New Approaches for Studying and Building Information Seeking Models: A Possible Hybrid Approach

For the SIG USE Research Symposium at ASIST '03 in Long Beach, CA. Oct 18, 2003.

Don Turnbull
School of Information
University of Texas at Austin
1 University Station, D7000
Austin, TX 78712
donturn@ischool.utexas.edu
512.232.3508

A significant amount of the contributions of information seeking theory and models are based on assumptions or smaller-scale studies of user information needs and searching. (N. J. Belkin, N.Oddy, & Brooks, 1982; N. J. Belkin, Oddy, & Brooks, 1982; Kuhlthau, 1991) These models suppose cognitive activity and the feedback given by the information search environment. With the more recent focus on World Wide Web browsing and search a new set of data collection and user modeling studies are possible. While these new Web-based models can be primarily based on detectable patterns in either individual interest (content) or situational activity (context), it is hoped that it may be possible to generalize a set of habits from these patterns that can then be either integrated with existing information seeking models or to form a new basis for information seeking studies.

It is possible that by combining these previous studies with new data collection methods and accounting for newer information seeking technologies such as the Web, we may be able to discover, measure and compare information seeking and information retrieval patterns. Over the last few years Choo, Detlor and Turnbull have worked on a hybrid information seeking framework (based on Ellis and Aguilar) which has potential to be extended or mapped to other information seeking activities such as information search. (Choo, Detlor, & Turnbull, 1998, 1999, 2000a, 2000b)

The research question proposed is it more useful to move towards a more purely quantitative model as seen in studies like Huberman, et al.'s "Law of Surfing" study of aggregate Web use or towards the Information Foraging work championed by Pirolli and Card that focuses more on individual users and their general information seeking habits? (Huberman, Pirolli, Pitkow, & Lukose, 1998; Pirolli & Card, 1995) Moreover, are these fruitful directions to move in light of the significant previous research into information seeking? Is there anything new to discover in information seeking except to apply existing models to Web environments?

This more behavioral-oriented research shows that habits and routines are dominant in terms of information (seeking) behavior and it may be useful to explore the efficacy of the above models towards more finitely describing user behavior, be it assumptions about cognitive activity or noting the impact of various information technologies. One technique increasingly employed in the recent years is a highly quantitative, Knowledge Discovery in Databases (which includes Data Mining) to analyze large-scale datasets of Web use activity. It would be useful to address the applicability of these techniques to either extend or confirm existing information models or devise altogether new models.

One barrier to understanding such large datasets is in visualization methods to make sense of quantitative information seeking data including new metaphors for envisioning and presenting results or altogether new models now more possible due to advances in multi-dimensional or real-time visualization.(Card, 1996; Card, Mackinlay, & Shneiderman, 1999) One application that may have merit is taken from ideas relating to faceted classification, where different views of the same data may be viewed according to context, confirmation of habits, prediction, content identification and classification or by individual or aggregate use. Also, at a simpler level, since much information behavior is repetitive and iterative, it may be possible to visualize both individual and aggregate information seeking sessions as a spiral model, where the user slowly gets to a centered "target" or solution to the information seeking problem at hand. Additionally, it would be useful to address and formalize a set of standards for data collection, analysis and presentation to make it possible to compare studies of Web information seeking. The current variety of presentation and graphing formats, not to mention the variability of data collection methods and datasets makes any comparison and consensus about Web information seeking problematic.

Recently, there has been an increased focus on modeling information retrieval activity, more specifically Web searching. By focusing first on Web searching, some of the above questions may be answered in an approachable manner and then be built upon for more complex, open-ended information seeking activities. These Web search studies show that users do have particular patterns of search use and yield insight into user mental models of search.(Jansen, Spink, Bateman, & Saracevic, 1998; Silverstein, Henzinger, Marais, & Moricz, 1998) It is possible that a system that can identify and track an individual user's search behavior, in comparison to aggregate patterns, and begin to build a profile of interests and habits can be used to personalize search results and interfaces to augment information seeking sessions. (Pitkow et al., 2002) This type of observation leads to an approach that has potential to both confirm or leverage current information seeking models or help to unveil altogether new models, that may be either general (e.g. an information seeking episode) or

contextual (e.g. eCommerce or content topic specific) to a particular user, a particular situation or as generally applicable to any information seeker.

Discussion Questions

1.The name(s) of the theory

Hybrid Information Seeking Model

2.Authoritative primary (and secondary if significant) publication sources

Choo, C. W., Detlor, B., & Turnbull, D. (1999). *Information Seeking on the Web - An Integrated Model of Browsing and Searching.* Paper presented at the Proceedings of the 62nd Annual Meeting of the American Society of Information Science, Washington, D.C.

Also see References at end of this document

3.Who developed it and from which field

Originally based on Choo's analysis of Aguilar and Ellis' models of environment scanning and information retrieval interfaces (respectively). Further refined in a series of studies by Choo, Detlor and Turnbull 1998-2000.

4.How/why the theory was developed

This theory is an attempt to characterize the information seeking behavior of Web users. It's goals are to determine critical Web use modes of environmental scanning combined with Web interaction via browser functionality to build towards a more complex understanding of the types of behaviors and their asssociated activities for Web information seeking. Turnbull seeks to advance these ideas into forming profiles of Web information seeking patterns of use that may be predicted and serve as a user-centered basis for augmenting Web information seeking by means of agents or other automated support systems.

5.The theory's premise and propositions

This hybrid model is a first step into developing a complex, multi-faceted view of the Web information seeking process. This includes what this type of behavior shares in common with traditional information seeking studies and models as well as what more recent models of Web access and use can be applied.

6.How the theory has been used in past work and in which fields, i.e., track record to-date, especially in IB or information science

In comparison to the history of IB studies, this area is very new. As such, part of the goals of this workshop is to explore the existing scholarship for IB and information seeking studies to find probable appropriate models to apply to Web use or possibly to combine with this current hybrid model.

7.How the theory is relevant to IB, i.e., topics and concepts it might be used to study, gaps it might fi ll, aid in translating IB findings to system design, etc.

See abstract.

8.The theory's accompanying methodological recommendations (if any; e.g., Dervin's sense-making suggests the micro-moment timeline approach)

Many traditional methods may be appropriate for this avenue of study including:

- Transaction Log Analysis
- Critical Incident Theory
- Bibliometrics
- Contextual Inquiry

It seems appropriate to use a combination of quantitative and qualitative methods to collect and then confirm (respectively) results from such a new area of study. The bulk of the ideas offered in this proposal do rely on more quantitative means which may be appropriate to locate areas of focus for more intensive, qualitative inquiriy using the methods noted.

9.Which other theories the discussion theory might connect

There are currently many areas that coincide with the hybrid information seeking model. Included are network traffic analysis, proxy algorithm determination, web workload characterization, Web browser interface design, value-added processes and informetrics to name a few.

10.Summary of theory's overall strengths for studying IB problems

This theory's main strengths may prove to be:

- Grounded focus on empirical data collected from Web users in the field
- Comparable, quantitative measures that can be applied over different user populations
- Sophisticated data collection instruments relying on recent advances in technology to collect user information seeking activity.

11.Summary of theory's overall weaknesses for studying IB problems

This theory's weaknesses may prove to be:

- A reliance on using past information behavior models information seeking models to describe a set of activities that are substantially different in scope and purpose.
- The ability to actually determine a user's mental model or intended use of a Web information resource may be reduced with large-scale data collection that inhibits contextual observation and inquiry.
- As Web search technologies advance, a shift towards information retrieval models may mean that traditional information seeking activities will be performed less or transformed into a less similar activity.

12.Recommendations for future use

See abstract.

References

Belkin, N. J., N.Oddy, R., & Brooks, H. M. (1982). ASK for Information Retrieval: Part I. Background and history. *Journal of Documentation, 38*(2), 61-71.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part II. Results of a design study. *Journal of Documentation, 38*(3), 145-164.

Card, S. K. (1996). Visualizing Retrieved Information: A Survey. *IEEE Computer Graphics and Applications, 16*(2).

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization : using vision to think*. San Francisco, Calif.: Morgan Kaufmann Publishers.

Choo, C. W., Detlor, B., & Turnbull, D. (1998). *A Behavioral Model of Information Seeking on the Web - Preliminary Results of a Study of How Managers and IT Specialists Use the Web*. Paper presented at the Proceedings of the 61st Annual Meeting of the American Society of Information Science, Pittsburgh, PA.

Choo, C. W., Detlor, B., & Turnbull, D. (1999). *Information Seeking on the Web - An Integrated Model of Browsing and Searching.* Paper presented at the Proceedings of the 62nd Annual Meeting of the American Society of Information Science, Washington, D.C.

Choo, C. W., Detlor, B., & Turnbull, D. (2000a). *Web Work: Information Seeking and Knowledge Work on the World Wide Web*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Choo, C. W., Detlor, B., & Turnbull, D. (2000b, Jan 4-7). *Working the Web: An Empirical Model of Web Use.* Paper presented at the 33rd Hawaii International Conference on System Science (HICSS), Maui, HI.

Huberman, B. A., Pirolli, P. L. T., Pitkow, J. E., & Lukose, R. (1998, April 3, 1998). Strong Regularities in World-Wide Web Surfing. *SCIENCE, 280,* 95-97.

Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. (1998, Spring, 1998). Real Life Information Retrieval: A Study of User Queries on the Web. *SIGIR Forum: A Publication of the Special Interest Group on Information Retrieval, 32,* 5-18.

Kuhlthau, C. C. (1991). Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society of Information Science, 42*(5), 361-371.

Pirolli, P., & Card, S. (1995). *Information Foraging in Information Access Environments.* Paper presented at the Conference on Human Factors in Computer Systems, CHI-95, Denver, Colorado, USA.

Pitkow, J., Schutze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., et al. (2002). Personalized Search: A Contextual Computing Approach May Prove a Breakthrough in Personalized Search Efficiency. *Communications of the ACM, 45*(9), 50-55.

Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1998). Analysis of a Very Large Web Search Engine Query Log. *SIGIR Forum: A Publication of the Special Interest Group on Information Retrieval,* 6-12.