and found the environment very appealing and engaging.

5 Discussion and Conclusions

Two experimental tasks were performed by 10 participants to explore the opportunistic discovery of information (ODI) of a virtual reality library. A new metric was conceived to quantify ODI, defined as the sum of distances of the links to websites of interest from the target websites retrieved in a conventional search. The ODI for one task was calculated as 156 links and for another task was 168 links. Equivalent tasks on Google show a maximum theoretical ODI of 45 links, assuming young users do not browse past the first result page. These results suggest that VRLibrary with average ODIs higher than the theoretical value calculated for Google may inspire young people to discover information more readily than conventional search engines.

In one instance during the experiment, a student chose through ODI two completely unrelated books because she was "curious". The element of curiosity or surprise is a prerequisite for ODI [18]. In one study, Bilal found that for self-generated tasks, most children in her experiment visited websites that were irrelevant to their task, and concluded that they likely wanted to "explore topics of interest other than those they were pursuing" [6, p. 1177]. While we did not build any triggering mechanism for discovery in VRLibrary, as recommended by some researchers [1] [10], it seems that the system supports and perhaps encourages ODI.

Caution must be exercised in generalizing from one small-scale study using one interface. First, the observed wide variations among the participants in searching and browsing in the VRLibrary may suggest different cognitive styles in information behaviour among the young users. Westerman et al [25] found in their study of testing a two and three-dimensional virtual information space that the adult participants with lower cognitive ability out-performed others in the three-dimensional environment, in terms of recall and precision, and performance efficiency. To use quantifiable measures and a metric such as the ODI $_{(v)}$, larger samples may produce more precise results in experimental studies. Second, the organization of the information (or the links), may affect opportunistic discovery. In the VRLibrary, we used the Dewey

Decimal Classification to organize a small collection of links on Canadian history to mimic the organizational structures of many real-world public and school libraries. Other clustering methods, such as automatic text categorization (TC) using machine learning algorithms, or self-organizing maps [20] may have different effects on ODI. Although logistically not feasible, for comparative analysis a similar algorithm as in Google's Page ranking should be utilized to organize the links in the VRLibrary to measure ODI. Finally, McCay-Peet and Toms discuss the challenges and tribulations of using experimental laboratory goal-oriented methods to measure ODI [18]. Ideally, a field study under operational setting, where students would have the option of using VRLibrary for completion of classroom projects, will be the ultimate test for measuring ODI in a virtual environment.

Despite the limitations of the study, however, a common theme emerged from the experiment. VRLibrary's basic premise of using a real-world metaphor seems to be an effective and exceptionally engaging method of presenting information to young users for opportunistic discovery of information.

Acknowledgements. The VRLibrary program was developed by Ian Clement; his expertise is gratefully acknowledged, as is the help of our research assistants, Nouf Khoshman and Megan Stecyk. The research could not have been conducted without the student volunteers. This project was funded by the Social Sciences and Humanities Research Council (Canada).

References

1. André, P., Teevan, J., Dumais, S.T.: From X-rays to silly putty via Uranus: Serendipity and its role in web search. In: CHI 2009: Proceedings of the 27th International Conference on Human Factors in Computing Systems, April 9. ACM, New York (2009)
2. Beheshti, J., Large, V., Bialek, M.: Public Access Catalogue Extension (PACE): A Browsable Graphical Interface. Information Technology and Libraries 15(4), 231–240 (1996)
3. Beheshti, J., Large, A., Julien, C.A.: Designing A Virtual Reality Interface for Children's Web Portals Data, Information, and Knowledge in a Networked World. In: Annual Conference of the Canadian Association for Information Science, London, Ontario, June 2-4 (2005)
4. Beheshti, J., Large, A.: Preliminary Design Indicators to Desktop Virtual Reality Environments. In: ASIST 2006 Annual Meeting, Information Realities: Shaping the Digital Future for All. SIG HCI Research Symposium: Human-Computer Interaction in Information Intensive Environments, Austin, Texas, November 5 (2006)
5. Beheshti, J., Large, A., Clement, I., Tabatabaei, N.: Evaluating the usability of a virtual reality information system for children Information Sharing in a Fragmented World: Crossing Boundaries. In: Annual Conference of the Canadian Association for Information Science, Montreal, Quebec, May 10-12 (2007)
6. Bilal, D.: Children's use of the Yahooligans! Web search engine. III. Cognitive and physical behaviors on fully self-generated search tasks. Journal of the American Society for Information Science and Technology 53(13), 1170–1183 (2002)
7. de Bruijn, O., Spence, R.: Serendipity within a ubiquitous computing environment: A case for opportunistic browsing. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 362–369. Springer, Heidelberg (2001)

8. Druin, A., Bederson, B.B., Hourcade, J.P., Sherman, L., Revelle, G., Platner, M., Weng, S.: Designing a digital library for young children: an intergenerational partership. In: Proceedings of JCDL, pp. 398–405. ACM Press, New York (2001)
9. Erdelez, S.: Information encountering: It's more than just bumping into information. Bulletin of the American Society for Information Science, 25–29 (Febrary/March 1999)
10. Erdelez, S.: Investigation of information encountering in the controlled research environment. Information Processing and Management 40(6), 1013–1025 (2004)
11. Guan, Z., Cutrell, E.: An eye tracking study of the effect of target rank on web search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York (2007)
12. Joachims, T., Granka, L.: Accurately Interpreting Clickthrough Data as Implicit Feedback. In: ACM SIGIR 2005, pp. 154–161 (2005)
13. Large, A., Beheshti, J.: The Web as a classroom resource: reactions from the users. Journal of American Society for Informational Science 51(12), 1069–1080 (2000)
14. Large, A., Beheshti, J., Nesset, V., Bowler, L.: Children's Web Portals: Can an Intergenerational Design Team Deliver the Goods? In: Chelton, M.K., Cool, C. (eds.) Youth Information-Seeking Behavior, vol. 2, pp. 279–311. Scarecrow, Lantham (2007)
15. Lehman, S., Schwanecke, U., Dörner, R.: Interactive visualization for opportunistic exploration of large document collections. Information Systems 35(2), 260–269 (2010)
16. Marchionini, G.: Information seeking in electronic environments. The Press Syndicate of the University of Cambridge, Cambridge (1995)
17. Marchionini, G.: Exploratory search: From finding to understanding. Communications of the ACM 49(4), 41–46 (2006)
18. McCay-Peet, L., Toms, E.G.: The Process of serendipity in knowledge work, August 18-21. ACM, New Brunswick (2010)
19. Pejtersen, A.M.: A Library system for information retrieval based on a cognitive task analysis and supported by an icon-based interface. In: SIGIR 1989 Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 40–47 (1989)
20. Rauber, A., Merkl, D.: The SOMLib digital library system. In: Abiteboul, S., Vercoustre, A.-M. (eds.) ECDL 1999. LNCS, vol. 1696, pp. 323–342. Springer, Heidelberg (1999)
21. Shiaw, H., Jacob, R.J.K., Crane, G.R.: The 3D vase museum: a new approach to context in a digital library. In: Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, June 7-11, pp. 125–134 (2004)
22. Toms, E.: Serendipitous information retrieval. In: Proc. of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries (2000)
23. Walter, V.A.: Public library service to children and teens: A research agenda. Library Trends 51(4), 571–589 (2003)
24. Ware, C.: Visual thinking for design. Morgan Kaufmann, Burlington (2008)
25. Westerman, S.J., Collins, J., Cribbin, T.: Browsing a document collection represented in two- and three-dimensional virtual information space. International Journal of Human-Computer Studies 62(6), 713–736 (2005)
26. Wilson, T.D.: Information behaviour: an interdisciplinary perspective. Information Processing and Management 33(4), 551–572 (1997)
27. Zavesky, E., Chang, S.F., Yang, C.C.: Visual Islands: Intuitive browsing of visual search results. In: CIVR 2008: Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval, Niagara Falls, Ont, July 7-9. ACM Press, New York (2008)