

Wrapper: An Application for Evaluating Exploratory Searching Outside of the Lab

Bernard J. Jansen

College of Information Sciences and Department of Computer Science and
Technology
The Pennsylvania State University
University Park PA 16802
jjansen@ist.psu.edu

Raghavan Ramadoss

College of Engineering
The Pennsylvania State University
University Park PA 16802
ramadoss@cse.psu.edu

Mimi Zhang and Nan Zang

College of Information Sciences and
Technology
The Pennsylvania State University
University Park PA 16802
mzhang@ist.psu.edu,
nzz101@psu.edu

ABSTRACT

In this position paper, we assert that a focus on evaluating individual exploratory searching systems misses a critical aspect of assessing the exploratory searching process. Namely, that in complex information environments, searchers use multiple systems over an extended period marked by specific episodes of interaction with online systems. We argue that the focus of the evaluation should be on the process, not a single system. However, evaluating an exploratory searching process can be a difficult task to conduct in a naturalistic setting (i.e., outside of a laboratory). In response, we have developed a client-server application for use in the study and evaluation of exploratory searching processes. We describe the application and demonstrate the ability of the application in a pilot study. The results from our evaluation show that exploratory searching is indeed a chaotic process, demonstrated by the use of multiple information systems and repeated episodes of searching. The implications are that by using this tool one can successfully evaluate exploratory searching processes. Assessment of the entire process rather than a single exploratory searching system could significantly further the advancement of system design for this critical searching context.

Categories and Subject Descriptors:

H.3.3 [1] Information Search and Retrieval – relevance feedback.

General Terms:

Performance, Design, Experimentation, Human Factors

Additional Key Words and Phrases:

Implicit user feedback, exploratory search evaluation

1. INTRODUCTION

In exploratory search, the situational context in which the user performs individual searching episodes is critically important in the evaluation of the overall process. The user's searching episode (i.e., a distinct period of interaction with an online system) may involve multiple queries related only at some high-level of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGR 2006 Conference, August 6 - 11, Seattle, Washington, USA
Copyright 2006 ACM 1-58113-000-0/00/0004...\$5.00.

information abstraction. There may be several searching episodes in close temporal proximity, or a considerable temporal span may separate the searching episodes. Additionally, these searching episodes may occur on multiple searching systems.

In such a complex situational environment, the evaluation of a single Exploratory Search System (ESS) could miss crucial elements of the user context, since the searching process may not occur on one ESS. In fact, our view is that once you have confined exploratory search to a single system, you have over simplified the problem. Our position is that the evaluation should not focus on a particular ESS but on the Exploratory Searching Process (ESP), which can span multiple searching episodes, multiple systems, and varied temporal spaces. However, evaluating ESPs has been nearly impossible or at least too costly in terms of effort due in part to the lack of automated methods of collecting ESP data.

In this paper, we present the Wrapper, a client-server application for the use in evaluating ESPs. The Wrapper is based on technology we developed for user evaluations on information retrieval (IR) systems [4]. We describe the Wrapper's design and value in terms of evaluating ESPs, and we show its value by discussing the results of a pilot study where we employed the Wrapper. With the Wrapper or similar client-server applications, one can conduct naturalistic studies of the entire ESP and not be limited to the study of a single ESS. We believe that such an approach provides much more realistic insight into the users' tasks, goals and behaviors.

Sections 2 present a brief literature review and the research objectives. We then discuss the structure of the proposed application, and research results to date in sections 3 and 4 respectively. Section 5 provides the concluding remarks, along with future aims and implications of the Wrapper for ESP evaluation.

2. LITERATURE REVIEW

Although there are open questions, the evaluation of a single ESS (or any single searching system) is relatively straightforward compared to evaluating ESPs. One can point to the series of Text REtrieval Conferences as an example. There are also good commercial applications for single system evaluation in labs, such as
Morae 1.1
(<http://www.techsmith.com/products/morae/default.asp>).

However, the evaluation of an ESP is much more difficult because the central actor is not the system but the user. During an ESP, the user may access multiple systems. The search topic is difficult, and the period of searching is longer. All of these factors point to the need for naturalistic (i.e., outside of a laboratory) studies of

the entire process rather than a system. However, there has not previously been an application to facilitate data collection and delivery available to the research community.

Researchers have relied on a variety of methods for data collection. Hancock-Beaulieu, Robertson, and Nielsen [2] used server-side transaction logs and online questionnaires. In their naturalistic and longitudinal study of professionals and their information seeking patterns, Choo, Betlor, and Turnbull [1] developed their own logging software but had to physically collection the logs. Kelly [5] used a spy software package and a proxy server. Spy software has inherent disadvantages including granularity of data capture, and privacy concerns. A proxy sever is limited to logging traffic only on one network. Toms, Freund, and Li [7] developed a system for conducting large-scale evaluations. However, the entire study must occur within the WiIRE framework and is limited to one server.

To address the need for an application to study ESPs, we developed the Wrapper, a client-side application to collect and gather user data. The application is coded in Visual Basic 6, is easy to install, collects a wide range of user-systems interactions, and is not limited to a single server.

In the following sections, we present a description of the features and output of the application. We aim to provide this version of the Wrapper to the research community via a uniform resource locator (URL). Interested researchers can download the application from the URL for use in their research projects and studies.

3. DEVELOPMENT METHODOLOGY

One of the essential aims in developing the Wrapper is to facilitate the collection and gathering of data over an extended period and over multiple information systems.

3.1 Structure of the Application

The software consists of (1) the client-side module and (2) the server-side module. The client and server modules of the Wrapper communicate using sockets. The client module sends a string message containing the user details to the server in the form of a plain text document. This text file contains the computer's Internet Protocol (IP), time-stamped, implicit feedback action, and object of that action [3]. Implicit feedback interactions, including copy, bookmark, print, save, etc., indicate relevance.

A daemon, capable of simultaneous reception of files from multiple clients, runs at the server end and waits for incoming files from the clients. The server dumps the received file locally for further analysis on the collected data.

3.2 The Client Wrapper Application

The client-side module is a self-installing executable that can be downloaded and installed over the Internet. The executable is generated from the Visual Basic programming environment. One

can activate the application manually, via a batch file, or from a browser toolbar. The application has a Window interface (Figure 1) for real time observation that can be hidden to allow for unobtrusive monitoring. The application logs interactions with the IR system, along with other applications, using Dynamic Data Exchange (DDE). Output is to a text file, with a specifiable location and an automatically generated unique filename. Additionally, the client module also sends this output directly to the server-side module for data collection.

Referring to Figure 1, we numbered each of the functional aspects of the application, which we describe below.

1. Log filename (generated automatically using date and time)
2. Running text of log file.
3. List of all processing running.
4. The current value of the clipboard.
5. Text to be appended to log file.
6. Current system time.
7. Title and URL of current page.
8. Running list of URLs.

The dialog box in Figure 1 can be set to hide during studies, so the participant will never see it. An example of the application output is:

```
20:58:21 Clipboard Use
20:58:21 https://mail.ist.psu.edu/exchange/ View URL
20:58:43 http://search.yahoo.com/search?ei=utf-8&fr=slv1-
&p=successive+searching View URL
20:59:49 http://search.yahoo.com/search?ei=UTF-
8&p=successive+searching View URL
21:00:07 http://search.yahoo.com/search?ei=UTF-
8&p=successive+searching View URL
21:00:21
http://edc.techleaders.org/LNT99/presentations/05_Thu/strat
egies.htm View URL
21:00:40 View URL
21:00:40 Bookmark URL
```

In its current version, the application logs a wide range of user interactions, include interactions with the browser tool bars, interactions with the system clipboard, scrolling of results listing or documents, and numerous implicit feedback actions [6], such as bookmark, copy, print, save, and scroll. The user activates the Wrapper prior to performing a search and when the Web browser closes the Wrapper automatically terminates. Figure 2 illustrates how the Wrapper integrates with the browser and computer operating system.

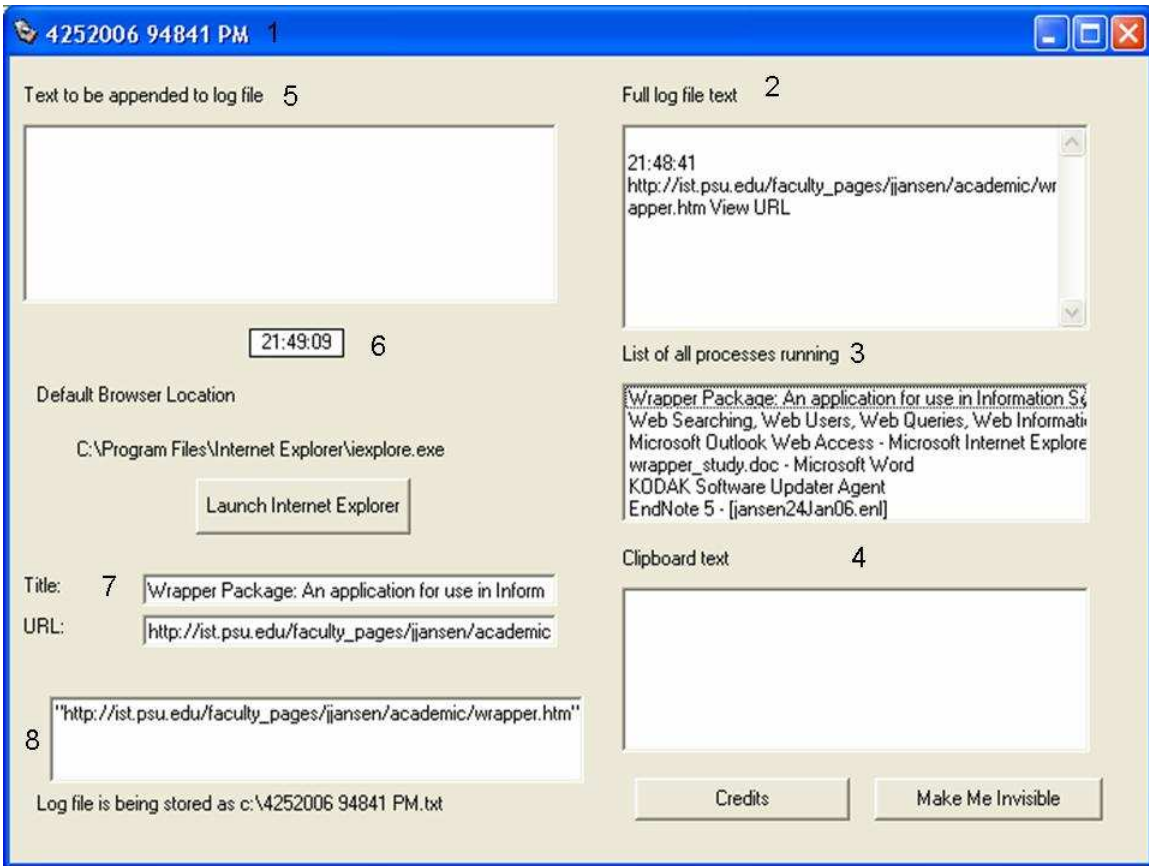


Figure 1. Visible Version of the Client-side Module of the Wrapper.

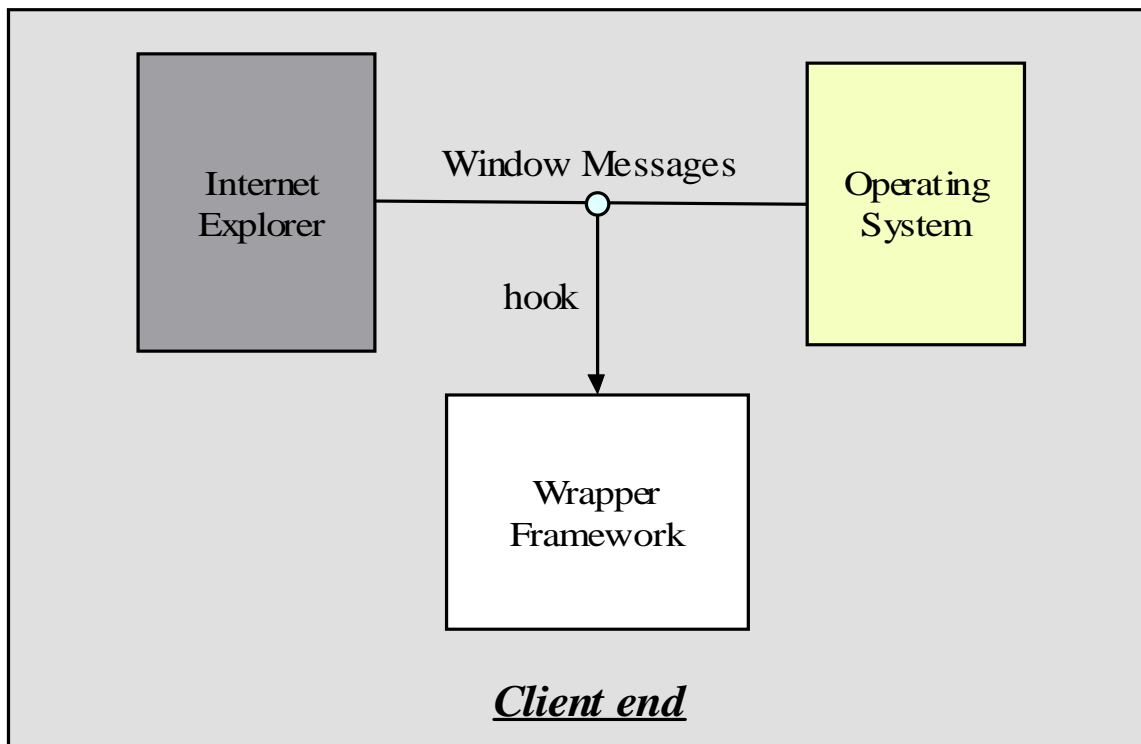


Figure 2. The Client-side Module Interfacing with the Browser and Search Engine.

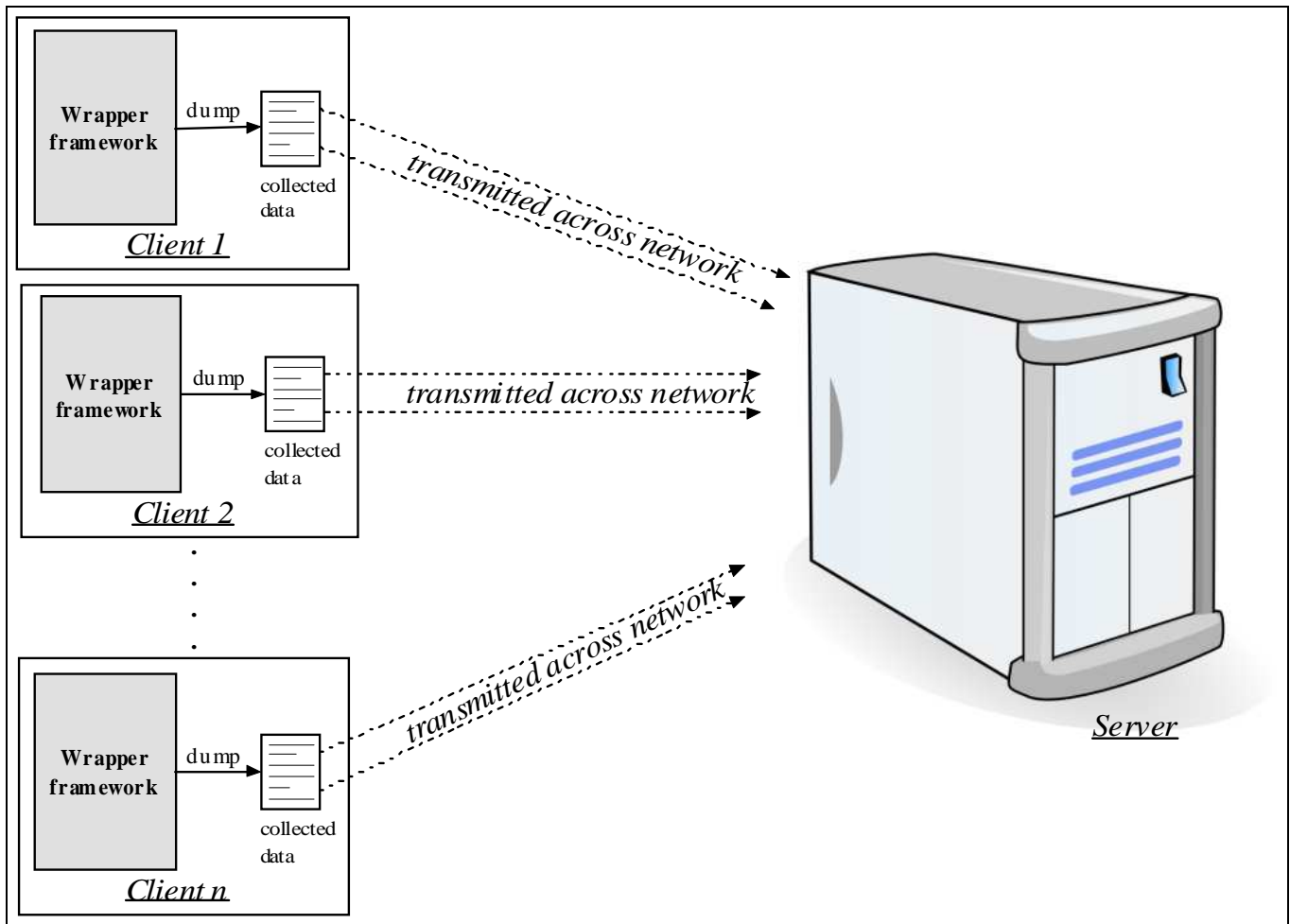


Figure 3. The Server-side Module Listening for Messages from the Client-side Module.

3.3 The Server Wrapper Application

The server-side module collects the information sent from one or more client-side applications, storing each information set into a file. The server-side module creates and names the file using the IP address of the computer on which the client-side module is running. The server-side module uses sockets to receive the messages from the client-side modules. The server-side application writes the received information values to the appropriate file at five-second intervals. These intervals can be adjusted with changing requirements by altering the server code. Figure 3 illustrates the communication among the server module and client modules.

4. WRAPPER EVALUATION

4.1 User Evaluation of Performance Tool

In order to gauge the effectiveness of the Wrapper, we conducted a pilot test of the application with 4 participants in order to determine the effectiveness of the software in light of users' differing searching experiences and searching needs. We conducted the study over a one-week period (i.e., 7 continuous 24-hour periods). The participants conducted their searching as they normally would, using whatever systems they deemed

appropriate. The only additionally needed action was to activate the Wrapper via a button on the browser toolbar at the start of each searching episode. The client-side module collected the data and sent it to the server-side module. The study did not tie the participants to any network, location, computer, or searching system.

4.2 Results from User Evaluation of Performance Tool

Aggregate results of the 7-day pilot study are presented in Table 1. As we can see from Table 1, searching is a disorganized task and does not conform to the logical sequence of events that one so often sees in the scenario approaches used in lab studies of IR systems. Instead, searchers employ a variety of searching and information sources and return to topics over multiple days. Users exhibited behaviors on certain topics that one would classify as being part of an ESP. In these ESP, users visited multiple search engines and searched on multiple Web sites. Some of these Web sites were directly off the search engine results listing. Others were browsed to from the search engine results or bookmarked Web sites.

Table 1. Aggregate Statistics Collected by the Wrapper During Pilot Study

User	Computers Used	Searching Episodes	Information Topics By Episode	Systems Used	Episode Duration
1	3	18	1. Medical, 2. Entertainment, 3. Parenting, 4. Topic Research A, 5. Topic Research B, 6. Technology, 7. Technology, 8. Entertainment, 9. Topic Research, 10. Topic Research A, 11. Topic Research A, 12. Religion, 13. Topic Research A, 14. Topic Research C, 15. Ecommerce (Housing), 16. Ecommerce (Housing), 17. Ecommerce (Housing), 18. Ecommerce (Housing)	5	min: 1 minute max: 31 minutes
2	2	4	1. Ecommerce, 2. Hobby, 3. Technology, 4. Ecommerce	12	min: 2 minute max: 147 minutes
3	1	4	1. Topic Research D, 2. Topic Research E, 3. Topic Research D and Sports, 4. Topic Research D	4	min: 1 minute max: 28 minutes
4	1	21	1. Ecommerce, 2. Technology, 3. Topic Research F, 4. Topic Research G, 5. Ecommerce, 6. Topic Research H, 7. Topic Research I, 8. ECommerce, 9. Entertainment, 10. Topic Research J, 11. Topic Research K 12. Topic Research L 13. Work Requirement 14. People Search 15. Topic Research M 16. Topic Research N 17. News, 18. Ecommerce, Technology, History, Art, ecommerce 19. Topic Research O, 20. Topic Research P, 21. Hobby	23	min: 2 minute max: 72 minutes

During ESPs, users also employed multiple queries with few query terms in common, but the queries were related at a higher information abstraction. However, searchers who engaged in ESPs conducted their searching over multiple days.

5. CONCLUSION

The Wrapper is an open source application for use during ESP studies. It addresses a fundamental issue in exploratory search evaluation in that user may seek information over an extended

period and on multiple information systems. The Wrapper collects and gathers, at a central location, the typical interactions of searchers from the client-side, thereby permitting studies of ESPs over extended durations and not limited to any one ESS. Therefore, the Wrapper directly supports the development of metrics to evaluate ESS performance and provides a focus on the searcher during evaluations. In future research, we aim to increase the number of user interactions the application logs.

6. REFERENCES

- [1] Choo, C. and Turnbull, D., Information Seeking on the Web: An Integrated Model of Browsing and Searching, *First Monday*, vol. 5, pp. http://firstmonday.org/issues/issue5_2/choo/index.html, 2000.
- [2] Hancock-Beaulieu, M., Robertson, S., and Nielsen, C., Evaluation of online catalogues: an assessment of methods (BL Research Paper 78), The British Library Research and Development Department, London 1990.
- [3] Jansen, B. J., Designing Automated Help Using Searcher System Dialogues, in *Proceedings of the 2003 IEEE International Conference on Systems, Man & Cybernetics*, 2003. Washington, D.C., USA. 5-8 October. pp. 10 - 16.
- [4] Jansen, B. J., Search log analysis: What is it; what's been done; how to do it, *Library and Information Science Research*, forthcoming.
- [5] Kelly, D., Understanding Implicit Feedback and Document Preference: A Naturalistic User Study, Rutgers, The State University of New Jersey, New Brunswick January 2004.
- [6] Oard, D. and Kim, J., Modeling Information Content Using Observable Behavior, in *Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology*, 2001. Washington, D.C., USA. 31 October - 4 November. pp. 38-45.
- [7] Toms, E. G., Freund, L., and Li, C., WiIRE: the Web interactive information retrieval experimentation system prototype, *Information Processing & Management*, vol. 40, pp. 655-675, 2004.