Using 'Search Transitions' to Study Searchers' Investment of Effort: Experiences with Client and Server Side Logging

Nils Pharo and Ragnar Nordlie

Faculty of Social sciences, Oslo and Akershus University College of Applied Sciences
PO Box 4, St. Olavs Plass, 0130 Oslo, Norway
{nils.pharo,ragnar.nordlie}@hioa.no

Abstract. We are investigating the value of using the concept 'search transition' for studying effort invested in information search processes. In this paper we present findings from a comparative study of data collected from client and server side loggings. The purpose is to see what factors of effort can be captured from the two logging methods. The data stems from studies of searchers interaction with an XML information retrieval system. The searchers interaction was simultaneously logged by a screen capturing software and the IR systems logging facility. In order to identify the advantages and disadvantages we have compared the data gathered from a selection of sessions. We believe there is value in identifying the effort investment in a search process, both to evaluate the quality of the search system and to suggest areas of system intervention in the search process, if effort investment can be detected dynamically.

Keywords: Information retrieval, methods, evaluation.

1 Introduction

Numerous studies have been performed on searchers' interaction with IR systems, in non-web systems [1–3], but in particular with the advent of the Web [4–6]. The study reported in this paper has as its point of departure the notion that effort invested in search processes can be investigated in the light of the concept 'search transition' [7]. Search transitions are constructed to take into consideration the mental effort invested by the searcher during a search process. Effort spent during information searching could be invested in learning to use the system, the adaption of specific system functionalities in the searcher's search strategy; the time spent investigating the details of query result lists etc. Search processes can be split into series of transitions, which in turn can be categorized into different types.

Search transitions can be identified and categorized by thorough analysis of information system transaction logs. There are, we believe, significant differences in the type of information that can be gathered from server and client logs, respectively. Hence it will also differ to what degree the log types provide details of effort invested in the search process. In the present paper we try to answer the following question:

M. Lupu, E. Kanoulas, and F. Loizides (Eds.): IRFC 2013, LNCS 8201, pp. 33–44, 2013.

What signs of effort can server and client side transaction logs reveal in different types of transitions?

The capturing of server and client logs [8] is a common way to gather data for analyzing IR interaction. One direction of research has been quantitatively oriented studies where researchers have performed analysis of server logs that have captured up to 1 billion queries [4] submitted to the IR system. These kinds of studies have revealed many interesting characteristics of searchers' query formulation and reformulation, e.g. that queries are typically quite short; the use of result lists, e.g. searchers' tendency to only look at a very limited set of documents; and the topicality of queries, e.g. that a large share of web queries are related to pornography.

On the other hand, several studies have been designed that use client side logging, where the goal has been to study e.g. the search processes of searchers across several web sites or to perform very detailed analysis of searchers' interaction with a particular information system. Client logs can be collected either by using client navigation software, such as a web browser, or by screen capturing software. Eye-tracking software can also log the searchers' eye movements over the screen [9].

Server logs reflect the complexity of the information system they capture, e.g. whether the system only contains document surrogates or if it also contains the documents themselves. In a typical web search engine the former will be the case, and primarily interaction with document surrogates will be covered. In order to understand what aspects of effort, as it is understood in our definition of search transitions, can optimally be identified from server logs, we have performed an analysis of a selection of server logs collected by the INEX 2008 interactive track [10], which also contains interaction with the documents. We have compared the server log data with data collected on the client side of the interaction, using the Morae screen capturing software [11]. From this comparison we can learn more on the factors that reflect the searchers' investment of mental effort, and which of these factors that can be identified using server logs, and which factors cannot.

1.1 Measuring Effort Invested in Information Searching

The term "effort", which received an early definition by Fenichel [12] as "a set of search variables [including] e.g. number of commands and descriptors [and] connect time", is quite often considered in the more general literature on information seeking behavior, with this or a variety of other, more or less similar definitions. Zippf's "law of least effort" is often invoked to explain users' choice of information channel[13] , which refers to a number of studies who take this perspective. When effort is considered in the more restricted environment of information search behavior, however, it is often relatively vaguely defined. Typically, it is treated as in [14], where, in an investigation of the influence of user experience on search outcomes, effort is considered as one of several "search language use patterns" and defined to consist of "mean number of cycles per topic, mean command frequency per topic, and mean number of documents visited per cycle" without any motivation for this choice of parameters. A number of authors invoke "cognitive effort" as distinct from observable, logged actions in their characterization of search [15]. Cognitive effort is a

concept well known from fields such as psychology and decision theory, but as a parameter of search effort it is often treated with a similar lack of specific definition as the concept of effort in general. Where it is defined the measurement definitions range widely, from "pupil dilation" in an eye-tracking study of search and evaluation behavior [16] to "number of iterations, i.e. queries in a search" [17].

The term *transition*, or parallel expressions such as shifts, state changes etc. is widely used in both the general literature on information seeking and more specifically in studies of information search behavior. It is generally defined in terms of a move from one state to another (or a sequence of such moves). Stages or patterns of stages appear in more and more fine-grained form in models of information seeking behavior from Ellis' and others' early models [18, 19], and are becoming more and more fine-grained, as in Xie [20], where the interest is in shifting patterns between search stages. Such stages may be identified for instance in information seeking mediation, as in [21] where stages are identified as sets of cognitive and operational elements and transitions between stages are identified through vocabulary changes in dialogue. Transitions have been of particular interest to studies of search system interactions, where it has been thought that being able to detect transitions or distinct shifts in interaction would enable the automatic detection of patterns that might engender some kind of machine assistance or inform interface design. Variants of Markov modeling have often been suggested for such modeling, in [22] weaknesses of this approach is discussed, and an alternative modeling approach with Petri nets are suggested. In this paper and many others the transitions themselves are vaguely defined, and this is a persistent problem in the literature.

We believe our suggested concept, *search transition*, can be used to measure effort in the form of number of search transitions and through analyzing the search transition patterns followed by searchers. Each transition represents a combination of factors involving searcher interaction with information items. Factors that represent mental effort invested during IR interaction include query formulation and reformulation, the selection of source and document types, the number of documents and/or other units of information viewed etc. The rationale behind using search transition as a measure of effort is to take into account the cognitive load required by searchers to deal with a variety of such challenges during interaction. Different IR systems facilitate different types of search transitions, e.g. ISI citation indexes exemplifies a complex IR system with many filtering and refinement options whereas the default search options of web search engines offer quite simple interaction options.

2 Method

The search system applied in the study is a java-based retrieval system built within the Daffodil framework [23], which resides on a server at and is maintained by the University of Duisburg. The search system interface (see Figures 1 and 2) is developed for the INEX (Initiative for the Evaluation of XML retrieval) 2008 interactive track [10]. The database consists of approximately 650 000 Wikipedia articles, which

have been indexed on three levels of granularity; as full article, section level, and subsection level.

Searchers were asked to perform two search sessions, to solve one fact-finding task and one research task, each task was formulated as a simulated work task [24]. Searchers were, for each task, asked to assess the relevance of any document (article) or document element (section or subsection) they viewed during the process. All sessions were logged by the IR system, in order to compare the server and client logs a selection of sessions were also screen captured on the client side using Morae. In our comparative analysis we have looked at 8 sessions in detail to compare the advantages and disadvantages of each of the two logging procedures in connection with identifying different expressions of effort during information searching, relating these to explicit search transitions. In addition we have studied the individual transition patterns of two selected sessions in order to study effort invested throughout the sessions.

2.1 Server Logs

Our server logs captures information about the query input, titles of retrieved information units (i.e. articles, sections and subsections), system-suggested terms for alternative query formulation, titles of information units selected from the result list and the articles table of contents, relevance assessments, internal interaction within individual articles and parts of articles and more. All transition types can be captured by the server logs.

All events in the logs are time-stamped, which means that we can trace the sessions in high detail with respect to the order and selection of events. This makes us able to recreate what interface functionalities were used by the searchers.

2.2 Client Logs

The logs captured at the client side contain all actions made by the searcher during the session, including traces of all mouse movements, highlighting of clicks, continuous time recording to the hundredth part of a second, etc. It is possible to record searchers' utterances/talking aloud, but we choose not to do so for this experiment.

2.3 Log Comparison

Our comparative analysis has focused on characteristics in the two log types that reflect searcher effort. Search transition type (see below) has been used as the organizing factor, i.e. for each transition type we have attempted to make explicit what traces of effort can be found using server and client logs respectively.

2.4 Search Transitions

The following list of search transition types were identified through a study of the server logs of the system used in our experiment:

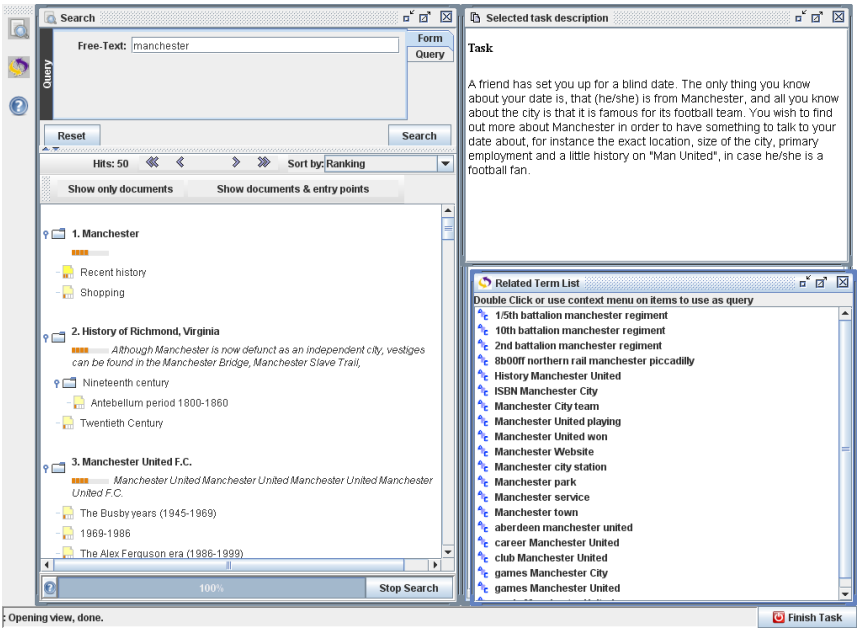

Fig. 1. Search interface of Daffodil

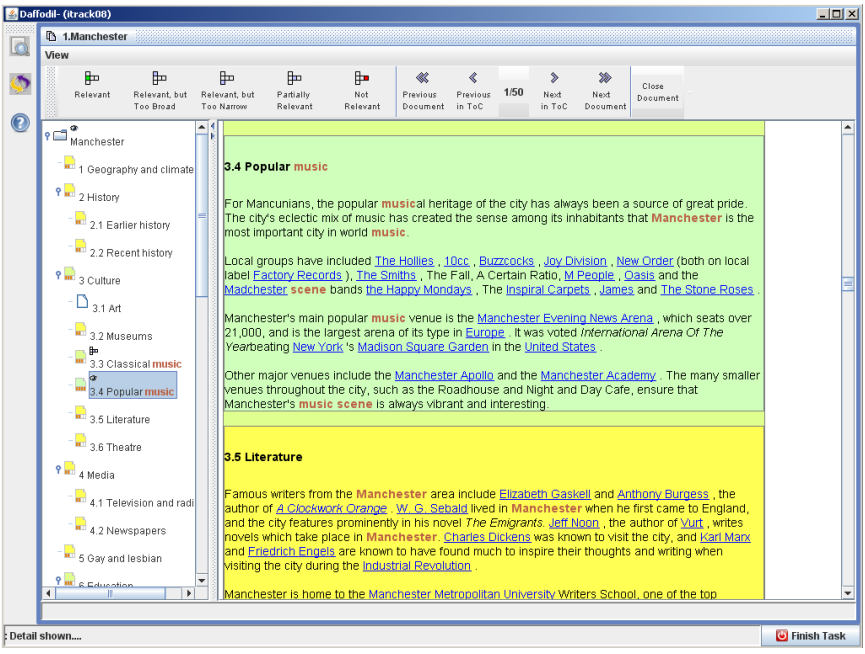

Fig. 2. Document interface of Daffodil

a) Query – result
b) Query – result – inspection
c) Result – inspection
d) Inspection – link to other page – inspection
e) Back button – link to other page – inspection
f) Use system suggested terms – results
g) Use system suggested terms – results – inspection
z) End interaction

Transition a) describes the searcher performing a query in the IR system, but no information unit is selected for further inspection (i.e. selected by a click in the result list, since this is the only expression of "inspection" identifiable in the server log). In transition b) the searcher performs a query, and then from the result list selects a unit of information (a document, a section of a document, or metadata representing the document). In transition type c) the searcher returns to the result list after having inspected a unit of information and then selects a new unit, without a new query being performed. In transition d) the searcher from within an article selects a link to another article. In transition type e) the searcher uses the system's back button to the previous page and then selects a link to another article (note that transition type e) is always preceded by transition type d)). The difference between transition types f) and g) is that in the former the searcher does not select any of the entries in the result list for further inspection after having performed the search on the search term from the suggestion list. Note that interaction within a document, for instance through the TOC, cannot be identified through server logs and is treated as part of the inspection process. Transition z) is used to indicate that system interaction stops, this could be provoked by the searcher logging out of or exiting the system in other ways or by system failure.

3 Findings

The 8 sessions contained a total of 85 server-log-identifiable transitions. Six out of the eight sessions lasted approximately 15 minutes, whereas one lasted seven minutes and 40 seconds and one ended after two minutes and 15 seconds (the latter two were both performed by the same searcher). Table 1 shows the distribution of transition types in the sessions. We see that the large majority of transitions are of types a, b and c, i.e. searchers entering queries, looking at result lists and selecting potentially relevant items. Few examples are found of the use of the system's suggested term feature. No examples of transitions d or e were found in our server log files, the client logs, however, showed many examples of searchers contemplating following internal links. In general the client logs, as expected, revealed many signs of searcher confusion in using the interface.

Table 1. Distribution of transitions in sample

Transition type	
a	25
b	24
c	20
d	0
e	0
f	3
g	2
z	8

We studied the server and client logs in order to find expressions of effort in the different types of transitions. Table 2 summarizes the findings from our log comparison study, and shows *additional* expressions of effort identifiable in the client logs.

Table 2. Server and client log comparison

Transition type	**Server log**	**Client log**
a	Duration in seconds Query terms used Number of items found Titles of the items found	Time spent waiting for system response Time spent contemplating actions (e.g. term selection) Browsing of result lists Reading of text snippets in lists Query reformulations considered, but not executed
b	Same as a) + Number of items looked at (i.e. clicked) Titles of the items looked at Relevance assessments	Same as a) + Browsing of items Reading of items Hesitation in relevance assessments
c	Number of items looked at Titles of the items looked at Relevance assessments	Browsing of items Reading of items Hesitation in relevance assessments
d		
e		
f	Available suggested terms Term(s) selected	Terms considered selected
g	Same as f + Titles of the items looked at	Same as f + Browsing of items Reading of items
z		

We see that the client logs reveal in much more detail how searchers are investing effort in interacting with the text, in particular on how the work load is divided between browsing and reading information items (articles and sections) and browsing metadata, such as titles and related terms (in transition types a, but particularly in type b and g), and on the variation of sources and on the dynamic process of query formulation and reformulation. Server logs, on the other hand, facilitate statistical data analysis due to the logs' capturing of the number of items found and used and the timestamps of all events. We have found, however, examples of server logs mixing up the order of events, making exact capturing of the process in the use of this particular system harder.

Table 3. Two session examples with transitions

Transition #	Session 1	Length	Session 2	Length
1	a	2 min 19 s	a	3 min 43 s
2	a	7 s	c	18 s
3	b	57 s	c	30 s
4	a	1 min 4 s	a	43 s
5	b	1 min 15 s	g	43 s
6	a	1 min 14 s	a	1 min 33 s
7	b	31 s	c	1 min 11 s
8	a	1 min 15 s	a	1 min 33 s
9	c	1 min 8 s	b	1 min 10 s
10	g	54 s	b	1 min 16 s
11	a	58 s	a	1 min 47 s
12	c	40 s	z	0 s
13	c	1 min 39 s		
14	z	0 s		

A close examination of the server log of two sessions showed the distribution of transition types as reported in Table 3. Here transitions are ordered chronologically and we see the time spent on each transition. Both sessions were approximately 15 minutes long.

We see from Table 2 that Type a transitions are the most common in our small sample, but also that the time spent in different transition types differ very much. If we analyze these transactions through the client logs, we find that a lot of the time spent during the session consists of waiting for query results to appear, thus time is not always a good effort indicator. In both sessions the searcher starts with a rather long transition, in Session 1 the searcher spends much time spent in inspecting the query results whereas in Session 2 a large amount of time is spent in inspecting one particular document, also here the searcher hesitates much in deciding whether the document is relevant or not. Also in the first transition in Session 2 the searcher inspect the system's related-term feature, perhaps in order to acquire inspiration for query formulation. No terms are however selected to generate new queries. Transition 10 in Session 2 (a Type b transition) contains an interesting sample of query formulation, here the searcher

spends the approximately first half a minute to formulate two different queries without submitting them before settling for a third version. This kind of effort investment cannot be captured by the server logs. In Session 1, transition 10 (Type g transition), it is interesting to observe that as the searcher is struggling in formulating an effective query with the help of the suggested-term feature, there are several relevant items visible in the result list, but these are overlooked by the searcher. This can perhaps be considered an example of "uni-tasking", i.e. inability to deal with several items in the system's interface (=multi-task) due to heavy effort investment in one particular task. Other examples of time spent during the sessions include waiting for query results to appear, and inspection of result lists to find ideas for query terms.

In Table 4, looking at the first a) + b)-type transition in more detail, we identify the following behavior (based on the premise that mouse movements to a large extent identify the focus of attention on the screen):

Table 4. Two transtitions in detail

00:00	Reads task
00:30	Writes search term (st): Presidential election
00:36	Hesitates
00:41	Extends st: Presidential election France
00:48	Corrects st: Presidential election Europe
00:51	Clicks *search*
00:51	Waits for result (rl) and suggestion (sl) lists
01:14	Inspects rl, finds no relevant item among 3 items shown
01:22	Inspects sl, finds no relevant suggestion
01:29	Changes st; European presidential election
01:35	Clicks *search*
01:35	Waits for result
01:53	Corrects st: Presidential election
01:54	Inspects sl: selects European Election official
01:55	Receives rl for search European presidential election, and continues working with this list

We see that the two identifiable server-side transitions involves several attempts and misunderstandings, waiting time and work time interspersed, all of which is invisible and unanalyzable in the server-side log.

4 Conclusion

We believe there is value in identifying the effort investment in a search process, both to evaluate the quality of the search system and to suggest areas of system intervention in the search process, if effort investment can be detected dynamically. This calls for a machine-identifiable measure of effort, however. Our concept of search transition is

describable and identifiable in server-side logs and should thus be possible to automatically detect and apply. Server-side logs have several advantages:

- It is easy to collect data on time spent on different activities
- Data on query formulation and reformulation can be easily collected for analysis
- Easier countability of events (page retrieval, link selection etc) allows for discovery of general patterns

This is only of value, however, if the effort implied in the server-side transitions are comparable to the effort identified in client-side application of the same definition of transition. The client-side analysis permits, among other things:

- making distinctions between time spent due to hesitation, browsing of content, inspection of interface functionalities, low system response time etc
- Capturing browsing of pages and result lists
- Identification of more details in the query formulation process (e.g. queries that are edited several times before they are submitted)
- Easier understanding of the order of events (server log events are sometimes presented in an un-predictive order)
- Capturing details in the use of system functionalities that are not included in the server logs

Client logs are usable for acquiring valuable data about searchers' effort unavailable from the server. Of importance, for instance, is data that distinguishes between time spent due to system and software problems and time spent by searcher trying to launch his/her search strategies various ways.

Additional data makes it possible to create a more fine-grained taxonomy of search transition types. Different transition types can, e.g. differentiate between the effort spent waiting for system response and the effort invested in creating queries that best match the searcher's current understanding of his/her information need. This fine-gradedness supplements, but does not necessarily replace the transition taxonomy of the server-side logs.

However, in our study of the client logs we have been able to identify several signs of intellectual effort investment in terms of searcher decisions. It may seem that effort, considered in this way, is distributed randomly across different transition types at different stages of the session. Analysis of larger data sets is necessary to identify if there are clear patterns, and whether it is these instances of "micro-effort" rather than the more comprehensive transitions identifiable in the server logs which are best suited to measure user search effort.

We believe that the value gained from analyzing client side logs for understanding more about the mental effort involved from searchers justifies its use. Lessons learnt from usability studies states that from studying a rather small number of users, usability experts are able to identify a large number of the system errors [25]. Is a corresponding pattern to be found when performing microanalysis of client side logs of information search behavior? Our findings at least indicate that the analysis of quite

small sets of data (8 sessions by four different searchers) can be used to identify interesting characteristics of effort investments.

In order to understand more about how searchers invest their mental effort in information searching we suggest that as much data as possible should be collected from the server logs, including timestamps of documents retrieved and accessed from the result lists. Client side logs should complement the analysis of the server logs to identify the specific challenges of the information system in use. Preferably client logs should be collected so that they cover different "environmental" factors as broadly as possible, e.g. search sessions from different times of the day, from different locations, with different client software (e.g. different web browsers) etc. The client log data could then be used to strengthen the understanding of the effort invested in the different search transitions types that are categorized from the server log data.

Acknowledgments. We would like to thank the participants of the INEX 2008 interactive track.

References

1. Ingwersen, P.: Search Procedures in the Library—Analysed from the Cognitive Point of View. J. Doc. 38, 165–191 (1982)
2. Fidel, R.: Moves in online searching. Online Inf. Rev. 9, 61–74 (1985)
3. Bates, M.J.: The design of browsing and berrypicking techniques for the online search interface. Online Inf. Rev. 13, 407–424 (1989)
4. Silverstein, C., Marais, H., Henzinger, M., Moricz, M.: Analysis of a very large web search engine query log. Sigir Forum 33, 6–12 (1999)
5. Spink, A., Jansen, B.J.: Web Search: Public Searching on the Web. Springer (2004)
6. Jansen, B.J., Spink, A.: How are we searching the World Wide Web? A comparison of nine search engine transaction logs. Inf. Process. Manag. 42, 248–263 (2006)
7. Pharo, N., Järvelin, K.: The SST method: a tool for analysing Web information search processes. Inf. Process. Manag. 40, 633–654 (2004)
8. Jansen, B.J., Spink, A., Taksa, I. (eds.): Handbook of Research on Web Log Analysis. IGI Global (2008)
9. Balatsoukas, P., Ruthven, I.: An eye-tracking approach to the analysis of relevance judgments on the Web: The case of Google search engine. J. Am. Soc. Inf. Sci. Technol. 63, 1728–1746 (2012)
10. Pharo, N., Nordlie, R., Fachry, K.N.: Overview of the INEX 2008 Interactive Track. In: Geva, S., Kamps, J., Trotman, A. (eds.) INEX 2008. LNCS, vol. 5631, pp. 300–313. Springer, Heidelberg (2009)
11. Morae usability testing software from TechSmith, `http://www.techsmith.com/morae.html`
12. Fenichel, C.H.: Online searching: Measures that discriminate among users with different types of experiences. J. Am. Soc. Inf. Sci. 32, 23–32 (1981)
13. Bronstein, J., Baruchson-Arbib, S.: The application of cost—benefit and least effort theories in studies of information seeking behavior of humanities scholars: the case of Jewish studies scholars in Israel. J. Inf. Sci. 34, 131–144 (2008)

14. Yuan, W.: End-user searching behavior in information retrieval: A longitudinal study. J. Am. Soc. Inf. Sci. 48, 218–234 (1997)
15. Thatcher, A.: Web search strategies: The influence of Web experience and task type. Inf. Process. Manag. 44, 1308–1329 (2008)
16. Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., Gay, G.: The influence of task and gender on search and evaluation behavior using Google. Inf. Process. Manag. 42, 1123–1131 (2006)
17. Belkin, N.J., Kelly, D., Kim, G., Kim, J.-Y., Lee, H.-J., Muresan, G., Tang, M.-C., Yuan, X.-J., Cool, C.: Query length in interactive information retrieval. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, pp. 205–212. ACM, New York (2003)
18. Belkin, N.J., Cool, C., Stein, A., Thiel, U.: Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. Expert Syst. Appl. 9, 379–395 (1995)
19. Ellis, D.: A behavioural approach to information retrieval system design. J. Doc. 45, 171–212 (1989)
20. Xie, H.: Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. J. Am. Soc. Inf. Sci. 51, 841–857 (2000)
21. Olah, J.: Shifts between search stages during task-performance in mediated information seeking interaction. Proc. Am. Soc. Inf. Sci. Technol. 42, n/a–n/a (2005)
22. Kantor, P.B., Nordlie, R.: Models of the Behavior of People Searching the Internet: A Petri Net Approach. Proc. Asis Annu. Meet. 36, 643–650 (1999)
23. Fuhr, N., Klas, C.-P., Schaefer, A., Mutschke, P.: Daffodil: An Integrated Desktop for Supporting High-Level Search Activities in Federated Digital Libraries. In: Agosti, M., Thanos, C. (eds.) ECDL 2002. LNCS, vol. 2458, pp. 597–612. Springer, Heidelberg (2002)
24. Borlund, P.: Evaluation of interactive information retrieval systems. Abo Akademis Forlag (2000)
25. Nielsen, J., Landauer, T.K.: A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT 1993 and CHI 1993 Conference on Human Factors in Computing Systems, pp. 206–213. ACM, New York (1993)