Information seeking: an overview of web tracking and the criteria for tracking software

The Authors

Ina Fourie, *Department of Information Science, University of Pretoria, South Africa*

Theo Bothma, *Department of Information Science, University of Pretoria, South Africa*

Abstract

Purpose – The purpose of this article is to alert researchers to software for web tracking of information seeking behaviour, and to offer a list of criteria that will make it easier to select software. A selection of research projects based on web tracking as well as the benefits and disadvantages of web tracking are also explored.

Design/methodology/approach – An overview of the literature, including clarification of key concepts, a brief overview of studies of web information seeking behaviour based on web tracking, identification of software used, as well as the strengths and short-comings noted for web tracking is used as a background to the identification of criteria for the selection of web tracking software.

Findings – Web tracking can offer very valuable information for the development of websites, portals, digital libraries, etc. It, however, needs to be supplemented by qualitative studies, and researchers need to ensure that the tracking software will collect the data required.

Research limitations/implications – The criteria is not applied to any software in particular.

Practical implications – The criteria can be used by researchers working on web usage and web information seeking behaviour to select suitable tracking software.

Originality/value – Although there are many reports on the use of web tracking (also reported in this article), nothing could be traced on criteria for the evaluation of web tracking software.

Article Type:
Literature review

Keyword(s):
Tracking; Computer software; Information retrieval; Worldwide web.

Introduction

The value of web information seeking studies in improving web information spaces, web-based training, our knowledge of the information seeking phenomenon, as well as our insight into the methodology used in information seeking studies has often been stressed. Knowledge gained can improve the design of user interfaces, search engines, digital libraries, menu items, navigational features, online help and intelligent software agents, information architecture, content description, metadata, and the teaching of information skills. The findings may be useful in web advertising, marketing and e-commerce, and can also be important to library managers, library and information professionals, knowledge managers, educators, software vendors, database designers, system designers and publishers. According to Hochheiser and Shneiderman (2001) it can also help us to collect data and gain an understanding of issues such as the depth versus the breadth of tree structures and incidental learning patterns. According to Nicholas *et al.* (1999c, p. 264) there are, however, also people who for various reasons do not want the data that can be collected since it might, for example, cramp their creativity and style, or lead to management problems.

Web information seeking behaviour is considered to be a subsection of the umbrella term information behaviour, which includes information seeking, information searching and information retrieval (Fourie, 2002; Wilson, 1999). According to Spink and Cole (2004, p. 657):

> Information behaviour is a broad term covering all aspects of information seeking, including passive or undetermined information behaviour. Information Seeking Behaviour is usually thought of as active or conscious information behaviour. Information searching behaviour describes the interactive elements between a user and an information system.

Pharo and Järvelin (2004, p. 633) refer to web information search processes. To deepen our understanding of information seeking behaviour specific data need to be collected, often through a variety of methods. In behaviour research quantitative research methods (e.g. web tracking) and qualitative research methods (e.g. think-aloud protocols, questionnaires and interviews) are distinguished. Wang (1999) gives an excellent overview of these methods in behaviour studies. Similar to many other authors (e.g. Fourie, 2002; Lazonder *et al.*, 2000; Pharo and Järvelin, 2004; Williams and Gunter, 2006) he argues for the combination of quantitative and qualitative methods to gain deeper and more meaningful data.

Web tracking, also known as web log analysis, web logging, and web log file analysis seems to be very popular in information seeking studies (Hsieh-Yee, 2001; Fourie, 2002; Jansen and Pooch, 2001; Pharo and Järvelin, 2004). Jansen and Pooch (2001), for example, did a literature review in which all but one of the studies were based on web tracking. Web tracking follows on from the earlier work on transaction log analysis (also known as TLA, log analysis, log file analysis, or log tracking) that has been used to monitor the use of databases, CD-ROM software and library catalogues (OPACs). The latest

term that can be found in the subject literature is deep log analysis (DLA) (Nicholas *et al.*, 2005; Nicholas *et al.*, 2006).

A 1993 edition of *Library Hi Tech* was dedicated to transaction log analysis and includes excellent articles on the history, rationale, strengths and shortcomings of transaction logs (e.g. Kaske, 1993; Kurth, 1993; Peters, 1993; Sandore, 1993; Sandore *et al.*, 1993), all of which argue for standardisation in transaction logs, and the need to popularise and improve the use of transaction logs. More recently Nicholas *et al.* (2006) report on deep log analysis (DLA) to monitor information seeking in very large digital libraries. It was therefore interesting to note that in his recent book on information seeking research Case (2002) refers only to qualitative methods, with no index entries for “transaction logs”, “log analysis”, “transactions” or “web tracking”. There also seems to be a growing number of studies relying mostly on qualitative methods (e.g. Palmquist, 2001).

The terms “web tracking” and “transaction logs” are used interchangeably in the web information seeking literature. This paper will use “web tracking” as the process and “transaction logs” as the result of web tracking. “Log analysis” will refer to the analysis of the content of the transaction logs. Software that has been used in reported information seeking studies include Camastia, FlashSite, GrandArt, Investigator, LittleBrother, Lotus ScreenCam, Netsnitch, SurfSpy, WebTrends and web browser logs, server logs and tailor made software such as WebTracker discussed by Choo *et al.* (2000a). An alternative to web tracking logs is interactive visualisations (e.g. Spotfire described by Hochheiser and Shneiderman, 2001) and data mining software. Although there are many recent articles on the use of web tracking (e.g. Nicholas *et al.*, 2004; Ozmutlu *et al.*, 2004; Nicholas *et al.*, 2006), nothing could be traced on the evaluation of web tracking software. Since it can play such an important role in supplementing qualitative methods to deepen our understanding of the web information seeking phenomena, it seemed timely to review the use of web tracking in information seeking studies to develop a criteria matrix. The purpose of this paper is to alert researchers to available software, and to offer a criteria matrix that will make it easier to select software. Often it seems easier to follow the same path as other researchers without considering the possibilities offered by software not yet explored in information seeking studies.

The paper will address the following:

1. clarification of key concepts;
2. brief overview of studies of web information seeking behaviour based on web tracking and identification of software used;
3. strengths and short-comings noted for web tracking;
4. explanation of the criteria; and
5. recommendations.

Clarification of key concepts

Web tracking builds on the earlier concept of transaction log analysis, which can be dated back to the mid 1960s according to Peters (1993). A number of definitions of transaction log analysis can be found in the subject literature. Nicholas *et al.* (2002, p. 65) explain log files as "… machine-generated records of user activity. The actual information collected by the logs depends partly on the software used and how the server was configured". Peters (1993) defines transaction log analysis as:

> The study of electronic recorded interactions between on-line information retrieval systems and the persons who search for the information found in those systems.

According to Blecic *et al.* (1998) (as quoted by Griffiths *et al.*, 2002) transaction logs should include system responses to user input to qualify as true transaction logging systems. Sandore (1993, p. 87) explains:

> Transaction logs supply unequivocal information about what a user typed while searching.

According to Davis (2004, p. 327) transaction log analysis is a "non-intrusive method for collecting data from a large number of individuals for the purpose of understanding online-user behaviour".

Based on these definitions we take web tracking to be machine-generated records of users' activities when interacting with a web space such as a web page, website, database, library catalogue, discussion group, search engine, intranet, electronic book, electronic journal, electronic newspaper or portal. Activities may include information searching (e.g. the use of search terms and the use of Boolean operators), information seeking (e.g. the selection of information sources to search, selection of links to follow), the capturing of information (e.g. printing and downloading), and the beginning and ending of search sessions. The actual information collected by the logs depends partly on the software used and how the server was configured. Such information may include the length of search sessions, referral sources (URLs), search queries, rate that web pages were visited, how users visited old and new web pages, distance in terms of URLs between repeated web page visits, frequency of web page visits, extent of browsing in one cluster of web pages, repeated sequences of "path-following behaviour", total number of hits, the most frequently requested documents, total number of unique addresses making requests, peak usage, number of user sessions and the system's responses. User activity is recorded on a real-time continuous basis. The before-mentioned builds on an earlier definition by Van der Berg and Fourie (2003).

When selecting web tracking software one should also get clarity on the meaning of the concepts that are monitored and measured, that is the metrics (e.g. search sessions, referral sources, and significant episodes of information seeking). Different interpretations are offered in the subject literature and complaints are often heard about such concepts not being clearly explained, leaving a reader in doubt as to how they were interpreted and making it

impossible to replicate a research project. Kurth (1993, p. 101), for example, declares:

> Researchers have not sufficiently described the data elements and measurement tools upon which their transactions log analyses have depended.

In his discussion of the shortcomings of web logs Dowling (2001) also explains that everything in a log file is fuzzier than it looks: what really is a hit, and a view or user count? Sandore *et al.* (1993) also argues for the standardisation of the definitions of common terms associated with transaction log analysis.

It is important to clarify the meaning of metrics before selecting software that can collect appropriate data. Although important, this paper will, however, not attempt such an analysis.

Brief overview of the use of transaction logs leading up to web tracking

In the context of libraries and information services (LIS) transaction log analysis has been used over many years to collect information on how information systems such as online library catalogues (OPACs), online and CD-ROM databases and web-based products and systems are used. The first reports on such transaction logs date back to the 1980s studies on OPACs (Nicholas *et al.*, 1999a, p. 265). Other examples include transaction logs of serials and their use (Sullenger, 1997), material availability to determine the success of information seekers in obtaining the material (Ciliberti *et al.*, 1998), search behaviour and the implications for bibliographic instruction and systems design (Wallace, 1993), searcher successes and failures (Hunter, 1991), and search failures on an OPAC (Peters, 1989). Zink (1991) reports on the use of transaction log analysis to monitor user search success, Wyly (1996) reports on access point value, Blecic *et al.* (1998) reports on the analysis of transaction logs to improve OPAC retrieval results, and Peters (1996) on its use in library management.

These studies are important when noting the strengths, weaknesses and potential of transaction logs.

Web tracking and studies on web information seeking behaviour

One method of studying web information seeking behaviour is to monitor transaction logs. Nicholas *et al.* (1999a, p. 263) declare:

> With the web being such a universally popular medium, accounting for ever more of people's information seeking behaviour, and with every move a person makes on the web being routinely monitored,

> web logs offer a treasure trove of data. This data is breathtaking in its sheer volume, detail and potential.

There has been an increase in the literature on web information seeking behaviour reporting the use of web tracking on its own, or in combination with other methods. The following is a brief overview of a selection of the literature that may shed light on the scope and variety of web information seeking studies, as well as the tracking software used.

One of the earlier studies often cited is Catledge and Pitkow (1995). Studies on search moves include the work of Whitmire (2004) who used ScreenCam to collect data on the search moves of undergraduate students and their decision-making. The latter was recorded through verbal protocol analysis. The purpose of the project was to determine the relationship between undergraduates' epistemological beliefs, reflective judgment, and their information-seeking behaviour.

Amanda Spink has been extensively involved with web information seeking research projects. A number of these projects were done with transaction logs from the Excite search engine, as well as a few other large search engines. The target group was the public at large – also referred to as “real life users with real needs”. Studies on the search logs of Excite are reported by Spink and Xu (2000), Jansen *et al.* (2000b) and Spink *et al.* (2001). Ozmutlu *et al.* (2003a) report on a study on multimedia web searching trends (1997-2001), multitasking information seeking and searching processes (Spink *et al.*, 2002), and an exploratory study of web searching for sexual information (Spink *et al.*, 2004). Ozmutlu *et al.* (2003b) analysed logs from Excite and AlltheWeb to re-examine the length of users' queries and their search sessions, while Ozmutlu *et al.* (2004) report on studies of the transaction logs of Excite and Fast Web. They found that there are fluctuations in web user behaviour over the day. User investigations of query results are much longer and submission of queries and the number of users are much higher in the mornings and some query characteristics, including terms per query and query reformulation, remain steady through the day. Jansen (2000) reports on the effect of query complexity on web searching results. He analysed the transaction logs of a major web search service, and then submitted 15 queries to five other major search engines. Goodrum and Spink (2001) report on image searching on Excite, Ross and Wolfram (2000) on an analysis of term pair topics submitted to Excite, Jansen *et al.* (2000a) report on searching for multimedia (audio, video and image queries), and Wolfram *et al.* (2001) on public searching of the web. More recently Huntington *et al.* (2004) report research on digital visibility and its impact on online usage, while Nicholas *et al.* (2006) report on the early findings of an exploratory deep log analysis of journal usage on OhioLink. Huntington and Nicholas (2006) report on improving the relevance of web menus. They classified search expressions on diabetes for the BBC website. Jamali *et al.* (2005) report on the use and users of scholarly e-journals.

Nachmias and Gilad (2002) used SurfSpy to track the time, web address, and searching/browsing string of each search step in order to determine the

search success of a group of graduate students. Thelwall (2001) explores the usefulness of web log file analysis as a source of information on visitor site use and navigation behaviour. Wang *et al.* (2000) did an exploratory study on users' interaction with the web. They used a combination of methods for data collection. The log file was created with NetSnitch. Cothey (2002) completed a longitudinal study of transaction logs in a “real world” study of web users' information searching behaviour. The daily individual history files that are automatically generated by the Netscape Navigator browser were collected. Each individual user had a unique cumulative history file that was maintained on a central server.

Bilal (2000, 2001) used ScreenCam, and a measure developed by her, namely web traversal, to study children's behaviour in using Yahooligans! She used a combination of quantitative and qualitative methods. Bilal and Kirby (2002) build on these projects to study differences and similarities in information seeking between children and adults. Palmquist (2001) gives an overview of usability for the study of users' web-based information retrieval behaviour. Usability testing techniques were part of the curriculum of Library and Information Science (LIS) students. She used Lotus' ScreenCam, TechSmith's Camastia and WinWhatWhere's Investigator to collect navigation data. The search sessions were recorded with a GrandArt video converter. A microphone was connected to the converter and the users were allowed to talk aloud. Pu *et al.* (2002) did a subject categorisation of query terms to explore web users' search interests. Data was collected from the search logs of three representative web search engines in Taiwan.

Ford *et al.* (2003) used a Java Script front-end to record all search data for further analysis to find a correlation between web search strategies and approaches to studying. It also logged “help access”. Ke *et al.* (2002) explored the behaviour of e-journal users in science and technology by monitoring the transaction logs of Elsevier's Science Direct OnSite in Taiwan. Although they used statistical methods for transaction log analysis, they conclude that they will in future rather use data-mining techniques (e.g. the data mining package DBMiner 2.0). Van der Berg (2000) and Van der Berg and Fourie (2003) report on a study of the Nexus database transaction logs, while Li (1999) deals with library web usage.

Kim and Allen (2002) investigated cognitive and task influences on web searching behaviour. They used a combination of qualitative and quantitative methods. In one of their experiments, a commercially available software program, LittleBrother, was used to record all the URLs of web pages visited and the duration of each visit. In a second experiment participants' comments were also recorded using a microphone connected to a videocassette recorder.

Moukdad and Large (2001) report on users' perceptions of the web as revealed by transaction log analysis. A random number of queries was captured from WebCrawler and loaded into a Microsoft Access database for further analysis. Large *et al.* (2002a) studied gender differences in collaborative web searching behaviour in an elementary school. They

captured the searches on videotape: a VCR was attached to the graphics card of each computer via an RCA plug and a table microphone was connected to the audio input of each VCR via an audio mixer, allowing all web searching behaviour to be directly captured onto video tape along with student conversation. A similar method was used in their study on the design criteria for children's web portals (Large *et al.*, 2002b). Both studies also relied heavily on qualitative methods such as interviews and focus group interviews. Lazonder *et al.* (2000) studied the differences between novice and experienced users in searching information on the web. The experiment was conducted on a Dutch search engine called Ilse (www.ilse.nl). Each computer was equipped with a registration program that captured the action from screen and saved it in a dribble file as an AVI movie. Lucas and Topi (2002) give a detailed description of their data analysis and findings in an experiment on the impact of query term and operator usage on web search results and their relevance. They, however, do not clearly explain how the data was collected (in the introduction brief reference is made to the shortcomings of transaction logs).

David Nicholas and his co-workers report on a number of web information seeking research projects. Nicholas *et al.* (1999a) report on problems related to web log analysis, while Nicholas *et al.* (1999b) discuss methods developed to monitor the uses of case study newspapers. Nicholas *et al.* (2004) build on their earlier studies (Nicholas *et al.*, 1999c, 2001, 2002) by collecting data through logs and questionnaires on the use of a range of consumer health digital information platforms to suggest a new typology of digital users, including bouncers, checkers and returnees (since they do not specify the software used, it is assumed that they used browser tracking).

Sormunen and Pennanen (2004) report on the findings of a log analysis on user problems in exercising Boolean and best-match queries. The analysis was part of their project to introduce the query performance analyser, an instructional tool for information retrieval learning environments. Hochheiser and Shneiderman (2001) report on the use of visualisations of WWW log data to inform site design, while Griffiths *et al.* (2002) offer an excellent discussion on combining transaction log analysis and protocol analysis when studying user-system interaction.

An aspect frequently addressed with regard to web tracking in these studies (and still without solution) is user or client privacy: what should remain confidential about a search: the user's identity or the user's moves or activities? (Sandore *et al.*, 1993, p. 106). Privacy and confidentiality become key issues when considering the advertisements on the web tracking software suppliers' websites: monitoring your employees' e-mails and online shopping, monitoring your spouse's surfing of pornographic websites, part taking in cyber sex, or monitoring your child's discussions in a chat group.

Value and shortcomings noted for web tracking

Articles in the field of information seeking normally give a brief overview of their rationale for using web tracking, the software that is used, and some

shortcomings foreseen. Nicholas and his colleagues and Choo, Detlor and Turnbull belong to the small group of authors who dwell on these in more detail, and who make a serious attempt to address at least some of the issues. Nicholas *et al.* (2002, p. 64), for example, noted that spider and robot use can account for up to a third of all website traffic use. They suggest that these can be identified by their visit to the “Robot.txt” file located on the server. It is, however, becoming more difficult due to the number of undeclared spiders.

Value of web tracking

Web tracking collects large amounts of data unobtrusively so that data can be analysed at a later stage. It can collect data 24/7 without the researcher's presence and without interrupting the information seeker. It can collect data to generate a variety of statistics, for example on referral URLs (Davis, 2004), navigation moves (Choo *et al.*, 1999, 2000a, b, c), number of visits to a website such as the Nexus database system, visitors to a website, general information seeking/searching behaviour, information channels used (e.g. specific search engines, websites, portals or databases), and the users' interaction with the system (Thelwall, 2001, p. 223). To appreciate the value of web tracking, one should also consider the potential and strengths noted for transaction logs in general. The benefits listed by Van der Berg and Fourie (2003) are therefore supplemented from Thelwall (2001), Sandore (1993) and a number of other sources. We are, however, not attempting to offer an exhaustive list. When considering the value, one should bear in mind that logs can be analysed from the servers' point-of-view (e.g. a search engine intending to draw more visitors), or from the client's point-of-view (e.g. a library user whose search skills need to be improved).

We can use web tracking to learn more about, for example, the analytical searching behaviour of users (i.e. search terms used, search strategies, spelling mistakes, etc.), how users browsed for information (e.g. using menu-driven options), or from their navigation of web spaces such as websites. For the purpose of this discussion we will distinguish between the value of web tracking for analytical searching and its value to learn more about browsing and navigation. The latter will be considered in the same category. Aspects mentioned under browsing and navigation may, however, also be of value for analytical searching.

***Learning more about analytical searching*.** Web tracking can, for example:

- Help to identify areas for database or website maintenance (e.g. authority files if users use different forms of a name, or spelling). Failed searches for common terms may also reveal data entry errors in the bibliographic records. This is also useful for web-enabled databases.
- Yield information on the types of semantic relationships that users posit among terms and headings, which can add to the body of cataloguing and classification research. Subject access, and also the use of metadata can thus be improved.

- Be used to offer feedback to users on their use of the information retrieval system (e.g. the database), which is especially important when information searching is part of their professional activities and when they need to improve their skills.
- Help to discover user needs by collecting data on the subjects and topics, and can get an idea of user searching patterns. Access points are important, as well as different search strategies.
- Be used to reveal repetitive problems in patterns of searching. How do users, for example, deal with problematic searches, null set searches and unmanageable search results? (Sandore, 1993, p. 91; Van der Berg, 2000; Van der Berg and Fourie, 2003).
- Collect data on the search length, the number of search terms used, the search techniques used.

Learning more about navigation and browsing. Web tracking can also:

- Support the development and improvement of the user interface and software by noting navigation and browsing behaviour.
- Be used to test the efficacy of changes to the system (e.g. changes to the user interface), as well as user preferences for experimental changes (Van der Berg and Fourie, 2003).
- Be used to anticipate the evolution of system use and demands. Shifts in users' searching behaviour over time can be picked up. It can give an aggregate view of the use of system resources.
- Act as a decision-making tool for networks and consortia. Sandore (1993, p. 93) writes: “In a network setting aggregate statistics reveal both similarities and differences in user searching patterns among the various institutions that comprise the network.”
- Be used for the development of marketing strategies for unutilised categories of the system and thus to draw new users. This also applies to underutilised features offered by the system, and is strongly stressed by the web tracking software developers.
- Be used to monitor usage of particular resources/channels, or moves.
- Be used to improve the human (and system) understanding of how the systems are used by the information seekers (Peters, 1993).
- Be used to collect data on any variety of user types and web spaces such as websites or search engines.
- Identify areas where training is necessary for staff as well as users.
- Give information on the use of menus and help screens.
- Give information on system response time and user “think time”.
- Provide useful information to site owners about sources of new visitors (Thelwall, 2001, p. 223): “Studies of this kind, based upon laboratory experiments or web log analyses, have helped to develop rules for improved site usability. One example of this is the identified need to minimise the number of clicks needed to get to any information from the home page.”

Web tracking can also be used to shed light on cognitive behaviour if used in combination with other methods such as protocol analysis, personal interviews, questionnaires or videotapes. Sandore (1993, p. 94) explains:

> The existing body of research in this area suggests that experienced searchers approach an IR system with a different set of assumptions from an inexperienced searcher regarding how the system will function.

Nicholas *et al.* (1999a, p. 264) declare:

> There is a mistaken belief among the profession that the future is all about sharing information (knowledge management style), or storing and distributing information (digital library style), but it is, in fact, about getting closer to what people need in the way of information and producing it in a processed, packaged form for the individual to consume at a particular point in time. Customisation, individualisation and segmentation can only come on the back of personal detail and knowledge. The logs (and their allies – the subscriber databases and the cookies) provide the means by which this data can be collected. Today's logs are therefore vital to understanding the world we find ourselves in. They are the mirror on the wall.

Although the data collected through transaction logs (quantitative data) is very useful, it does not offer insight into qualitative aspects such as users' preferences, affective experiences and rationale for choices. To really draw benefit from the data collected by transaction logs, we should combine the measurement of the data (quantitative measurements) with an understanding of the rationale behind these measurements (qualitative data). Griffiths *et al.* (2002) have the following to say about transaction logs:

> Whilst transaction log analysis has considerable value as a data collection method, it has its limitations and it is best used in conjunction with a method which captures data regarding users' real information needs, comments and reactions whilst using a system and satisfaction with the system.

Pharo and Järvelin's (2004) search situation transition (SST) is an excellent example of the combined use of transaction logs with other methods. Bates (2004) considers the use of narrative interviewing in collecting data on everyday information behaviour research, while Kari and Savolainen (2001) also cover very interesting approaches to data collection concerning everyday life.

Shortcomings of web tracking

The shortcomings of web tracking is also widely acknowledged in web information seeking literature, with the main shortcoming that it only sheds light on one part of the phenomenon, namely the actual physical moves made by the information seeker. It cannot offer any information on the information seeker's real intentions, motivation, rationale for decisions, emotional experiences, or any background on personal characteristics, learning and personality styles.

> Existing metrics really just skate the surface and rarely deliver the quality information that web managers and sponsors are looking for to appraise their not inconsiderable investments. Essentially, they are being used simply as a palliative (Nicholas *et al.*, 2002, p. 64).

There is also reasonable agreement to Thelwall's (2001) statement that web logs offer only snapshots of the actions of a particular set of users (i.e. the activities at a particular point in time). Thelwall (2001, p. 217) continues to explain:

> It is contended here that there is, nevertheless, the potential for a close analysis of the data to reveal information that is unexpected in the context of navigation between sites and, therefore, to provide a novel hypothesis about aspects of online activity.

In addition to the value of web tracking, we should also acknowledge the shortcomings. Van der Berg and Fourie (2003) list the following:

- User groups are often undefined, without distinction between novice users and information intermediaries trained to use a system.
- Users' levels of information literacy, education and experience with the system or the subject domain of the search strategy are not indicated.
- Reasons for the search or search strategies are not indicated.
- Users' beliefs about the information retrieval system that are logical preconditions for a search, such as that the system is a possible place to find what is wanted, are not explored.
- Users' understanding of fundamental aspects regarding multiple discourses on a topic over time within a domain of knowledge is not considered (Iivonen and Sonnenwald, 1998).
- Users' linguistic expressive ability is not revealed.
- The social aspects of search behaviour are not revealed (Kurth, 1993).
- Transaction logs sometimes do not correlate with the users' observations of their behaviour. According to Kurth (1993) logs may show results from the point of view of the system, but may not accurately capture the users' experience and perceptions. They can help to identify only certain types of errors.
- It is hard (or even impossible) to identify users' real information needs.
- Actual uses of search results are unknown.
- Boundaries of searches (starting and ending points) are unclear if public computer workstations are used. It is also difficult to identify individual search sessions.
- Volume of data generated can be very difficult to analyse (e.g. the Excite studies).
- Users' perceptions of and satisfaction with their searches are not recorded, and the logs cannot measure the information needs that users are unable to express in their search statements.
- Statistics/data gathered should be interpreted with great care. A number of factors that can influence and skew the data have been discussed (e.g. internal site architecture).

Nicholas *et al.* (1999a, p. 265) also highlight some further problems:

- Nobody logs off on the web (you have to allow for a suitable time of inactivity, for example, 30 minutes), and then assume that they are no longer there.
- People can be locked on to the web but are not using it (for example, taking a coffee break).
- The fact that the page was downloaded does not mean that anyone actually wanted it (the person was, for example, provided with an irrelevant link).
- When selecting web tracking software, one should consider the strengths associated with web tracking to identify features that should be supported by the software, as well as the shortcomings, to identify software features that may address these.

Criteria for evaluation of web tracking software

There are many commercial programs available, some of which are very expensive and can be difficult and sophisticated to handle and/or manage, for example, programs such as WebTrends live enterprise edition which can cost as much as a thousand dollars a month. Others such as SurfSpy are inexpensive and easy to use, almost invisible, but the retrieved information is also a minimum. Some of these programs are discussed by Kemp (2001). Software that has been noted as potentially useful for web information seeking studies include the following:

- Activity Monitor (www.softactivity.com).
- Grand Art (www.keyzone.co.uk/frame_pc.html).
- Lotus' ScreenCam (www.lotus.com/home.nsf/welcome/screencam).
- NetTracker (www.sane.com).
- SpeedTracker (www.speedtracker.de).
- Spotfire (www.spotfire.com).
- TechSmith's Camtasia (www.camtasia.com).
- WinWhatWhere's Investigator (www.winwhatwhere.com).
- Lotus ScreenCam (www.Lotus.com).
- NetSnitch (www.netsnitch.com).
- SurfSpy (www.bysoft.se/sureshot/surfspy/index.html).
- WebTrends (www.netiq.com/webtrends/default.asp).
- Spector Pro 5.0 (www.spectorsoft.com).
- TheCounter.com (www.thecounter.com).
- DeepMetrix (www.deepmetrix.com).
- ScreenCam (www.smartguyz.com).

When evaluating web tracking software we need to distinguish between software that monitors activities on individual computers, including computers connected to a LAN or company network (e.g. Spector Pro 5.0 (www.spectorsoft.com)) and software monitoring activities on a specific website or browser server (e.g. TheCounter.com (www.thecounter.com) or DeepMetrix (www.deepmetrix.com)). In the case of the latter the data is sometimes referred to as visitor intelligence.

For the evaluation criteria we will draw from the literature on the evaluation of multimedia software (CIDOC Multimedia Working Group, 1997), CD-ROMs and databases (Fourie and Behrens, 1997), user interfaces (Shneiderman, 2005) and information retrieval systems in general (Chu, 2003), as well as the requirements set by web information seeking studies. According to Griffiths *et al.* (2002) it is, for example, a requirement that web tracking software should be able to show: elapsed time, mouse movement of the end user, input of the end user, and response of the system (e.g. error messages, retrieved items). It should also be unobtrusive, the researcher should have control, it should be possible to synchronise transaction logs with verbal protocols, and it should be possible to use the software across different interfaces.

In the use of web tracking software for information seeking studies, the functionality of the software in collecting meaningful data is very important. Although the purpose of this paper is to encourage users to make informed selections, and to also consider software not reported in the information seeking literature, it should be born in mind that the selection process should not be too tedious and cumbersome. It should not take more time than the actual research! The list of criteria presented here is, therefore, limited to what is considered a pragmatic list of criteria for studies of web information seeking behaviour. It can be used as a checklist.

Before applying the criteria, the researcher should get clarity on the following:

- Purpose of the research: what does he/she want to do with the data collected (e.g. improving teaching practices and assignment results, increasing website use, increasing visitors to the website, increasing the usefulness of the access points to the website)?
- Target group.
- Experiential situation (e.g. real-life such as in the Excite studies or a controlled situation such as Bilal's (2000, 2001) and Bilal and Kirby's (2002) studies with school children).
- Which data does he/she need to “do the something”? This will lead to the next question: which metrics will be measured and how does the researcher define these? The next question is then whether the software will be able to collect the data.
- Can all the data be collected through web tracking, or should it be complimented by other methods (e.g. think aloud protocols) (i.e. should the software allow for the use of such complementary methods)?
- Which problems are foreseen that might be addressed by the software?
- How much funds are available for purchasing the software? (Perhaps this should be the first question to ask).

The following criteria are suggested:

1. Supplier related criteria: this includes supplier reliability (as far as known), cost (this is often the key factor on which a decision will depend), licensing agreement (e.g. single license, multiple license), support (e.g. online help, e-mail helpdesk), availability of manuals and

supporting documentation, ease of ordering and delivery (e.g. via the WWW and credit cards and downloading of software), agreement on updating software.

2. Support available to make an informed selection: this includes demonstration versions, advertising information available on WWW, online documentation available for downloading, feedback from other users (e.g. information seeking literature or from supplier website), reviews.
3. Technical requirements: this includes hardware requirements, software requirements, compatibility with other software, ease of installation, security, protection of information seeker's privacy and confidentiality, remote monitoring, and multi-platform support.
4. User interface and user friendliness: this includes online help, ease of use, cognitive load on the researcher, customisation and interactivity. Although the user interface is normally considered extremely important in software selection (Shneiderman, 2005), such things as nice colours, buttons and uncluttered screen layout may be a benefit, but is not essential for transaction logs software for information seeking studies. The cognitive load on the researcher in learning and using the software is, however, very important: it should have an intuitive interface.
5. Functionality: (this includes the data captured, metrics that can be monitored, statistics gathered, report generation). A distinction is drawn between:
    - Functionality required for information seeking studies (the researcher can highlight specific criteria that are essential to data collection for a particular project). Literature reports on web information seeking studies were used to identify functionality criteria (e.g. the reports by Choo *et al.* (1999) and Bilal (2000, 2001)).
    - Additional functionality offered by the software (these might perhaps be relevant to information seeking studies, but is not considered essential). As the scope of information seeking studies change these criteria might, however, be added to the criteria for (a). Additional functionality can, for example, include multi-lingual language support, multiple time zone support, collection of statistics about search engine performance, ability to deal with false traffic, visualisation abilities and access blocking.

The functionality required for information seeking studies will be explained in more detail. It should be possible to collect information on the following:

- Search string or search query: this can indicate the topic or subject, the search terms, the use of search syntax, searches for specific media types such as PDF, audio or video, limitation to specific web document parts (e.g. URL, date, language). It can give an indication of the average length of search strings/search queries, the number of search terms used, the use of advanced search syntax, topics of interest, and perhaps a very limited indication of information needs.
- Search sessions: this can provide information on the length of a search session, the duration of the session, time of day (e.g. morning,

afternoon or evening), number of search sessions, and switching between applications. Choo *et al.* (1999, 2000a, b, c) refer to significant episodes of information seeking. To be a useful metric search sessions should, however, be clearly defined at the onset of the project.

- Search paths: this can provide information on the web point of departure (e.g. the website of the academic information service, a search engine, a professional organisation, bookmarks), links followed, backtracks, attempts at downloads, moves between different search engines or different databases. This links closely to the monitoring of keystrokes. This can offer information on the extent of browsing in one cluster of web pages, as well as the distances in terms of URLs between repeated websites. This will also include addresses visited.
- Keystrokes or moves: this can include, for example, the use of the back button, enter key, navigation behaviour, use of web applications, use of menus.
- User identification: this information can be used to determine frequency of visits, diversity of websites visited, return visits to specific websites, user demographics, as well as the number of pages or screens viewed (page use or page impressions). The question, however, arises whether it is possible to link the information seeking to a user that can be linked to a specific computer workstation.
- Screen captures: this information is very useful to determine the information provided to the information seeker, and which he/she could use to make a decision on the next move, as well as on files downloaded.
- System responses (can be included with screen captures): this can offer information on the number of hits and system error messages.
- Combined use: with other methods such as verbal protocol analysis.
- User control of tracking software (e.g. activating or deactivating software, sending messages, logging users off and copying files): this can also include control by the end-user (person searching the web to protect his/her privacy and confidentiality).
- System related data: this can include data such as the number of visitors to a website, the time spent viewing a page, number of pages viewed per search session, session time length, number of screens printed or downloaded, and re-visits to a site. The more times a user returns to a site, the more satisfied the user is assumed to be with the site. Coming back to a site appears to constitute conscious and directed use. According to Morris, as cited by Nicholas *et al.* (2002, p. 69), people return to a site because they like what they find. Nicholas *et al.* (2002, p. 74) refer to this as the “returnees metric”. Systems data can also be collected on single page sessions: users view only one page during the time they spend at the website, as well as on the number of sessions, and time of page view. Nicholas *et al.* (2002, p. 74) also refer to “bouncers' metrics”: the percentage of users accessing and leaving a site without looking at an information page.
- Report and statistics generation (e.g. the ability to export logs to HTML, MSExcel, CSV files, combination of log files from individual computers into one file).

The above are reflected in Table I, which can be adapted as a criteria matrix for the evaluation of specific software.

Supplier related criteria	Server monitoring or client side monitoring
	Supplier reliability
	Cost
	Licensing agreement
	Supporting documentation
	Technical support
	Ordering and delivery
	Updating software
Support available to make an informed selection of software	Demonstration version
	Advertising information
	Online documentation
	User feedback
Technical aspects	Hardware
	Software
	Compatibility with other software
	Ease of installation
	Security
	Information seeker confidentiality/privacy
User interface	Cognitive load (i.e. how difficult is it to use the software)
	User friendliness
	Ease of navigation
Functionality	(a) Functionality required for information seeking studies
	Search string/search query
	Search sessions
	Search paths
	Keystrokes/moves
	User identification
	Screen captures
	System responses
	Combination with other methods
	User control
	Report generation
	(b) Additional functionality (need to be filled in as required by the researcher)

Table I.
Criteria for evaluation of tracking software

Conclusion

If we recognise the shortcomings of transaction logs, and supplement it with qualitative research methods, it can be a very useful method to enhance our understanding of information seeking behaviour. The quality of the data collected will depend on the software used. In this regard, a criteria matrix can be used to evaluate the software of choice, or a selection of software. The application of the criteria should, however, not take more time than the actual data collection and interpretation. For future research the use of interactive visualisations of WWW log data should be investigated (e.g. as discussed by Hochheiser and Shneiderman, 2001). With regard to the more traditional transaction log analysis software they remark:

Although useful, these summaries obscure useful information and restrict users to passive interpretation of static displays. Interactive visualisations can be used to provide users with greater abilities to interpret and explore web log data (p. 331).

References

Bates, J.A. (2004), "Use of narrative interviewing in everyday information behaviour research", *Library & Information Science Research*, Vol. 26 No.1, pp.15-28.

Bilal, D. (2000), "Children's use of the Yahooligans! Web search engine. Part 1. Cognitive, physical, and affective behaviours on fact-based search tasks", *Journal of the American Society for Information Science*, Vol. 51 No.7, pp.646-65.

Bilal, D. (2001), "Children's use of the Yahooligans! Web search engine. Part 2. Cognitive and physical behaviours on complex research tasks", *Journal of the American Society for Information Science*, Vol. 52 No.2, pp.118-36.

Bilal, D., Kirby, J. (2002), "Differences and similarities in information seeking: children and adults as web users", *Information Processing and Management*, Vol. 38 No.5, pp.649-70.

Blecic, D.D., Bangalore, N.S., Dorsch, J.L., Henderson, C.L., Koenig, M.H., Weller, A.C. (1998), "Using transaction log analysis to improve OPAC retrieval results", *College and Research Libraries*, No.January, pp.39-50.

Case, D.O. (2002), *Looking for Information: A Survey of Research on Information Seeking, Needs and Behaviour*, Academic Press, Amsterdam.

Catledge, L.D., Pitkow, J.E. (1995), "Characterizing browsing strategies in the world wide web", *Computer Networks and ISDN*, Vol. 27 No.6, pp.1065-73.
Choo, C.W., Detlor, B., Turnbull, D. (1999), "Information seeking on the web: an integrated model of browsing and searching", paper presented at the ASIS Annual Meeting, available at: http://0-choo.fis.utoronto.ca.innopac.up.ac.za:80/fis/respub/asis99/ (accessed 28 July 2004).

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

Choo, C.W., Detlor, B., Turnbull, D. (2000b), "Information seeking on the Web: an integrated model of browsing and searching", *First Monday*, available at: http://firstmonday.org/issues/issuess_2/choo/ index.html (accessed 6 March 2007), Vol. 5 No.2.

Choo, C.W., Detlor, B., Turnbull, D. (2000c), "Working on the web: an empirical model of web use", paper presented at the HICSS 33 (Hawaii International Conference on Systems Science), 4-7 January, available at: http://0choo.fis.utoronto.ca.innopac.up.ac.za:80/fis/respub/HICSS/default.html / (accessed 28 July 2004).

Chu, H. (2003), *Information Representation and Retrieval in the Digital Age*, Information Today, Medford, NJ.

CIDOC Multimedia Working Group (1997), "Multimedia evaluation criteria", available at: www.archimuse.com/papers/cidoc/cidoc.mmwg.eval.crit.html (accessed 6 March 2007).

Ciliberti, A., Radford, M.L., Radford, G.P., Ballard, T. (1998), "Empty handed? A material availability study and transaction log analysis verification", *The Journal of Academic Librarianship*, Vol. 24 No.4, pp.282-9.

Cothey, V. (2002), "A longitudinal study of world wide web users' information-searching behaviour", *Journal of the American Society for Information Science and Technology*, Vol. 53 No.2, pp.67-78.

Davis, P.M. (2004), "Information-seeking behaviour of chemists: a transaction log analysis of referral URLs", *Journal of the American Society for Information Science and Technology*, Vol. 55 No.4, pp.326-32.

Dowling, T. (2001), "Lies, damned lies, and web logs", *Library Journal*, No.Spring, pp.34-5.

Ford, N., Miller, D., Moss, N. (2003), "Web search strategies and approaches to studying", *Journal of the American Society for Information Science*, Vol. 54 No.6, pp.473-89.

Fourie, I. (2002), "A review of web information-seeking/searching studies (2000-2002): implications for research in the South African context", in Bothma, T., Kaniki, A. (Eds),Progress in Library and Information Science in Southern Africa: Proceedings of the Second Biennial DISSAnet Conference, PROLISSA Conference, 24-25 October, Pretoria, Infuse, Pretoria, available at: http://www.dissanet.com (accessed 28 July 2004), pp.49-75.

Fourie, I., Behrens, S.J. (1997), "South African studies: a review of NISC's anthology CD-ROM", *South African Journal of Library and Information Science*, Vol. 65 No.2, pp.116-23.

Goodrum, A., Spink, A. (2001), "Image searching on the Excite web search engines", *Information Processing and Management*, Vol. 37 No.2, pp.295-311.

Griffiths, J.R., Hartley, R.J., Wilson, J.P. (2002), "An improved method of studying user-system interaction by combining transaction log analysis and protocol analysis", *Information Research*, available at: http://InformationR.net/ir/7-4/paper139.html (accessed 28 July 2004), Vol. 7 No.4.

Hochheiser, H., Shneiderman, B. (2001), "Using interactive visualizations of WWW log data to characterize access patterns and inform site design", *Journal of the American Society for Information Science and Technology*, Vol. 52 No.4, pp.331-43.

Hsieh-Yee, I. (2001), "Research on web search behaviour", *Library and Information Science Research*, Vol. 23 pp.167-85.

Hunter, R.N. (1991), "Successes and failures of patrons searching the online catalogue at a large academic library: a transaction log analysis", *RQ*, No.Spring, pp.395-402.

Huntington, P., Nicholas, D. (2006), "Improving the relevance of web menus using search logs: a BBCi case study", *Aslib Proceedings*, Vol. 58 No.1/2, pp.118-28.

Huntington, P., Nicholas, D., Warren, D. (2004), "Digital visibility and its impact upon online usage: case study of a health website", *Libri*, Vol. 54 No.4, pp.211-20.

Iivonen, M., Sonnenwald, D.H. (1998), "From translation to navigation of different discourses: a model of search term selection during the pre-online stage of the search process", *Journal of the American Society for Information Science*, Vol. 49 No.4, pp.312-26.

Jamali, H.R., Nicholas, D., Huntington, P. (2005), "The use and users of scholarly e-journals: a review of log analysis studies", *Aslib Proceedings*, Vol. 57 No.6, pp.554-71.

Jansen, B.J. (2000), "The effect of query complexity on web searching results", *Information Research: Electronic Networking Applications and Policy*, available at: http://0-informationr.net.innopac.up.ac.za:80/ir/6-1/paper87.html (accessed 28 July 2004), Vol. 6 No.1.
Jansen, B.J., Pooch, U. (2001), "A review of web searching studies and a framework for future research", *Journal of the American Society for Information Science and Technology*, Vol. 52 No.3, pp.235-46.

Jansen, B.J., Goodrum, A., Spink, A. (2000a), "Searching for multimedia: analysis of audio, video and image queries", *World Wide Web*, Vol. 3 No.4, pp.249-54.

Jansen, B.J., Spink, A., Saracevic, T. (2000b), "Real life, real users, and real needs: a study and analysis of user queries on the web", *Information Processing and Management*, Vol. 36 No.2, pp.207-27.

Kari, J., Savolainen, R. (2001), "Web searching in the context of information seeking in everyday life: the cases of civic and spiritual action", available at: www.uta.fi/~csjakar/kari-savolainen.pdf (accessed 28 July 2004).

Kaske, N.K. (1993), "Research methodologies and transaction log analysis: issues, questions, and a proposed model", *Library Hi Tech*, Vol. 119 No.2, pp.70-86.

Ke, H.-R., Kwakkelaar, R., Tai, Y.-M., Chen, L.-C. (2002), "Exploring behaviour of e-journal users in science and technology: transaction log

analysis of Elsevier's ScienceDirect OnSite in Taiwan", *Library and Information Science Research*, Vol. 24 pp.265-91.

Kemp, T. (2001), "The many flavours of web tracking – varying tools and techniques let companies gauge site performance and visitors' interaction", *Internet week*, 26 November, No.887.

Kim, K.-S., Allen, B. (2002), "Cognitive and task influences on web searching behaviour", *Journal of the American Society for Information Science and Technology*, Vol. 53 No.2, pp.109-19.

Kurth, M. (1993), "The limits and limitations of transaction log analysis", *Library Hi Tech*, Vol. 11 No.2, pp.98-104.

Large, A., Beheshti, J., Rahman, T. (2002a), "Gender differences in collaborative web searching behaviour: an elementary school study", *Information Processing and Management*, Vol. 38 No.3, pp.427-43.

Large, A., Beheshti, J., Rahman, T. (2002b), "Design criteria for children's web portals: the users speak out", *Journal of the American Society for Information Science and Technology*, Vol. 53 No.2, pp.79-94.

Lazonder, A.W., Biemans, H.J.A., Wopereis, I.G.J.H. (2000), "Differences between novice and experienced users in searching information on the world wide web", *Journal of the American Society for Information Science*, Vol. 51 No.6, pp.576-81.

Li, X. (1999), "Library web page usage: a statistical analysis", *The Bottom Line*, Vol. 12 No.4, pp.153-9.

Lucas, W., Topi, H. (2002), "Form and function: the impact of query term and operator usage on web search results", *Journal of the American Society for Information Science and Technology*, Vol. 53 No.2, pp.95-108.

Moukdad, H., Large, A. (2001), "Users' perceptions of the web as revealed by transaction log analysis", *Online Information Review*, Vol. 25 No.6, pp.349-58.

Nachmias, R., Gilad, A. (2002), "Needle in a hyperstack: searching for information on the world wide web", *Journal of Research on Technology in Education*, Vol. 34 No.4, pp.475.

Nicholas, D., Huntington, P., Lievesley, N., Withey, R. (1999a), "Cracking the code: web log analysis", *Online and CD-ROM Review*, Vol. 23 No.5, pp.263-9.

Nicholas, D., Huntington, P., Williams, P., Lievesley, N., Dobrowolski, T., Withey, R. (1999b), "Developing and testing methods to determine the use of web sites: case study newspapers", *Aslib Proceedings*, Vol. 51 No.5, pp.144-54.

Nicholas, D., Huntington, P., Williams, P. (1999c), "Developing and testing methods to determine the use of the web sites: case study newspapers", *Aslib Proceedings*, Vol. 51 No.5, pp.144-54.

Nicholas, D., Huntington, P., Williams, P. (2002), "Evaluating metrics for comparing the use of websites: a case study of two consumer health web sites", *Journal of Information Science*, Vol. 28 No.1, pp.63-75.

Nicholas, D., Huntington, P., Williams, P., Dobrowolski, T. (2004), "Re-appraising of information seeking behaviour in a digital environment: bouncers, checkers, returnees and the like", *Journal of Documentation*, Vol. 60 No.1, pp.24-43.

Nicholas, D., Huntington, P., Watkinson, A. (2005), "Scholarly journal usage: the results of deep log analysis", *Journal of Documentation*, Vol. 61 No.2, pp.248-80.

Nicholas, D., Huntington, P., Jamali, H.R., Tenopir, C. (2006), "Finding information in (very large) digital libraries: a deep log approach to determining in use according to method of access", *Journal of Academic Librarianship*, Vol. 32 No.2, pp.119-26.

Ozmutlu, S., Spink, A., Ozmutlu, H.C. (2003a), "Multimedia web searching trends: 1997-2001", *Information Processing and Management*, Vol. 39 No.4, pp.611-21.

Ozmutlu, S., Spink, A., Ozmutlu, H.C. (2003b), "Are people asking questions of general web search engines", *Online Information Review*, Vol. 27 No.6, pp.396-406.

Ozmutlu, S., Spink, A., Ozmutlu, H.C. (2004), "A day in the life of web searching: an exploratory study", *Information Processing and Management*, Vol. 40 pp.319-45.

Palmquist, R.A. (2001), "An overview of usability for the study of users' web-based information retrieval behaviour", *Journal of Education for Library and Information Science*, Vol. 42 No.2, pp.123-36.

Peters, T.A. (1989), "When smart people fail: an analysis of the transaction log of an online public access catalogue", *The Journal of Academic Librarianship*, Vol. 15 No.5, pp.267-73.

Peters, T.A. (1996), "Using transaction log analysis for library management information", *Library Administration and Management*, Vol. 10 No.1, pp.20-5.

Peters, T.A. (1993), "The history and development of transaction log analysis", *Library Hi Tech*, Vol. 11 No.2, pp.41-66.

Pharo, N., Järvelin, K. (2004), "The SST method: a tool for analysing web information search processes", *Information Processing and Management*, Vol. 40 pp.633-54.

Ross, N., Wolfram, D. (2000), "End user searching on the internet: an analysis of term pair topics submitted to the Excite search engine", *Journal of the American Society for Information Science*, Vol. 51 No.10, pp.949-58.

Sandore, B. (1993), "Applying the results of transaction log analysis", *Library Hi Tech*, Vol. 11 No.2, pp.87-97.

Sandore, B., Flaherty, P., Kaske, N.K., Kurth, M., Peters, T. (1993), "A manifesto regarding the future of transaction log analysis", *Library Hi Tech*, Vol. 11 No.2, pp.105-6.

Shneiderman, B. (2005), *Designing the User Interface: Strategies for Effective Human-computer Interaction*, 4th ed., Pearson-Addison-Wesley, Reading, MA.

Sormunen, E., Pennanen, S. (2004), "The challenge of automated tutoring in web-based learning environments for information retrieval instruction", *Information Research*, available at: http://InformationR.net/ir/9-2/paper169.html (accessed 28 July 2004), Vol. 9 No.2, pp.169.

Spink, A., Xu, J.L. (2000), "Selected results from a large study of web searching: the Excite study", *Internet Research: Electronic Networking Applications and Policy*, available at: http://0-informationr.net.innopac.up.ac.za:80/ir/6-1/paper90.html (accessed 28 July 2004), Vol. 6 No.1.

Spink, A., Wolfram, D., Jansen, M.B.J., Saracevic, T. (2001), "Searching the web: the public and their queries", *Journal of the American Society for Information Science and Technology*, Vol. 52 No.3, pp.226-34.

Spink, A., Ozmutlu, H.C., Ozmutlu, S. (2002), "Multitasking information seeking and searching processes", *Journal of the American Society for Information Science and Technology*, Vol. 53 No.8, pp.639-52.

Spink, A., Ozmutlu, H.C., Lorence, D.P. (2004), "Web searching for sexual information: an exploratory study", *Information Processing and Management*, Vol. 40 pp.113-23.

Sullenger, P. (1997), "A serials transaction log analysis", *Serials Review*, Vol. 23 No.3, pp.21-6.

Thelwall, M. (2001), "Web log file analysis: backlinks and queries", *Aslib Proceedings*, Vol. 53 No.6, pp.217-23.

Van der Berg, H. (2000), "An analysis of search strategy trends in order to enhance the web search interface of the Nexus database system", paper

presented at the CRIS2000 Conference, Espoo, 25-27 May 2000, available at: www.cordis.lu/cris2000/src/product.htm (accessed 28 July 2004).

Van der Berg, H., Fourie, I. (2003), "A story told by Nexus transaction logs: what to make of it", *Mousaion*, Vol. 21 No.2, pp.11-40.

Wallace, P.M. (1993), "How do patrons search the online catalogue when no one's looking? Transaction log analysis and implications for bibliographic instruction and system design", *RQ*, Vol. 33 pp.239-51.

Wang, P. (1999), "Methodologies and methods for user behavioural research", in Williams, M.E. (Eds),*Annual Review of Information Science and Technology*, Information Today, Medford, NJ, Vol. Vol. 34 pp.53-99.

Wang, P., Hawk, W.B., Tenopir, C. (2000), "Users' interaction with world wide web resources: an exploratory study: using a holistic approach", *Information Processing and Management*, Vol. 36 No.2, pp.229-51.

Whitmire, E. (2004), "The relationship between undergraduates' epistemological beliefs, reflective judgment, and their information-seeking behaviour", *Information Professing and Management*, Vol. 40 No.1, pp.97-111.

Williams, P., Gunter, B. (2006), "Triangulating qualitative research and computer transaction logs in health information studies", *Aslib Proceedings*, Vol. 58 No.1/2, pp.129-39.

Wilson, T.D. (1999), "Models in information behaviour research", *Journal of Documentation*, Vol. 55 No.3, pp.249-70.

Wolfram, D., Spink, A., Jansen, B.J., Saracevic, T. (2001), "Vox populi: the public searching of the web", *Journal of the American Society for Information Science and Technology*, Vol. 52 No.12, pp.1073-4.

Wyly, B.J. (1996), "From access points to materials: a transaction log analysis of access point value for online catalog users", *LRST*, Vol. 40 No.3, pp.211-36.

Zink, S.D. (1991), "Monitoring user search success through transaction log analysis: the WolfPAC example", *Reference Services Review*, Vol. 19 No.1, pp.49-56.

Corresponding author

Ina Fourie can be contacted at: ina.fourie@up.ac.za