Examining the effect of task stage and topic knowledge on searcher interaction with a "digital bookstore"

Nils Pharo
Department of Archivistics, Library and Information Science, Oslo and Akershus university college of applied sciences
Postboks 4, St. Olavs plass
NO-0130 Oslo
+47 22452684

nils.pharo@hioa.no

Ragnar Nordlie
Department of Archivistics, Library and Information Science, Oslo and Akershus university college of applied sciences
Postboks 4, St. Olavs plass
NO-0130 Oslo
+47 22452688

ragnar.nordlie@hioa.no

ABSTRACT
This paper reports some results from the experiment of the 2010 INEX interactive track. The experiment was designed to let searchers simulate being at two distinct stages of a work task process. Data were also collected on the test participants' topic knowledge. We have performed statistical analysis of the collected data to study differences with respect to relevance judgments and use of different types of metadata, at the different stages and for users with high and low topic knowledge.

Keywords
Interactive information retrieval, relevance judgments, information seeking.

1. INTRODUCTION
Several studies have established that the processes of information gathering, interpreting, and use changes throughout a work task [5,11]. The coupling of work task stages to actual IR interaction has also gained interest lately. In their book "The Turn", Ingwersen and Järvelin [8] call for an integrated view of information seeking and retrieval. The Information Interaction in Context (IIiX) symposium series has addressed these issues, and empirical work on these matters has been published in the central LIS journals. Particular interest has been on the relevance criteria used at different stages [23,24].

Previous theoretical and experimental work has shed some

light on how factors such as the searchers' work tasks, search tasks, task or topic knowledge, and social environment affect the search process [e.g. 16], but there are still many issues that have not been systematically evaluated.

The Initiative for the Evaluation of XML retrieval (INEX) has included an interactive track in which the focus has been to conduct experiments on searchers' interaction with XML-encoded documents of various kinds. In the 2010 interactive track the experiment was designed with the goal of studying interaction taking place at two different stages of simulated work task processes. The experiment also collected data concerning the test persons' topical knowledge about the tasks they were asked to perform.

This paper reports some of the results from the INEX 2010 interactive track. We report some findings on the effect of work task stage and topic knowledge on searchers' use of document metadata. Two research questions are dealt with:

1. How does work task stage influence the number of document surrogates used and assessed as relevant?

2. How does topic knowledge influence the number of document surrogates used and assessed as relevant?

We study these quantifiable indicators of search behavior because we consider them to constitute an indicator for the kinds of information needs a searcher encounters during the stages of a search process, and thus may indicate the kind of support a search system should offer at these different stages.

2. PREVIOUS RESEARCH
There is a plethora of literature which deals with issues that are relevant for different aspects of this study. In the literature review the reader will be introduced to a selection of very central texts supplemented by studies that illustrate issues particularly relevant for this study.

2.1 Information seeking processes

Ellis [5,3,4] studied researchers' information seeking and use behavior and found that their activities during a project can be classified as belonging to six to eight categories of behavior (starting chaining, browsing, differentiating, monitoring, extracting, verifying and ending) [4]. The model has been tested and verified to fit researchers from different disciplines, and although the order of the stages may differ they will typically follow the order listed above. The studies show that information needs are unequally distributed and dealt with across the research process. Compared with Kuhlthau's model (below) this model is more focused on actions performed by the information searcher, but it does not describe any effects with respect to actual system interaction.

Kuhlthau [10,11] developed a model of the 'information search process' consisting of six different stages which the searcher goes through; the initiation, selection, exploration, formulation, collection and presentation stages. At the various stages the searcher experiences different cognitive and emotional states. The model has been verified as relevant for many different groups of searchers, and in the second edition of her book [11] she also includes findings from other researchers. In particular, the emotional states have been the focus of research, and, even if is it included in the model, the searchers' activities at the different stages have not been subject to many investigations. Remarkably little weight has been given to searchers' interaction with information systems at the different stages.

In his comparison of the two models Wilson [28] points out that they are quite similar with respect to their level of analysis, and in both being based on empirical investigations. Other models of information seeking processes exist; see e.g. Don Case's [2] review of different models of information behavior. Since we want to examine specifically the influence of different stages on user-system interaction we have, however, emphasized these two generic and stage-oriented models here.

Some attempts have been made to study the effects of seeking process stage on the interaction with information retrieval systems [e.g. 25,19], but very few [24] have studied more than a small sample of users.

Vakkari conducted an empirical study aimed at refining Kuhlthau's theory and found that "the stages of Kuhlthau's model […] has a systematic impact on the information types sought, on the choice of search terms and tactics as well as on the assessment of relevance and contribution of the references found and full texts acquired in the task performance process" [25:55]. Pharo and Järvelin [19] studied web searchers and found examples of how the work task stage influenced the relevance level applied by searchers during interaction.

Taylor and his colleagues [24] have performed the largest study of the influence of seeking process stages on IR interaction. They used a random sample of 40 undergraduate students to investigate the types of relevance categories assigned by users at different stages of the seeking process. Their analysis proposes, e.g., that there is a relationship between the writing stage (what Kuhlthau calls the presentation stage) of a process and "source novelty" as relevance criterion.

2.2 Topic knowledge

Along with search knowledge, searchers' topic knowledge [16] is the factor most commonly investigated by information behavior researchers. Several authors have studied the effect of searchers' topic knowledge on their interaction with information systems, a few of whom are reviewed below. There may be subtle nuances between what is called subject knowledge, domain knowledge, task knowledge and topic knowledge. There is however a clear distinction between these "knowledge types", which all refer to the searchers' factual expertise, and "search knowledge", which denotes expertise in using search systems and formulating search strategies in general.

Marchionini was an early investigator of the effect of domain expertise on search behavior in hypertext systems. In [15] he compared the search behavior and efficiency of third and fourth graders with the behavior of sixth graders. He found that the older (and hence more experienced) searchers were more efficient, in that they spent less time to find more useful information. In [14] Marchionini, Lin and Dwiggins compared the search behavior of subject experts, search experts and library students in a hypertext information system and found that both kinds of experts performed better than the non-experts, exploiting their respective knowledge. Marchionini and his colleagues [13] also found differences in the approach of search experts and domain experts with respect to their approach. Hölscher and Strube [6,7] compared domain knowledge and web expertise and concluded that "while successful search performance requires the combination of the two types of expertise, specific strategies directly related to Web experience or domain knowledge can be identified." Wildemuth [27] studied students of medicine and how the increase in domain knowledge over time influenced the kind of search tactics (i.e. "sequential combinations of moves") they used. White, Dumais and Teevan [26] have performed a longitudinal, large-scale transaction log-based study of web searchers within different domains. They found that domain experts used more successful search strategies (i.e. sessions ending with a document look-up) than non-experts.

2.3 Relevance

Relevance assessment is a central tool in information science [17], and of particular importance for the evaluation of information retrieval (IR) systems. The traditional IR system evaluation method is the test collection approach,

where experts assess the topical relevance of documents in a collection and IR systems are evaluated by how efficient they are in retrieving the relevant documents and dismissing the irrelevant ones.

Saracevic's [22] taxonomy of relevance levels depicts relevance on a scale representing different degrees of user involvement. On the most "objective" level relevance is strictly term dependent in that assessed relevance is based on the match between query terms and index terms (algorithmic relevance). On the next level the match is made between the subjects expressed in query and texts (topical relevance). Then there are several levels involving the query formulator's intention, including motivational relevance, defined as the "relation between the intents, goals, and motivations of a user, and texts retrieved by a system or in the file of a system, or even in existence. Satisfaction, success, accomplishment, and the like are criteria for inferring motivational relevance" [22:214]. The different levels of relevance can be said to reflect how the three revolutions (the cognitive, the relevance and the interactive revolution) put forward by Robertson and Hancock-Beaulieu [21] have brought the user's perspective into IR research.

The recognition of relevance as a more subtle and dynamic feature in IR has also led to the introduction of non-binary relevance assessments in IR system evaluation [9]. In the INEX 2006 interactive IR experiments a two-dimensional relevance scale was used. This scale represented both the topical relevance of the item as well as "how much [textual] content is needed to understand the element" [12:390]. However, these kinds of scale have been criticized as being too complex for users and assessors to apply in experimental search behavior investigations [20]

3. METHOD
The INEX interactive track is organized with a distributed data collection procedure [18]. The research groups that want to take part in the project are given a set of guidelines to secure that data are collected in the same way by all participants. The guidelines include:

- a common recruiting procedure for experimental subjects

- a common set of user tasks and data collection instruments such as interview guides and questionnaires

- a common logging procedure for user/system interaction

- an understanding that collected data should be made available to all participants for analysis

An experimental information system has been developed and made available for use to all participating groups. The software is built within the ezDL-framework and is stored on a server at the University of Duisburg-Essen. The collection used for the experiments consists of 2.7 million records from the digital bookstore Amazon.com coupled with corresponding bibliographic records from the social cataloguing tool, LibraryThing. Amazon fields in the database includes traditional bibliographic metadata such as ISBN, title and Dewey classification as well as user reviews and editorial reviews (source, content), from LibraryThing users' tags (including occurrence frequency), "blurbs", dedications, epigraphs, first words, last words, quotations, series, awards, browse nodes, characters, places and subjects have been extracted.

The data are indexed using Apache Solr 1.4, which is based on Apache Lucene. Lucene applies a variation of the vector space retrieval model. All participating groups decided to use the system provided by the organizers.
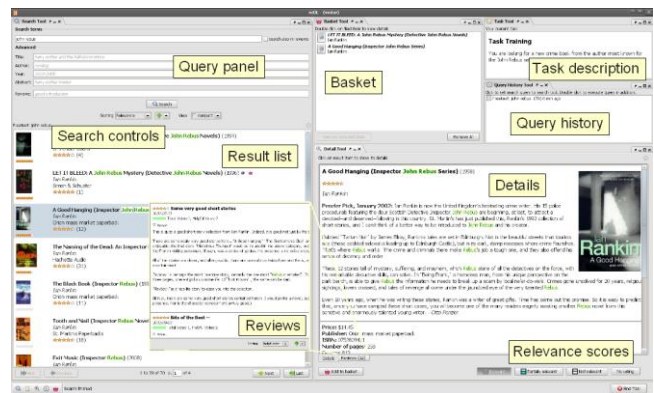

Figure 1 The search system interface

Figure 1 shows the interface of the search system. The system features includes description of the current search task, result list, bibliographic details and book reviews. The book reviews could be accessed via a list of the reviews, where the scores given by each reviewer was directly available. The experiment used a three level relevance scale and participants were instructed to determine the relevance of any examined book as "Relevant", "Partially relevant" or "Not relevant" by clicking markers at the bottom of the screen. Participants could also use the system to add wanted items into a basket. There were two slightly different interfaces to the system in use (Interface A and Interface B), the difference being that in Interface B it was not possible to query the text within book reviews or abstracts. Each participant was allocated to either Interface A or Interface B for the experiment. The different interfaces was not expected to have any effect on the research topics investigated in this paper, and we have performed independent t tests to learn whether the different interfaces has had any effects on the use of different "document surrogates" or on relevance assessment and found no significant differences.

49 participants at three different data collection sites volunteered to take part in the experiment. The participants were almost all students (at bachelor or master level) within the fields of computer science or library and information science.Each participant was asked to perform three tasks in the system; the tasks were rotated in order to control for learning effects. Two of the three tasks were formulated as simulated work tasks [1], where . the participant was allowed to choose between three tasks in each of two separate categories. One task category represented tasks that simulated a searcher being at an early "exploration" stage of a research project; the second category represented the "data gathering stage" of a research project. Both these stages have been found to represent distinct stages in information seeking processes by Kuhlthau [11] . The simulation of stages is, of course, difficult and we will discuss limitations this poses on the validity of our results in the Discussion section.

In addition the participants were asked to define a self-selected task with some constraints, i.e. they were asked not to look for known item. In this paper we primarily investigate user performance in the two assigned tasks.

The task groups were introduced as follows:

Task Group 1: The Explorative Tasks
You are at an early stage of working on an assignment, and have decided to start exploring the literature of your topic. Your initial idea has led to one of the following three research needs:
1. Find trustworthy books discussing the conspiracy theories which developed after the 9/11 terrorist attacks in New York.
2. Find controversial books discussing the climate change and whether it is man-made or not.
3. Find highly acclaimed novels that treat issues related to racial discrimination.

Task Group 2: The Data Gathering Tasks
You are in a data gathering stage of an assignment and need to collect a series of books for further analysis. This has led to one of the following three research needs:
4. Find novels that won the Nobel Prize during the 1990's.
5. Find bestseller crime novels by female authors.
6. Find biographies on athletes active in the 1990's.

The Semi self-selected Task
7. Try to find books about a specific topic or of a certain type, but do not look for a specific title you already know.

For each of the tasks, the participants were asked to assess the relevance of any book they viewed during the process. In addition they were asked to place books they would have bought in a "shopping basket". All sessions were logged by the IR system, which collected time stamped data about items viewed, relevance judgments, books added to the basket, book reviews viewed etc.

Prior to and after the experiment and before and after each task the participants answered questionnaires. The pre-experiment questionnaire included demographic questions as well as questions concerning participants' experience in information searching in general and in searching and using digital bookstores. Pre-task questions concerned, among other things, the participants' topic familiarity, i.e. *topic knowledge* and post-task questions included their evaluation of system features.

 In order to answer our research questions we have used the statistical analysis tool SPSS.

4. RESULTS
 Data from a total of 147 sessions performed by the 49 test subjects were collected from October 2010 to January 2011. The distribution of tasks performed is found in Table 1.

We see that very few of our test subjects chose Task 3 (novels treating the topic of racial discrimination), but the other tasks were fairly evenly distributed.

Table 1 Distribution of tasks between subjects

		Task							
		1	2	3	4	5	6	7	Total
Interface	A	14	10	2	6	8	11	26	77
	B	11	10	2	13	5	6	23	70
Total		25	20	4	19	13	17	49	147

4.1 Work task stage influence on user behavior
In order to answer research question 1 " How does work task stage influence the number of document surrogates used and assessed as relevant?" a number of statistical measures were applied. Table 2 contains the mean number of items visited for each task group. "Book views" represents the number of times a participant looked at the bibliographic details of an item (book) from the result list. The participant could further choose to click on a tab in the bibliographic view to retrieve a list of all reviews of the book, this contained the title of the review and the score (in stars) given by the reviewer, in Table 2 we call this "Review list". Finally the participant could also choose to retrieve full reviews from the list, this is simply called "Reviews" in Table 2.

Table 2 Items viewed per task group

Task group	Book views	Review lists	Reviews

1 Explore	17.5	4.0	5.1
2 Gathering	21.4	3.5	4.3

The findings indicate that our test subjects are more interested in looking at formal metadata about books at the data gathering stage, whereas they inspect more "third party" metadata, i.e. book reviews, at the early exploration stage. An independent t test shows, however, that the differences are not significant.

We also calculated mean values for relevance assessments for the two task groups. First we merged the "relevant" and "partially relevant" assessments to compare these with the non-relevant ones. In Table 3 we find the mean number of relevant (incl. partially relevant) and non-relevant items assessed by the test subjects. In addition we report the number of books added to the basket, indicating that the searcher would consider it an answer to the task.

Table 3 Relevance assessed per task group

Task group	Relevant	Not-relevant	Basket
1 Explore	7.5	0.65	6,4
2 Gathering	4.5	1.2	5,1

Our results show that the test subjects assessed more books as relevant at the simulated early stage of the information seeking process than at a simulated later stage (this is significant at p<0.05). We see that the number of books in the basket also differs, but not significantly. The findings are interesting when contrasted with the mean number of books viewed per task group, as they indicate that at an early stage of the work task process searchers view fewer books, but consider a larger number of them as relevant. One reason for this could be that the searchers at an early stage of a project needs to get a comprehensive overview of the topic area and plan to do so by reading a large share of the document (s)he finds.

If we look in more detail at the relevance assessments, and compare relevant and partially relevant judgments separately, we get the results showed in Table 4.

Table 4 Comparing assessments of Relevant and Partially relevant per task group

Task group	Relevant	Partially relevant
1 Explore	4.75	2.7
2 Gathering	3.2	1.3

We see that during the exploration stage, books are assessed more often both as relevant and as partially relevant, compared to during the data gathering stage. The difference is, however, only significant for the partially relevant assessments (p<0.005). This indicates that searchers at an early stage of the process tend to be more uncertain with respect to assessing if a book is relevant or not.

4.2 Topic knowledge influence on user behavior

Research question 2 was formulated: "How does topic knowledge influence the number of document surrogates used and assessed as relevant?"

Prior to each search session, the searchers were asked to indicate on a five point scale how familiar they were with the topic of the task they had chosen. In the questionnaire, 1 was indicated to mean "not at all", 5 "extremely" and 3 "somewhat" familiar. For the following discussion those who indicated 4 or 5 on the self-assessment scale were categorized as having high topic knowledge, a score of 3 indicated medium topic knowledge, the rest were categorized as having low topic knowledge. In this analysis the self-selected tasks were also included. We have performed statistical analysis to compare the different topic knowledge categories with respect to their use of different items and their relevance assessments. Table 5 shows the distribution of participants in the topic knowledge categories.

Table 5 Distribution of topic knowledge

Topic knowledge	No. of test subjects
Low	62
Medium	54
High	33

We started by comparing the groups with respect to their use of different items (Table 6). We only compared the high and low topic knowledge-groups.

Table 6 Items viewed per topic knowledge group

Topic knowledge	Book views	Review lists	Reviews
Low	20.6	3.6	4.8
High	15.3	4.5	5.8

We see that the test subjects with little knowledge about the topic prefer to look at significantly more books (p<0.05) than the members of the high topic knowledge-group. On the other hand, high topic knowledge-subjects seem to look at more lists of reviews and full reviews, but the difference is not significant.

Next we compared the relevance assessments made by the two groups (Table 7).

Table 7 Relevance assessed per topic knowledge group

Topic knowledge	Relevant	Not-relevant	Basket
Low	5	0.9	5.3
High	7.6	0.3	6,9

The data reported in Table 7 tell a rather obvious story, the

table shows that searchers with high topic knowledge judge more documents as relevant compared to low topic knowledge-searchers (significant at $p<0.05$), who, in return, assess more documents as non-relevant (significant at $p<0.05$). We also see that high topic knowledge also leads participants to put significantly more books in the basket ($p<0.05$). In other words we have clear indications that high topic knowledge makes it easier for searchers to find usable books and discard books that are not relevant for the task. Perhaps the most surprising finding is that the participants with low topic knowledge added more books to their baskets than they had found relevant. This is perhaps a result of the insecurity felt by this group of participants.

We have also divided the "relevant" assessments into fully relevant and partially relevant. The results can be found in Table 8.

Table 8 Comparing assessments of Relevant and Partially relevant per topic knowledge group

Topic knowledge	Relevant	Partially relevant
Low	3.1	1.9
High	5.2	2.4

The test subjects with high topic knowledge assessed more documents as fully relevant, compared to the persons with little topic knowledge ($p<0.05$), whereas no significant difference was found for partially relevant assessments.

5. DISCUSSION AND CONCLUSIONS

We have investigated the effect of work task stage and topic knowledge among a group of test subjects. The major weakness of the study is that tasks were imposed on the participants. We know from studies of real-life information behavior that the problems or tasks can be very complex and that search task formulation, information gathering, interpretation of results and search task reformulation often takes place over a long time. The design of the tasks in this study for these reasons cannot fully reflect stages in work task processes; thus the two task categories must be considered very simplified representatives of two different work task stages. It may even be argued that they could rather be seen as representing two task types. To get more representative data it would be necessary to perform studies of searchers engaged in their own tasks at different stages [19, 23].

Also, since our data were gathered from a rather homogenous groups of searchers (students within LIS and computer science) our findings are not generalizable. We have, however, examined relationships between factors that have not been investigated earlier. This has been made possible through the design of the INEX interactive track experiment, where transaction logs facilitates the identification of interaction with different parts of a bibliographic record.

The experiment used two slightly different interfaces to the system. The versions did, however, not differ with respect to the metadata items available for consultation and interaction. We therefore do not think the interface design had any effect on the participants' access to the data, their use of reviews or on their relevance judgments.

On the other hand, the system was previously unknown to all participants. Even if a training session was provided before the start of the experiment proper, the participants may have understood both system functionality and the nature of their experimental task in different ways. Individual differences in search style may also have influenced the use of the system features and the frequency of clicks (book views) on items in the result list.

The distinction between traditional metadata and book reviews as "content clues" can be used to learn how searchers trust different content producers. The data set showed some difference in the types of metadata used, but a larger scale-experiment is probably necessary in order to clarify whether work task stage or topic knowledge have any effect on the use of book reviews or not.

We have seen that the participants behaved significantly differently dependent on the task stage they were in. Our findings indicate that the searchers at an early stage judge more books as relevant, and that they also tend to be more uncertain with respect to assessing if a book is relevant or not. Kuhlthau found that searchers are typically more uncertain at an early stage than later on in the work task process, and our findings are compatible with this. The results from our analysis also complements the findings of Taylor, Cool, Belkin and Amadio (2007) who looked at the criteria used to assess relevance at different stages of the process, but said nothing about the effects this had on the number of relevant documents found. This relationship could be the purpose of further research.

We have also seen that searchers with little topic knowledge look at significantly more books than more experienced searchers, but the latter group still finds a higher number of books that they judge as relevant. This can also be explained by the uncertainty factor, if we presume that lack of knowledge makes searchers uncertain. The fact that we observed participants with low topic knowledge adding more books than they judged relevant is yet another indicator of uncertainty. Since searchers with high topic knowledge find more books they judge as relevant, it also indicates that they exploit their expertise in becoming more efficient searchers.

With respect to the effects the findings could have on IR systems design, it would be reasonable to consider how systems could be designed to adapt to the task stage of searchers by offering different kind of metadata at different stages.

This study calls for further qualitative investigations. It

would be of interest to study the kinds of sources seen and judged at the different stages. It would also be interesting to consider the effect of possible identifiable search styles on the decision process from a book view, via use of additional metadata, to a decision on relevance. This will increase our understanding of the use of quantifiable measures of search activity as means of predicting and interpreting information needs.

6. ACKNOWLEDGMENTS

7. REFERENCES

1. Borlund, P. *Evaluation of interactive information retrieval systems*. Abo Akademis Forlag, 2000.
2. Case, D. *Looking for Information, Second Edition: A Survey of Research on Information Seeking, Needs, and Behavior*. Emerald Group Publishing Limited, 2006.
3. Ellis, D., Cox, D., and Hall, K. A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation 49*, 4 (1993), 356-369.
4. Ellis, D. and Haugan, M. Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation 53*, 4 (1997), 384-403.
5. Ellis, D. A behavioural approach to information retrieval system design. *Journal of Documentation 45*, 3 (1989), 171-212.
6. Hölscher, C. and Strube, G. Searching on the Web (poster abstract). *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '99*, (1999), 305-306.
7. Hölscher, C. and Strube, G. Web search behavior of Internet experts and newbies. *Computer Networks 33*, 1-6 (2000), 337-346.
8. Ingwersen, P. and Järvelin, K. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, 2005.
9. Järvelin, K. and Kekäläinen, J. IR evaluation methods for retrieving highly relevant documents. *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '00*, (2000), 41-48.
10. Kuhlthau, C.C. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science 42*, 5 (1991), 361-371.
11. Kuhlthau, C.C. *Seeking Meaning*. Libraries Unlimited, 2003.
12. Malik, S., Tombros, A., and Larsen, B. The Interactive Track at INEX 2006. In N. Fuhr, M. Lalmas and A. Trotman, eds., *Comparative Evaluation of XML Information Retrieval Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, 387-399.
13. Marchionini, G., Dwiggins, S., Katz, A., and Lin, X. Information seeking in full-text end-user-oriented search systems: The roles of domain and search expertise. *Library & information science research 15*, 1 (1993), 35-69.
14. Marchionini, G., Lin, X., and Dwiggins, S. Effects of search and subject expertise on information seeking in a hypertext environment. In *Proceedings of the 53rd Annual Meeting of the American Society for Information Science*. ASIS, Silver Springs, MD, 1990, 129-142.
15. Marchionini, G. Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science 40*, (1989), 54-66.
16. Marchionini, G. *Information Seeking in Electronic Environments*. Cambridge University Press, 1995.
17. Mizzaro, S. Relevance: The whole history. *Journal of the American Society for Information Science 48*, 9 (1997), 810-832.
18. Pharo, N., Beckers, T., Nordlie, R., and Fuhr, N. Overview of the INEX 2010 Interactive Track. In S. Geva, J. Kamps, R. Schenkel and A. Trotman, eds., *Comparative Evaluation of Focused Retrieval*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, 227-235.
19. Pharo, N. and Järvelin, K. The SST method: a tool for analysing Web information search processes. *Information Processing & Management 40*, 4 (2004), 633-654.
20. Pharo, N. and Nordlie, R. Context Matters: An Analysis of Assessments of XML Documents. In F. Crestani and I. Ruthven, eds., *Context: Nature, Impact, and Role*. Springer Berlin / Heidelberg, 2005, 1911-1912.
21. Robertson, S.E. and Hancock-Beaulieu, M.M. On the evaluation of IR systems. *Information Processing & Management 28*, 4 (1992), 457-466.
22. Saracevic, T. Relevance reconsidered '96. In *CoLIS 2 : second international conference on conceptions of library science : integration in perspective : (Copenhagen, October 13-16, 1996)*. Copenhagen, 1996, 201-218.
23. Serola, S. and Vakkari, P. The anticipated and assessed contribution of information types in references retrieved for preparing a research proposal. *Journal of the American Society for Information Science and Technology 56*, 4 (2005), 373-381.
24. Taylor, A., Cool, C., Belkin, N., and Amadio, W. Relationships between categories of relevance criteria and stage in task completion. *Information Processing & Management 43*, 4 (2007), 1071-1084.
25. Vakkari, P. A theory of the task-based information

retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation 57*, 1 (2001), 44-60.

26. White, R.W., Dumais, S.T., and Teevan, J. Characterizing the influence of domain expertise on web search behavior. *Proceedings of the Second ACM International Conference on Web Search and Data Mining - WSDM '09*, (2009), 132.

27. Wildemuth, B.M. The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology 55*, 3 (2004), 246-258.

28. Wilson, T.D. Models in information behaviour research. *Journal of Documentation 55*, 3 (1999), 249-270.