

# SeekSeer: A Tool for Distributed Ethnomethodological Study of Information Retrieval

Andy Carle

Computer Science Division & Berkeley Institute of Design  
University of California – Berkeley  
Berkeley, CA 94720-1764  
1-510-642-1268

acarle@cs.berkeley.edu

## ABSTRACT

In this paper we describe the design of SeekSeer, a new software tool for gathering qualitative data on how people look for information. We posit that a general misunderstanding of how users look for information is the core issue that plagues the current design methodology for information retrieval systems and interfaces and diminishes chances for widespread acceptance of new products. Past attempts at remedying this issue are examined and critiqued. The lessons learned from these methods and studies are then used to inform the design of SeekSeer, in which descriptions of information retrieval activities are gathered at the user's pace and comfort level through wizard-based interviews. The software and additional information are available online at <http://www.seekseer.net>

## Keywords

Information retrieval, information seeking, search, ethnography, ethnomethodology, user experience

## 1. INTRODUCTION

Over the past fifteen years our society has become increasingly information driven. To make the management of this information more tractable many individuals have shifted significant portions of their information to digital mediums [1]. Whether stored locally on a personal computer or remotely in the cloud of the World Wide Web, much of this information is located via information retrieval (IR) and search tools [2, 3]. These tools influence the way people find information to such an extreme that if an item can not be found using them it is not likely to be found at all. This dependence places a great burden on the interaction between the user and the IR tool as, ultimately, the gatekeeper to this wealth of knowledge is the tool's user interface.

Unfortunately, the evolution of search tools from novelty to primary mediator of information has not been matched by a corresponding change in interface design techniques. Driven by a fear of catching loyal users by surprise, the single textbox and list of ranked results of the search interface has remained virtually unaltered since the mid-90s. Since then, much has changed about the way researchers think about interface design. Designers in the field of human-centered computing/design have produced

numerous examples of radically redesigned IR interfaces. However, not one of these has had significant commercial impact: Google<sup>1</sup> search today looks much like Yahoo!<sup>2</sup> search five years ago looks much like AltaVista search ten years ago<sup>3</sup>.

Reasons for this stagnation vary from instance to instance. However, we speculate that an underlying cause in many cases is a lack of thorough understanding of how and why users go about finding information. This deficiency affects the design and implementation phases of a project in two major ways. First, it is difficult to accommodate a user's needs without a very robust understanding of them. Second, it is a risky move to replace a successful interface with a new one in the absence of concrete evidence that there is something wrong with the existing solution. As multi-billion-dollar companies like Google are not in a position to take major risks with their flagship products, we are unlikely to see significant changes in user experience until a richer description of the average IR user is established

## 2. PRIOR WORK

Several methods of user study have been applied to the problem of information retrieval. Present space requirements preclude a complete summary of efforts to document user interaction with IR tools and information systems. However, a discussion of the more prominent methodologies and the studies that used them is necessary to reveal why the body of work has, as a whole, been uninformative.

One approach to studying information systems is to carefully categorize the activities of a user so that the data can be used with quantitative methods. The fundamental belief behind this approach is that an understanding of *what* users are doing (and how frequently) is enough information to tailor systems and software to the majority. Influential studies on user browsing habits were first conducted in the mid-1990s [4-6]. More recently, Byrne, John, and Crow proposed and implemented a taskonomy in which all user behaviors while browsing the web are coded. Unfortunately, none of these studies explicitly addresses information seeking.

This paper was prepared for SIMS 290-2: Search Engines: Technology, Society, and Business taught by Professor Marti Hearst. Permissions to make digital/hard copies of all or part of this material for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage and that title and authorship information is included. Available online at [www.seekseer.net](http://www.seekseer.net)

<sup>1</sup> <http://www.google.com>

<sup>2</sup> <http://search.yahoo.com>

<sup>3</sup> <http://web.archive.org/web/19961023234817/http://www.altavista.com/>

Many researchers believe that the categorizing and coding style of user study is too simplistic. By sacrificing the relative simplicity of quantitative measures in favor of qualitative evaluations a richer picture of a user's experience while working with an information system can be developed. The most commonly used technique for qualitative user study is ethnography. Nearly all ethnographic methods revolve around informal interviews with participants and observation of participant behavior. However, even within the relatively narrow field of information retrieval, ethnographic methodology varies considerably from study to study. Some, such as Solomon's elaborate three year study [7-9], follow a traditional ethnographic roadmap with the goal of thoroughly understanding the social context in which IR activities take place. Others advocate an ethnomethodologically informed view of ethnography that focuses on how users go about performing tasks within the social context – a view in which participant reasoning and methods become the locus of attention rather than the participants themselves [10]. Still others advocate a “cyber-ethnographic” approach in which the lines between virtual spaces and physical spaces is blurred when considering both the social aspects and methodology of tasks completed online [11].

O'Day and Jeffries's work of 1993 [12] is widely considered the seminal work of ethnography applied to information seeking. This paper first documented (via a traditional ethnographic method) an array of search modes and methods, including conditions under which information seekers move between interconnected searches. Unfortunately, this study, like others of its time [13], was based on an assumption that information retrieval tasks are mediated by an expert librarian – a scenario that is rarely accurate twelve years later.

More modern studies of information retrieval have critical weaknesses of their own. Many attempts at ethnomethodologically informed ethnographies rely on artificial, researcher-supplied tasks which have been shown to affect search performance [14]. Other studies blur the boundary between ethnography and quantitative methods by inspecting query logs [15] or query logs and user surveys [16]. Such studies attempt to reconstruct the user's state of mind and context from the information at hand, but usually either miss the context of and the reasoning behind the user's process entirely or limit their scope to keyword-based search engines, omitting a vast array of information seeking strategies. Even studies that get everything else right are of questionable value due to limited scope. Examples in this category include Holscher and Strube's work of 2000 [17] which was limited to specific domains and Teevan et al.'s study of directed search [18], the participant population of which was limited to 15 computer science graduate students in the authors' department.

### 3. OUTLINE OF A NEW APPROACH

The weaknesses of previous studies are significant enough to cast doubt on their conclusions. This doubt diminishes the authority of the field and limits the potential impact and influence of this body of work. We propose the following desirable attributes for a new approach to IR user study that is informed by the strengths and addresses the problems of prior works:

- Based on an ethnomethodologically-informed ethnographic technique
- Informal interviews guide the data-collection process
- Methodology is not bound to specific IR tools
- Highly scalable
- Very accommodating and easily distributed to a diverse body of participants
- Capable of long-term, longitudinal observation of users

We further posit that the most reasonable way to ensure high scalability, accommodation, and distribution of a study is to encapsulate the methodology in a highly mobile, replicable artifact such as a software application. The remainder of this paper will examine one model application: SeekSeer.

## 4. DESIGNING SEEKSEER

The SeekSeer application was created using a standard iterative design process. First a “best guess” at an appropriate interview script was compiled as described below. This script was then encoded in a low-fidelity mockup<sup>4</sup>. Continuous user feedback on the mockup led to successive versions of the interview script and mockup. The end result of this process was then encoded in a full-fledged prototype as described in section 4.2.

### 4.1 Creating the Interview Script

The initial design of the interview script was based on a combination of the most compelling aspects of several previous ethnographies. The basic questions relating to how, why, and with what tools the user is proceeding in their IR task come from O'Day and Jeffries [12]. O'Day and Jeffries also stress the importance of understanding how and why the user decides that they have reached an acceptable resolution for their self-defined task, inspiring the final question of the SeekSeer interview.

Several questions were derived from Spink's interesting proposal of a mental model of information seeking problems and tracking changes between stages of this model during IR tasks [19]. Teevan's directed search study urged the use of the phrase “look for” rather than “search” to avoid biasing responses and inspired questions about information type and orienteering behavior [18].

Most of the question content remained stable through all iterations of the interview design. However, user feedback resulted in the compression of many related questions into single, larger questions to shorten the overall length of the interview.

### 4.2 Creating the Application

It was decided early in the design process that a simple, wizard-based system for conducting the interviews should be developed so that more time could be spent tuning the interview script itself. The SeekSeer client is implemented as a rudimentary Java Swing application<sup>5</sup> that conducts the interview at the user's pace. Upon completion of an interview, the user's responses are recorded

<sup>4</sup> The final iteration of the mockup (in .ppt format) is available at [www.seekseer.net](http://www.seekseer.net)

<sup>5</sup> The SeekSeer client application and source code are available at [www.seekseer.net](http://www.seekseer.net)

locally in a transparent XML format to allow for user inspection, helping to build user trust. At the user's request, these files are uploaded to the central SeekSeer repository for analysis. See Appendix A for an example user interview generated using the application. It is worth noting that the client and server applications were implemented in a very general manner, making it quite easy for future developers to implement their own scripts.

## 5. EVALUATION

The SeekSeer application is still in a very primitive prototype version. However, initial results are encouraging. Study participants with no knowledge of ethnographic methods or information retrieval theory have responded as desired to the interview system. Full interview responses have been in line with those given in similar face-to-face informal-interview based studies. This is indicative of a promising line of study that merits further investigation.

## 6. FUTURE WORK

### 6.1 Data Analysis

At this stage of development very little analysis of the data gathered by SeekSeer has taken place. Many projects have shown how to manage small sets of qualitative data, but few have dealt with the scale of responses that we hope to one day possess. Crabtree et al.'s discussion of making use of the results of ethnomethodologically informed ethnography in the design process [10] will be useful in this stage of the project.

### 6.2 Application Design

The SeekSeer client is not yet viable for widespread release. Now that the interview script is relatively stable significant user interface improvements are possible for the client. Such improvements may include making the question interface more aesthetically pleasing and making the client window easier to hide in the system tray. But this is just speculation; an iterative design process will be the best way to determine the necessary changes.

### 6.3 Study Design

Many questions have been left unanswered by our initial investigation of a distributed ethnographic tool. For example, a major barrier to widespread release is user incentive. Designing an incentive system to gather participants will be an interesting challenge. An additional venue for extension is augmenting user interviews with automatically captured data, such as web browser logs. Having this information would afford more quantitative methods of analysis than we can currently accommodate.

## 7. CONCLUSIONS

In this paper we have presented a novel mechanism for conducting ethnomethodologically informed ethnographic studies. Our approach addresses reoccurring struggles with scope, participant diversity, and breadth of coverage experienced in prior studies while incorporating many of their successful attributes. We introduced a simple software application that meets the basic goals of our approach and has produced encouraging results at a prototype stage of development. We believe that this type of study, while facing a number of obstacles to widespread adoption, is fundamentally feasible and could lead to exciting results.

## 8. ACKNOWLEDGMENTS

We are grateful to all of our SeekSeer prototype study participants for their time and the feedback that they provided. Special thanks to Will Dimmit for insightful discussions related to the project.

## 9. REFERENCES

- [1] N. Negroponte, *Being Digital*, New York: Alfred A. Knopf, Inc., 1995.
- [2] S. Lawrence and C.L. Giles, "Accessibility of information on the Web," *Intelligence*, vol. 11, pp. 32-39, 2000.
- [3] B. Cole, "Search Engines Tackle the Desktop," *Computer*, vol. 38, pp. 14-17, 2005.
- [4] L.D. Catledge and J.E. Pitkow, "Characterizing browsing strategies in the World-Wide Web," in *Proceedings of the Third International World-Wide Web conference on Technology, tools and applications*, 1995, pp. 1065-1073.
- [5] A. Cockburn and S. Jones, "Which way now? Analysing and easing inadequacies in WWW navigation," *Int.J.Hum.-Comput.Stud.*, vol. 45, pp. 105-129, 1996.
- [6] L. Tauscher and S. Greenberg, "How people revisit web pages: empirical findings and implications for the design of history systems," *Int.J.Hum.-Comput.Stud.*, vol. 47, pp. 97-137, 1997.
- [7] P. Solomon, "Discovering Information Behavior in Sense Making: Time and Timing," *Journal of the American Society for Information Science*, vol. 48, pp. 1097-1108, 1997.
- [8] P. Solomon, "Discovering Information Behavior in Sense Making: The Social," *Journal of the American Society for Information Science*, vol. 48, pp. 1109-1126, 1997.
- [9] P. Solomon, "Discovering Information Behavior in Sense Making: The Person," *Journal of the American Society for Information Science*, vol. 48, pp. 1127-1138, 1997.

- [10] A. Crabtree, D.M. Nichols, J. O'Brien, M. Rouncefield and M.B. Twidale, "Ethnomethodologically informed ethnography and information system design," *Journal of the American Society for Information Science*, vol. 51, pp. 666-682, 18 May 2000. 2000.
- [11] K.J. Ward, "Cyber-ethnography and the emergence of the virtually new community," *Journal of Information Technology*, vol. 14, pp. 95-105, 1999.
- [12] V.L. O'Day and R. Jeffries, "Orienteering in an Information Landscape: How Information Seekers Get From Here to There," 1993, pp. 438.
- [13] N.J. Belkin, C. Cool, A. Stein and U. Thiel, "Case, Scripts, and Information-Seeking Strategies: On the Design of Interactive Information Retrieval Systems," in *Expert Systems with Applications*, vol. 9, 1995, pp. 379-395.
- [14] K. Kim and B. Allen, "Cognitive and task influences on Web searching behavior," *J.Am.Soc.Inf.Sci.Technol.*, vol. 53, pp. 109-119, 2002.
- [15] A. Spink, D. Wolfram, M.B.J. Jansen and T. Saracevic, "Searching the Web: the public and their queries," *J.Am.Soc.Inf.Sci.Technol.*, vol. 52, pp. 226-234, 2001.
- [16] A. Broder, "A taxonomy of web search," *SIGIR Forum*, vol. 36, pp. 3-10, 2002.
- [17] C. Holscher and G. Strube, "Web search behavior of Internet experts and newbies," in Proceedings of the 9th international World Wide Web conference on Computer networks : the international journal of computer and telecommunications netowrking, 2000, pp. 337-346.
- [18] J. Teevan, C. Alvarado, M.S. Ackerman and D.R. Karger, "The Perfect Search Engine is Not Enough: A Study of Orienteering Behavior in Directed Search," 2004, pp. 415.
- [19] A. Spink, "A user-centered approach to evaluating human interaction with Web search engines: an exploratory study," *Information Processing and Management*, vol. 38, pp. 401-426, 2002.

# Appendix A – An Interview Submitted Using SeekSeer

<Session>

<time>Wed Dec 7 05:20:54 CST 2005</time>

<LookingFor>I am looking for the average weight of a bobcat in Northern Missouri. </LookingFor>

<StartingStage>

I am identifying specific information sources that I think will be of use to me

</StartingStage>

<WhyLook>

I am looking for the average weight of a bobcat in Northern Missouri because I very recently hit a bobcat while driving near Moberly, Missouri. The state patrol officer said it did more damage to my car than he would have expected. I am curious how large the average bobcat is.

</WhyLook>

<HowStart>

I am most likely going to begin my search by using the search engine at [www.google.com](http://www.google.com). I will type in the phrase "average bobcat weight".

</HowStart>

<Steps>

<WhyChange>

I was receiving information about the average weight of bobcats in other specific states, but not Missouri.

</WhyChange>

<HowProceed>I changed my search to the broader phrase "missouri bobcat".</HowProceed>

</Steps>

<DefinitionChange>

My interest in the topic of weight decreased when I found information about the bobcat's predation style. I found a page that was created by the Missouri Department of Conservation, and it gave me information which is proving to be more applicable to my situation.

</DefinitionChange>

<EndingStage>

I am in the process of finishing the collection of information for this stage of my work

</EndingStage>

<Results>

I found the answer to my question and even more relevant information. Yes, there was additional information provided on the subject. I did not pursue it because I am very tired, and I found the specific information which I needed. I believe this information is trustworthy because it is from a state-created website.

</Results>

</Session>