Exploiting Browsing History for Exploratory Search

Wei-Lin Chen and Wei-Guang Teng

Department of Engineering Science,
National Cheng Kung University,
Tainan, Taiwan
realseven@gmail.com, wgteng@mail.ncku.edu.tw

Abstract. With the advance of information technologies, web search has become a necessary activity for most Internet users. Although current search engines are powerful enough to respond required search results in seconds, information seekers may still feel arduous to digest them when conducting the exploratory search. In other words, when users are unfamiliar with the domain of their goals or unsure about the ways to achieve their goals, they may need to read numerous pages before fully understand what they are searching for. In this work, we propose to fully exploit the browsing history to help users clarify their thoughts and discover new insights during the process of exploratory search. Specifically, interactive user interfaces of two different modes, i.e., the timeline mode and the relevance mode, are devised to provide users a vivid impression of their browsing history. Hopefully, the feasibility and usability of web search for users can be further improved.

Keywords: Exploratory search, information seeking, interactive interface, timeline.

1 Introduction

As the data amount gathered on the Internet increases at a very rapid speed, web search is becoming one of the most frequent tasks performed by the users. With proper query terms, one may reach required information or perceive some knowledge by exploring web pages in the search results. Moreover, as is investigated in previous studies, user goals in web search can be classified in to three types, i.e., navigational, transactional and informational search [6,14,21]. For users conducting either a navigational or a transactional search, there are always specific targets, e.g., a website to be visited or an item to be purchased, in their mind. On the other hand, for users conducting an informational search, they may experience a learning process to discover desired information from a number of web pages from corresponding search results.

It is thus observed that user activities in web search vary with their different goals [7]. With current search engines, a simple search activity can usually be completed in seconds. Nevertheless, when conducting the search in an unfamiliar domain, a user tends to browse a few search results and then recursively issue the next query for obtaining some further information. Such a process of *exploratory search* is a time-consuming and tentative task for most of the users since they may need to read numerous pages before fully understand what they are searching for. In other words, if a

M.J. Smith and G. Salvendy (Eds.): Human Interface, Part I, HCII 2009, LNCS 5617, pp. 355–364, 2009.
© Springer-Verlag Berlin Heidelberg 2009

user does not know which keywords to use or looks for several answers, he usually needs to spend much time on querying and browsing activities to accomplish an exploratory search.

When users try to investigate, evaluate, compare and synthesize information pieces grabbed during their exploratory search processes, they may encounter the data overloading problem due to the limited short-term memory of human beings. In view of this, an interesting concept which is inspired from the Pensieve in the famous series of novels "Harry Potter" is noted in this work. Specifically, the Pensieve is a fictional item used to store memories and to provide a near-omniscient perspective of the events to be relived later in the novels. To realize such a concept in practical applications, we propose to develop a proper approach for users to retain and to organize necessary information.

Furthermore, we note that search sessions of a user are not totally irrelevant, especially when the goal of web search is for obtaining more and more knowledge, or more specifically, *learning* something new. Consequently, the browsing history of users can help them on not only reviewing previous search activities but also clarifying next search targets. Nevertheless, to the best of our knowledge, the web browsing history is simply recorded in sequence and is usually loosely organized in previous works. For supporting the user needs of retrospecting their previous activities in exploratory search, we propose in this work to present and to organize browsing history of users in more feasible forms. Specifically, our goals are twofold. First, we expect to provide users a novel interface to support their historical retrospect. Moreover, relationship among search sessions should be identified to help users clarifying their thoughts and discover new insights.

The rest of this paper is organized as follows. Relevant works on exploratory search and usage of browsing logs are generally reviewed in Section 2. The concept and advantages of our approach are explored in Section3. In Section 4, implementation details of our approach are presented. This paper concludes with Section 5.

2 Preliminaries

The process of exploratory search and relevant information theories are generally discussed in Section 2.1. Moreover, in Section 2.2, we show how browsing logs can help to improve the user interactivity in web search.

2.1 Exploratory Search

Search can be considered as a specific type of information seeking behavior. Generally speaking, there are different stages for this goal-oriented and problem-solving process, i.e., problem recognition, problem definition, problem resolution, and solution statement [27]. With the advance of information technologies in recent years, web search has become an increasingly important part of most computer users. There are several previous works on improving search results or achieving personalized search. Typical techniques include reranking search results based on the personal profile [2,22,24], comparing search results of different search engines [18], and offering faceted search interfaces [12]. In addition, there are more and more works focusing on the visualization of search results [7,8].

A search engine can handle numerous user requests at a same time and returns more than a few matching results to the users, respectively. On the other hand, for users who are unfamiliar with the domain of their goals or unsure about the ways to achieve their goals, they may need to spend much more efforts to explore and to digest these documents on the web. An illustrative example is that if a user does not know much about classical music, how should he even begin to find a piece that he might like. This reveals that typical keyword search scenarios are insufficient. Specifically, relevant studies on exploratory search emphasize the circumstances that users start with vague information needs [15,17,25].

When conducting an exploratory search task, a user acquires not only specific documents but also the knowledge discovered during the interaction process with a search engine since the user may keep revising his query terms and thus the search engine returns some new results. When facing a large amount of retrieved data, users need to spent time and effort filtering useful information and organizing their own thoughts. Thus, a good tool supporting in exploratory search should help users to discovering new association and kinds of knowledge, resolving complex information problems, or developing an understanding of the terminology and the information space structures [25].

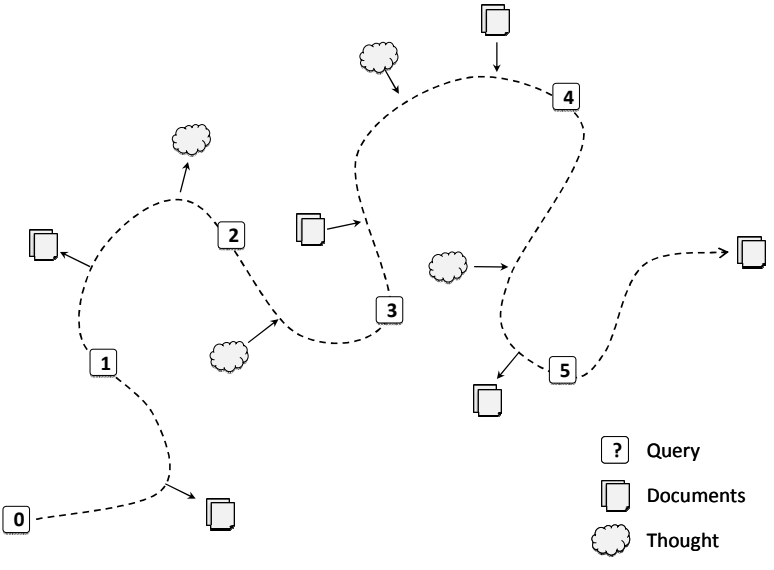

Fig. 1. An evolving search path as described in the berrypicking model (adapted from [5])

An informational model of *berrypicking* [5] best describes the actual behavior of information searchers, especially for the process of exploratory search. Unlike the comments made in conventional models, the search behavior is not merely a single movement for users to reach the best retrieved result, but a berrypicking or an evolving pattern. As shown in Fig. 1, the path of a searcher is usually a curve instead of a straight line. Also, searchers may iteratively change their direction of the query after gaining new knowledge from retrieved documents and clarifying their own thoughts.

In other words, the process of exploratory search usually consists of several sessions of keyword search and is significantly influenced by the current state of the searcher.

2.2 Usage of Browsing Logs

Logs of user activities can be feasible in discovering usage patterns and other applications. For seeking legal information, user interface tools based on search histories are developed [13] and the corresponding results show that the log information is useful for many information-seeking tasks. On the other hand, although current web browsers are usually capable to offer users simple logs of their browsing history, there are only a few basic functions provided for users to access their logs. Usually the most recently-viewed websites are listed so that a user can easily revisit a site even if he forgets the corresponding URL. Also, a user can search the log to find some sites that he may have forgotten he visited.

Note that time is the explicit information recorded with all entries in the browsing log. Thus, the timeline is a common approach to present time sequences [16,6] so that corresponding events can be organized in order. To provide a better visualization form, the utilization of timeline is adopted in many relevant works so that users can understand the relationship among information segments in a much easier way [1,3,6]. For example, news photos on Flickr can also be visualized as a time flow to demonstrate the evolving history [9]. Research papers can also be presented as a timeline with the tool SIMILE [23]. Generally speaking, how to use the temporal information effectively is already a crucial issue in the field of information retrieval [4,20]. For our purpose of exploratory search, the use of time for clustering and browsing online documents may provide information seekers a much distinct way than simply returning some documents in response to a query.

To the best of our knowledge, the records of web browsing history are usually loosely organized in current web browsers. Thus, it is difficult to recognize dissimilarities among browsing records, not to mention to discover the valuable relevance among them. By analyzing the browsing history, browsing patterns of users can be used to recommend websites which users might be interested in [10]. Moreover, for users to get more personalized search results or to follow interesting trends in their web activities, Google also provides similar services for users to review their search history online [11]. In this work, we thus propose to exploit two significant types of information embedded within the browsing history, i.e., record time and document relevance, to provide users better interactivities in conducting the exploratory search.

3 Exploiting Explicit and Implicit Information in Browsing Logs

As users may feel lost in the large quantity of gathered data in exploratory search, we propose to organize the search activities based on both the explicit and the implicit information in browsing logs. Our approach of utilizing the explicit record time is illustrated in Section 3.1. Moreover, the discovery of implicit relevance among search sessions and the corresponding usage are presented in Section 3.2.

3.1 Usage of Timeline and Temporal Granularities

As mentioned in previous sections, logs offered by current web browsers are usually simple and are not well-structured. The arrangement of records is simply ordered by browsing time in conventional approaches. Moreover, the basic unit of the browsing log is either a website or a web page. Nevertheless, for the purpose of exploratory search, we notice that a search session should be a more appropriate unit for tracking the browsing history. Also, we believe that more efforts can be elaborated on utilizing the corresponding temporal information.

Once a user's browsing history is tracked, search sessions along with some timestamps can be defined. Specifically, each keyword search and the corresponding browsing behavior of the result pages are regarded as a *session* in this work. Also, each search result is considered as an *item*. Also, the start timestamp and the duration of each session are crucial in our approach. Instead of presenting a long list of previously visited websites, we choose to organize previous search sessions on a timeline with variable temporal *granularities*.

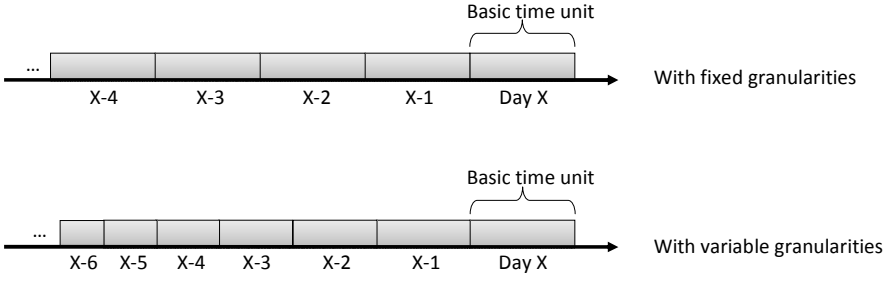

Fig. 2. Timeline with either fixed granularities or variable granularities

Granularities of a timeline can be adjusted according to the temporal distance between a previous search session and the present time. It is observed that a more recent event usually results in a higher impact to people. In other words, the search session issued yesterday is considered more impressive than the one issued a month ago for the same user. The difference of using fixed granularities and variable granularities on a timeline is depicted in Fig. 2. Note that although every time unit stands for a day in Fig. 2, the corresponding width is not fixed in the timeline with variable granularities.

We further incorporate the concept of *tilted-time window* to present the whole timeline in different granularities. In general, recent search sessions are presented in a finer granularity while older ones are presented in a coarser granularity. In addition to such a recent-biased arrangement, search sessions of more interestingness to the user can also be presented in a finer granularity. Such an extension of using variable granularities can be even more illustrative for users. This is similar to the memory of human beings since some significant events can be remembered for a quite long time.

3.2 Usage of Relevance among Search Sessions

Browsing logs not only reveal the behavior of users, but also help to establish the fundamental knowledge of users. Through reviewing the histories, users can reformulate

their thoughts to confirm their next directions in searching. In other words, relationship among previous sessions may help to facilitate the exploratory search since it can be a long-term learning process. This process is just like playing jigsaw puzzles, a user may gradually obtain some clues to place more pieces properly or suddenly encounter a bottleneck preventing him from moving forward. Consequently, search sessions of a user may not occur in series during a complete process of exploratory search. For example, many users search for the business affairs in office hours and search for their casual hobbits after work. Thus, two successive search sessions can be quite irrelevant.

Obviously, if the implicit relationship among search sessions can be identified from browsing logs, the search history can be further organized. Note that since each search session contains the corresponding query terms and a number of items, i.e., web pages in the search result, two relevant sessions may contain duplicate items. In other words, the number of duplicate items between two different sessions is used to measure the strength of their relevance. Moreover, browsed items and non-browsed ones should be of different interestingness to the user. Specifically, if there exists a duplicate item among relevance sessions and is not browsed by the user, then it is regarded as an interesting item.

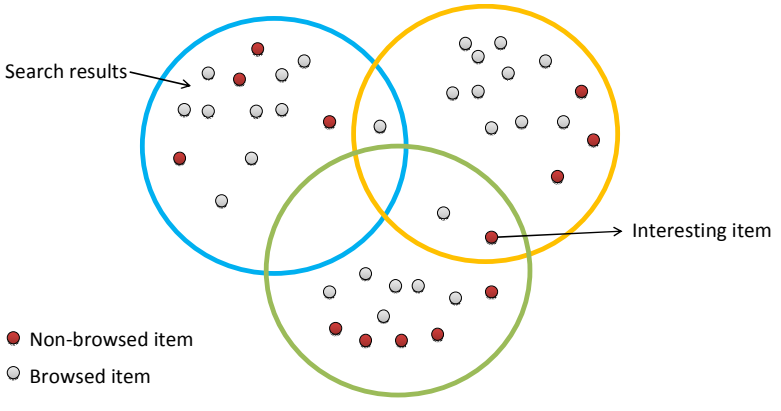

Fig. 3. Duplicate items is used for measuring the relevance among search sessions

As shown in Fig. 3, there are three different search sessions represented as hollow circles in the example browsing history of a user. Grey and red dots represent the browsed items and non-browsed items, respectively. Dots in the overlap region stand for duplicate items. Therefore, the relevance between the yellow and the green sessions is higher than that between the yellow and the blue ones. Also, an interesting item of this example is indicated in Fig. 3.

4 Implementation of Proposed Interfaces

To make previous concepts into practice, we propose new user interfaces to present the organized browsing history. Specifically, users can review their previous search activities in either the timeline mode or the relevance mode. Details of these two modes are illustrated in Section 4.1 and Section 4.2, respectively.

4.1 Timeline Mode

The interface of the timeline mode is as shown in Fig. 4. Search sessions are specified with different colors. On the left is the overall browsing history presented on a timeline with variable granularities. After the user specifies a desirable period to review his or her browsing history, details of all corresponding search sessions are presented on the right.

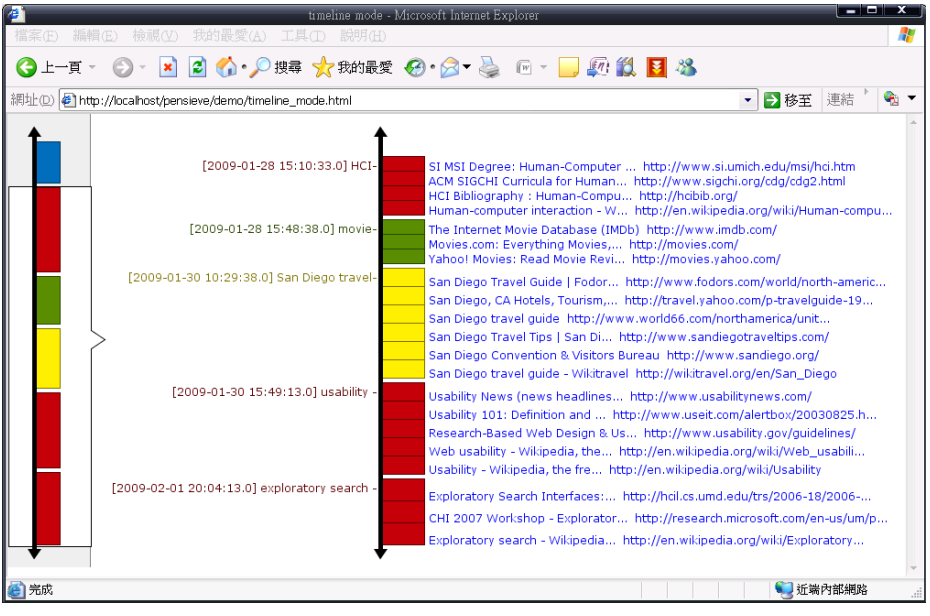

Fig. 4. Visualization of the browsing history in the timeline mode

In the timeline mode, all search sessions are arranged according to their timestamps. The granularities of the timeline vary with the recency and the interestingness of search sessions. Besides, it is noted that relevant sessions are in similar colors so that the user can distinguish them easily. As compared to the usage of a typical timeline of fixed granularities, our approach in timeline mode offers users a vivid impression of their browsing history.

4.2 Relevance Mode

The interface of the relevance mode is as shown in Fig. 5. Our implementation is built on the basis of an interactive visualization toolkit, i.e., prefuse [19]. Search sessions are represented as circles marked with the corresponding query terms. Also note that relevant sessions are in similar colors as the way in the timeline mode.

In the relevance mode, the relationship among search sessions can be easily identified since relevant sessions are connected. Note that user behavior can affect the strength of a relationship since not all search results are browsed by the user.

Generally speaking, our approach in the relevance mode breaks through the hedge of the timeline. Hence, users may obtain more insights about their own browsing history. For the purpose of exploratory search, in addition to what users already know, they can extend the current knowledge base by creating more search sessions of relevant query terms.

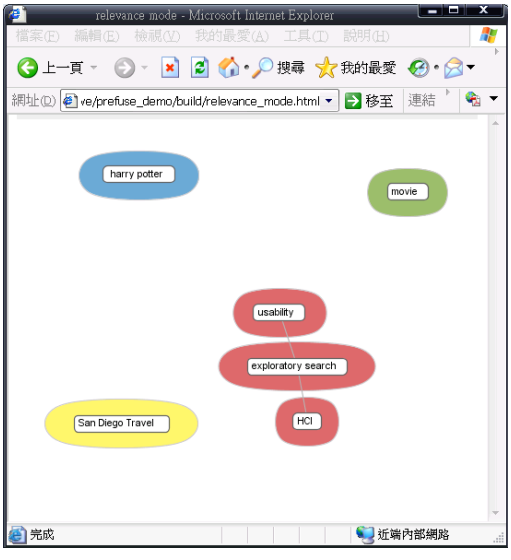

Fig. 5. Visualization of the browsing history in the relevance mode

5 Conclusions

As the tasks and goals involved with exploratory search are probably undefined and unpredictable, it is difficult and subjective for users to obtain insights during the search process. Much effort has been spent on studies in relevant fields, e.g., information seeking and human-computer interaction. In this work, we have proposed to fully exploit the browsing history to help users clarify their thoughts and discover new insights during the process of exploratory search. With one of the motivations being to support users when keyword search is not enough, we have focused on developing novel user interfaces and interaction models that support the user in different ways. Hopefully, the feasibility and usability of web search for users can be further improved as both the explicit and the implicit information of browsing history are properly utilized.

Acknowledgments. The work was supported in part by the National Science Council of Taiwan, R.O.C., under Contracts NSC95-2221-E-006-319-MY3.

References

1. Aigner, W., Bertone, A., Miksch, S., Tominski, C., Schumann, H.: Toward a Conceptual Framework for Visual Analytics of Time and Time-oriented Data. In: Proceedings of the 2007 Winter Simulation Conference, December 2007, pp. 721–729 (2007)
2. Agichtein, E., Brill, E., Dumais, S.: Improving Web Search Ranking by Incorporating User Behavior Information. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 2006, pp. 19–26 (2006)
3. Alonso, O., Baeza-Yates, R., Gertz, M.: Exploratory Search Using Timelines. In: Proceeding of the 25th Annual SIGCHI Conference on Human Factors in Computing Systems, April 2007, pp. 23–26 (2007)
4. Alonso, O., Gertz, M., Baeza-Yates, R.: On the Value of Temporal Information in Information Retrieval. ACM SIGIR Forum 41(2), 35–41 (2007)
5. Bates, M.J.: The Design of Browsing and Berrypicking Techniques for the Online Search Interface. Online Information Review 13(5), 407–424 (1989)
6. Broder, A.: A Taxonomy of Web Search. ACM SIGIR Forum 36(2), 3–10 (2002)
7. Chen, T.-W., Chang, P.-L., Teng, W.-G.: Supporting Informational Web Search with Interactive Explorations. In: Proceedings of the 4th International Conference on Signal Image Technology and Internet Based Systems, November 2008, pp. 153–160 (2008)
8. Dörk, M., Carpendale, S., Collins, C., Williamson, C.: VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery. IEEE Transactions on Visualization and Computer Graphics 14(6), 1205–1212 (2008)
9. Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., Tomkins, A.: Visualizing Tags over Time. In: Proceedings of the 15th International Conference on World Wide Web, May 2006, pp. 193–202 (2006)
10. Fu, X.-B., Sudzik, J., Hammond, K.J.: Mining Navigation History for Recommendation. In: Proceedings of the 5th International Conference on Intelligent User Interfaces, January 2000, pp. 106–112 (2000)
11. Google Web History, `http://www.google.com/searchhistory/`
12. Hearst, M.: Design Recommendations for Hierarchical Faceted Search Interfaces. In: Proceedings of the ACM SIGIR Workshop on Faceted Search, August 2006, pp. 26–30 (2006)
13. Komlodi, A., Soergel, D., Marchionini, G.: Search Histories for User Support in User Interfaces. Journal of the American Society for Information Science and Technology 57(6), 803–807 (2006)
14. Lee, U., Liu, Z., Cho, J.: Automatic Identification of User Goals in Web Search. In: Proceedings of the 14th International Conference on World Wide Web, May 2005, pp. 391–400 (2005)
15. Marchionini, G.: Exploratory Search: from Finding to Understanding. Communications of the ACM 49(4), 41–46 (2006)
16. Müller, W., Schumann, H.: Visualization Method for Time-dependent Data - an Overview. In: Proceedings of the 2003 Winter Simulation Conference, December 2003, pp. 737–745 (2003)
17. Park, H., Myaeng, S.H., Jang, G., Choi, J., Jo, S., Roh, H.: Interactive Information Seeking Interface for Exploratory Search. In: Proceedings of the 10th International Conference on Enterprise Information System, June 2008, pp. 276–285 (2008)

18. Petratos, P.: Informing through User-centered Exploratory Search and Human-computer Interaction Strategies. Proceedings of the Informing Science and Information Technology 5, 705–727 (2008)
19. Prefuse - Interactive Information Visualization Toolkit, http://prefuse.org/
20. Ringel, M., Cutrell, E., Dumais, S.T., Horvitz, E.: Milestones in Time: the Value of Landmarks in Retrieving Information from Personal Stores. In: Proceedings of the 9th IFIP TC13 International Conference on Human-Computer Interaction, September 2003, pp. 184–191 (2003)
21. Rose, D.E., Levinson, D.: Understanding user goals in web search. In: Proceedings of the 13th International Conference on World Wide Web, May 2004, pp. 13–19 (2004)
22. Shen, X., Zhai, C.: Exploiting Query History for Document Ranking in Interactive Information Retrieval. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2003, pp. 377–378 (2003)
23. SIMILE Timeline Toolkit, http://simile.mit.edu/timeline/
24. Speretta, M., Gauch, S.: Personalized Search Based on User Search Histories. In: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence, September 2005, pp. 622–628 (2005)
25. White, R.W., Kules, B., Drucker, S.M., Schraefel, M.C.: Supporting Exploratory Search. Communications of the ACM 49(4), 36–39 (2006)
26. Wild, P.J., Johnson, P., Roast, C., Czerwinski, M.: The Temporal Aspects of Work for HCI. In: Proceeding of the 22nd Annual SIGCHI Conference on Human Factors in Computing Systems, April 2004, pp. 1735–1736 (2004)
27. Wilson, T.: Human Information Behavior. Informing Science the International Journal of an Emerging Transdiscipline 3(2), 49–56 (2000)