A Qualitative Look at Eye-tracking for Implicit Relevance Feedback

Kirsten Kirkegaard Moe, Jeanette M. Jensen and Birger Larsen

Dept. of Information Studies, Royal School of Library and Information Science
Copenhagen, Denmark
kirstenkj@adr.dk, jeanettemjensen@gmail.com, blar@db.dk

Abstract. Our goal in this study was to explore the potentials of extracting features from eye-tracking data that have the potential to improve performance in implicit relevance feedback. We view this type of data as an example of the searcher' immediate context and as containing useful clues of the indications of the interaction between the searcher and the IR system. In particular, we explored if we could qualitatively identify features have potential to improve performance in implicit relevance feedback, and how such features correlate with document elements assessed as relevant or non-relevant. The results point to so-called thorough reading as one of the most promising features for identifying relevant information as input for implicit relevance feedback – in particular when it is related to the total time the searcher has looked an element.

Keywords. Eye-tracking ; Implicit Relevance Feedback ; Thorough reading ; INEX Interactive Track

1 Introduction

The core Information Retrieval (IR) techniques have reached a level of high maturity and do quite a good job of matching document content to a user given query. This is apparent from the widespread use of these techniques in Internet search engines and other search environments. As the core matching techniques have perhaps reached a plateau in terms of performance there has been an increased interest in exploiting the *context* of IR systems more fully to improve search performance[1]. The expectation is

[1] Contextual topics has also appeared at several IR related workshops and conferences recently, e.g., SIGIR 2004 & 2005 workshops on Information Retrieval in Context, the 2006 Information Interaction in Context Symposium, and the Context-Based Information Retrieval workshops at the recent CONTEXT conferences.

that better retrieval performance can be achieved by taking advantage of the context surrounding the IR systems, e.g., from documents or searchers [e.g., 1].

One well-known technique in IR for exploiting the immediate user context is relevance feedback [2], where relevance assessments provided by searchers can be used to modify subsequent queries, e.g., by increasing the weights of query terms found predominantly in relevant documents and decreasing the weights of terms mostly found in irrelevant ones. *Explicit* feedback techniques, based on the active marking of relevant documents, have been studied in some detail, e.g., [3, 4], and generally show that performance gains can be achieved. However, earlier studies have shown that it may be difficult to get explicit relevance assessments from searchers, as the active marking of relevant documents is not a part of the natural workflow in search systems [5, 6]. As a consequence, *implicit* relevance feedback where the feedback data are obtained indirectly from searchers' natural interaction with the system have received increased attention recently. Examples of such contextual behavioural data include: the amount of time that searchers have a document open [7], whether the document is printed [8], or saved [9].

In this paper we work with a type of contextual data that has so far not received much attention for implicit relevance feedback: eye-tracking data of how searchers look at search results. Outside IR research Human Computer Interaction studies have shown that eye-movements can be correlated to human's perception of relevance of read text [e.g., 10, 11]. The studies have, however, been carried out in very controlled (and thus rather unrealistic) environments with quite restricted and simple tasks that are very far from the complexity of realistic information seeking. For instance, in [11] the test persons were asked to identify an answer to a given question from 12 news headlines, or in [10] from 10 sentences. In this paper we investigate if eye-tracking data gathered from a less controlled, interactive IR experiment has potential value as source for implicit relevance feedback. We use a setting where the test persons have a choice of tasks, use a search system similar to an Internet search engine and are free to search and examine any documents in the collection as they wish (see Section 2 for details). In contrast to other studies [e.g., 12] we have chosen to take a qualitative and exploratory approach to identifying potentially useful features from the eye-tracking data rather than an algorithmic one. Apart from not having resources to implement the algorithmic approaches, the main reason is that we wish to study the potentials of eye-tracking data for implicit relevance feedback regardless of whether current algorithmic approaches can identify the observed features. The scope of the study is preliminary and the purpose is to attempt to identify promising features that can be tested empirically in future work. If good performance is obtained with certain features it can then be attempted to implement these algorithmically.

The overall goal of this study is to explore the potential of extracting features from eye-tracking data, regarded here as a type of contextual data, that can improve performance in implicit relevance feedback. Our research questions are:

- Is it possible by qualitative inspection to consistently identify features that have good potential for improving implicit relevance feedback performance from eye-tracking data of an interactive IR experiment?
- How do such features correlate with document elements that have been explicitly assessed as either relevant or non-relevant?

The paper is structured as follows: section 2 gives details of the experimental setting and the methods used in the analysis, section 3 presents the results, and section 4 concludes with a discussion.

2 Experimental setting

The study was carried out as a part of our research group's participation in the INEX2006 Interactive Track experiments [see 13 for more information]. INEX is the INitiative for the Evaluation of XML retrieval which studies the potential of providing more focussed retrieval results (i.e. document *elements*) to searchers by exploiting document structure, e.g., in XML documents [14]. This is mainly done by constructing laboratory test collections. The purpose of the INEX interactive track is to more broadly investigate the behaviour of users when interacting with elements of XML documents, and to investigate and develop approaches for XML retrieval which are effective in user-based environments [13].

In the INEX2006 interactive track the following test material was provided: an element retrieval system backend[2] containing a corpus more than 600,000 XMLified documents from the English version of Wikipedia [15], a prototype element retrieval interface including detailed transaction logging, 12 search tasks, questionnaires and experimental protocols [See 13 for more information]. The test persons acting as searchers were asked to search six of the 12 search tasks (they were given the tasks in pairs and could choose one of them), and were given up to 10 minutes to search for as much relevant information as possible to solve each task. The prototype element retrieval interface (a version of the Daffodil system adapted for element retrieval[3]) displayed the retrieved elements grouped by document, and allowed the searchers to access any full text part of the documents. Searchers could, e.g., access a section directly from the ranked hit list and navigate within the document using an automatically generated table of contents. Searchers were asked to provide explicit relevance assessment of any elements viewed, but were not forced to do so by the system as this might affect their natural interaction behaviour [22]. Assessments could be given using one of five categories [13]:

- **Relevant answer (RA) –** contains highly relevant information, and is just right in size to be understandable
- **Relevant, but too broad (TB) –** contains relevant information, but also a substantial amount of other information

[2] Both an element retrieval backend and a passage retrieval backend were made available. In the present paper we only analyse the tasks searched in the element retrieval backend due to technical problems with the passage backend.

[3] See [16] and http://www.daffodil.de/ for more information on Daffodil.

- **Relevant, but too narrow (TN)** – contains relevant information, but needs more context to be understood
- **Partial answer (PA)** – has enough context to be understandable, but contains only partially relevant information
- **Not relevant (NR)** – does not contain any information that is useful in solving the task

The 'too broad' and 'too narrow' categories are useful when experimenting with elements retrieval systems because they allow searchers to express that a result has some value but an inappropriate granularity.

INEX and its interactive track are particularly interesting in relation to our study: the retrieved and assessed units consisted of parts of documents. This is appropriate in relation to implicit relevance feedback and the eye-tracking data we use as we would typically be able to study patterns of gazing at the level of parts of documents rather than entire documents. In addition, compared to the experimental settings of earlier studies of perception of relevant text in, e.g., [10] or [11] the IENX interactive track is much less controlled and closer to a realistic search situation: the wikipedia corpus is of a general nature that a broad group of searchers would be able to relate to, the tasks were designed to fit the corpus and to be of general interest [20], the test persons could choose between several tasks and were free to interact with the system as they wished – querying, viewing, navigating and assessing any documents or elements from the ranked list, and to stop when they wished. Of course this is still an artificial experimental setting because it was not the searchers' own, real tasks and because the experiment took place in a controlled environment due to the need to use the eye-tracker. In particular, the time limit restriction of only 10 minutes per task is a factor that may affect the results.

In the present paper we analyse data collected from six searchers. In addition to the standard data collected in the interactive track we also collected eye-tracking recordings of all tasks being searched. The Tobii 1750 eye-tracker used provides a large amount of data types, including tracking of the searcher's gaze coordinates recorded at pixel level 50 times per second on both eyes, video screen capture, web cam recording of the test person and logging of keystrokes. As argued above, we have chosen to identify features qualitatively rather than attempting to find useful patterns algorithmically using, e.g., the gaze coordinates as done in some other studies. For our analysis we used a gaze replay visualisation, where the gaze data are overlaid on the video screen capture in real time (See figure 1): A dot shows the current focus on the screen and the trailing line after the dot shows the previous focus. The gaze replay allows us to qualitatively explore any patterns in the searchers' focus on the screen. The interpretation is aided by the web cam recording of the test person ad the tracking status window (Fig. 1).

We limited the analysis to the elements assessed as either Relevant (RA) or Not Relevant (NR). By focussing on the extreme poles of the relevance assessments we hope to get clearer indications that could tell us if some of the observed features can be correlated to relevant and irrelevant elements respectively.

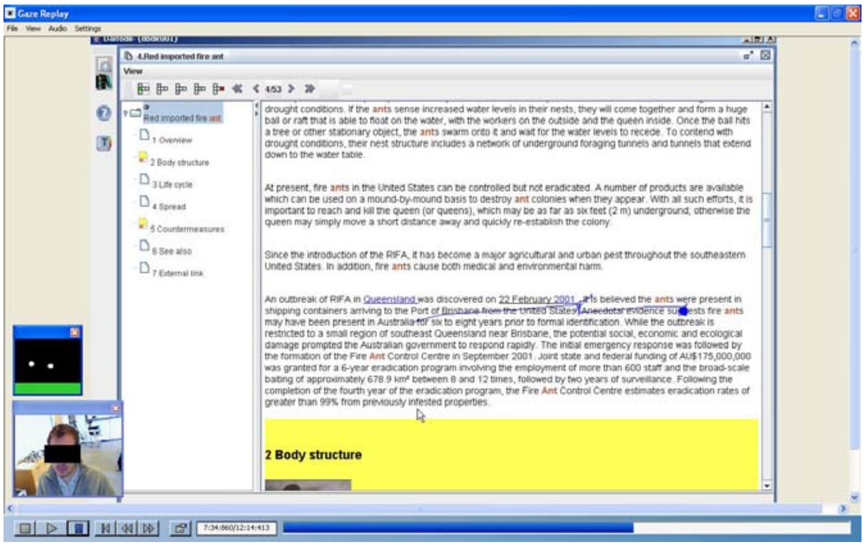

Fig. 1. The gaze replay visualization used for the qualitative identification of features. The blue dot shows the current focus and the trailing line the foci immediately preceding this. A tracker status window with information about eyes and web cam recording of the test person is included to the left.

3 Results

3.1 Overall browsing behaviour

The six searchers completed a total of 18 search tasks in the elements retrieval system. Due to technical problems only 15 of these could be analysed. In these 15 sessions a total of 201 elements were interacted with, and the searchers provided 97 explicit relevance assessments. Searchers were instructed to assess all viewed elements, and this rather low share of assessed elements supports earlier results that it may be difficult to obtain explicit relevance assessments from searchers [5, 6]. The distribution of assessments can be seen in Table 1. In the following only the 33 elements assessed as relevant and the 36 as irrelevant are analysed. In one of the tasks a searcher did not use these categories, which leaves 15 tasks for the analysis.

Table 1. Distribution of the 101 assessed elements.

Assessment	Frequency
Relevant answer (RA)	33
Relevant, but too broad (TB)	17
Relevant, but too narrow (TN)	2
Partial answer (PN)	9
Not relevant (NR)	36

3.2 Identified eye-tracking features

Inspired by existing studies and after an initial screening, where we observed the gaze replay visualisation (Fig. 1) for any gaze behaviour that could be correlated with relevance or non-relevance we choose to focus on three features observed in the gaze replay:

- **Total viewing time**, defined as the total time spent looking at an element relative to the length of the document
- **Thorough reading**, defined as mainly horizontal eye movements, with many fixations per line relative to the number of words on the line and at least half a line read
- **Regressions**, defined as the number of times where searchers regress, i.e., turn back to, an already seen element.

Total viewing time is interesting because some earlier studies have found indications that searchers spend longer time on relevant documents compared to irrelevant documents [e.g., 17, 18]. Kelly & Belkin did, on the other hand, not find any relation between display time and the usefulness of documents [7]. These studies were, however, based on transaction log analysis without the use of eye-tracking, where it was not possible to study if searchers actually did read (or at least looked at) the document content. In our analysis we only include the amount of time actually spent looking at an element. We normalise this with the element length measured in number of lines as it would intuitively take longer time to read a large element than a short and vice versa.

Thorough reading is a central notion because we would expect that any information that has been read, rather than just skimmed or glanced over rapidly, could be useful for implicit relevance feedback. A number of HCI studies have shown that it is possible to differentiate between relevant and irrelevant sentences in text by using eye-tracking [e.g., 11, 19]. Our definition, and our data collection, can be said to be qualitative in that the exact number of fixations in relation to the number of words on the line is not counted – rather it is interpreted qualitatively. We did do some comparisons of inter-coder consistency and found that two coders would agree in the vast majority of cases.

Pfeiffer, Saffari & Juffinger also found that test persons made more *regressions* back to relevant sentences [19]. These studies were carried out in rather restricted settings and with very narrow and simple tasks. In the present paper we use a much more open setting and study if these features can aid also in identifying relevant information in a less controlled information seeking situation with more complex documents and more open tasks.

A number of additional features were considered: skim-reading, skimming and orientation/navigation. It turned out that it was very hard to differentiate between these qualitatively and to define them consistently because they had very short durations. We also attempted to identify thorough reading behaviour from the number of fixations and the duration of these directly from gaze coordinate data, but did not succeed using simple measures.

The results for the three selected features are summarised below.

3.3 Regressions

We counted regressions made back to elements assessed as relevant and irrelevant. As the searchers gaze skip rapidly over the screen when skimming and navigating the documents we counted only those regressions where the searchers had left the element more than one second and the return to look at the element for more than one second afterwards.

A total of 70 regressions were made to the 33 elements assessed as relevant (2.1 per element) in the 15 tasks, and 42 regressions to the 36 irrelevant elements (1.2 per element). That is, we find almost twice as many regressions to relevant elements. However, this is only one more regression per element for relevant compared to irrelevant, and the distribution over searchers is heavily skewed (22 of the elements assessed as irrelevant were given by one searcher). Looking closer at the data no clear pattern emerges from the regressions and based on the present data there is no clear indication that regressions can be exploited to identify relevant elements.

3.4 Total viewing time

We calculated the total viewing time that searchers looked at elements assessed as relevant and irrelevant respectively. Any gazing for more than one second is included, and averaged over the elements. The total viewing time for relevant elements was 43.8 seconds for relevant elements and 9.1 seconds for irrelevant elements, that is, noticeably longer in relevant elements. When we normalise for element length, that is the number of lines in the element, the normalised total viewing time is 3.5 seconds per line for relevant elements on average and 1.1 seconds for irrelevant. Again the distribution is heavily skewed. 5 out of 6 searchers did spend much more time in relevant elements, but one searcher with a large amount of elements assessed as irrelevant spent slightly more time in these. Thus there is a tendency for searchers to spend more time in relevant elements, but it is not unambiguous.

3.5 Thorough reading

In the analysis of thorough reading we have calculated how large a part of the total reading time in seconds that was taken up by thorough reading. On average, the searchers read thoroughly 69 % of the time they spent in relevant elements, and 28 % of the time they spent in irrelevant elements. That is, about 2.5 as much time was spent reading thoroughly in relevant elements compared to irrelevant elements. This varies across test persons, but there is an unambiguous trend that they all spent more time reading thoroughly in relevant elements.

To further analyse this result we have related thorough reading to a contextual task variable given by the INEX interactive track. Each of the 12 search tasks were constructed to be one of three task types (*Decision making*; *Fact finding*; *Information Gathering*) [see 20 for details]. The distribution of thorough reading in relation total viewing time across the three task types can be seen in Figure 2 (note that the percentage can reach 100 for both relevant and irrelevant). In all three cases there is a clear tendency for searchers to spend a larger share of the time reading thoroughly in relevant compared to irrelevant elements. Comparing the tasks types this trend is strongest in *Fact finding* tasks with 77 % thorough reading in relevant elements, and only 15 % in irrelevant. Information gathering also has a large share of thorough reading in relevant elements, 70 % versus 34 % in irrelevant. For these two task types it seems that irrelevant information can be identified fairly easy as elements without much thorough reading, whereas information that searchers need to read more thoroughly tends to be judged relevant. For the Decision making tasks the trend is not so strong with 60 % thorough reading in relevant and 38 % in irrelevant. This is perhaps to be expected as searchers would have had to weight up several alternatives in the process of making decisions.

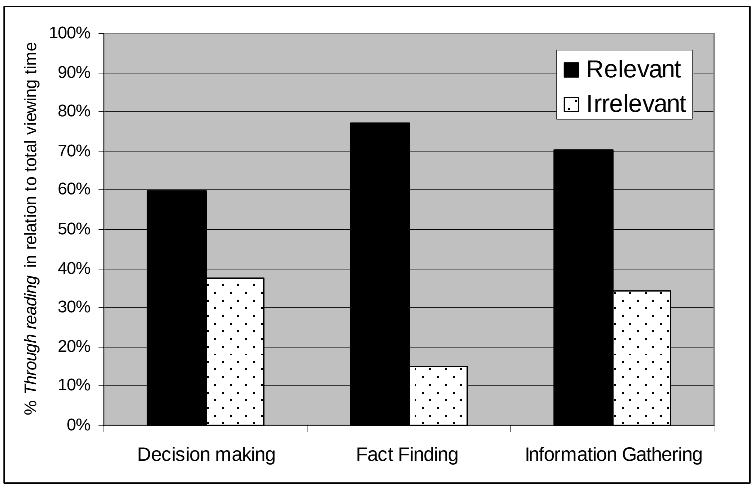

Fig. 2. Share of *Thorough reading* in relation to *Total viewing time* across task types.

4. Discussion

Inspired by features reported in HCI-studies we were able to identify some features by inspecting the eye-tracking gaze replays qualitatively. The features that we could consistently identify were those that were not of very short durations, e.g., less than a second or involving less than half a line of text. Behaviour involving shorter gazing or smaller pieces of text, such as skimming or navigation, proved hard to observe consistently using qualitative inspection.

The feature *thorough reading* was the most promising for identifying relevant information across the six searchers and across task types. Intuitively, through reading corresponds to the notion of having read text as opposed to just skimmed it or glanced over it, and intuitively this would be a good candidate for identifying relevant text.

Our way of operationalising through reading based on qualitative inspection is to the best of our knowledge novel.

However, as some thorough reading did occur also in elements assessed as irrelevant it may be necessary to filter out elements that are read thoroughly, but have a high risk of being irrelevant; including too much irrelevant information in the implicit relevance feedback may seriously harm performance. One way to filter out irrelevant elements could be to set a threshold on the percentage of thorough reading over total viewing time. By ranking the elements by share of thorough reading (not shown) we can observe that the data splits roughly in three bins: the top 33% contains almost only relevant elements, the bottom 33% almost only irrelevant elements and the middle 33% a mixture of both. The implication is that, by setting a threshold of 66% thorough reading out of total viewing time, almost only relevant elements would be included in the relevance feedback. In addition all elements where there was no thorough reading (i.e., corresponding to a threshold of 0 %) were assessed as irrelevant. These elements could thus be used as indications of irrelevance in implicit relevance feedback techniques.

A focus for future research could be to make an algorithmic implementation that can automatically identify thorough reading behaviour. Compared to the other, simpler features analysed in this study (regressions, total viewing time, the number of fixations and the duration of these) thorough reading is a composite concept where several conditions must be satisfied. Thorough reading will thus perhaps take a larger effort to implement, but the gains would also seem to be higher in terms of a better identification of relevant information for the implicit feedback. In addition, the implementation might draw on research already done on reading detection from eye-tracking [e.g., 21], and the output of existing algorithms can be compared to our more qualitative approach. It must be noted that in this study thorough reading has only been analysed in relation to the elements judged relevant or irrelevant. The clear trends shown by thorough reading may be blurred somewhat when thorough reading in elements with intermediate assessments (too broad, too narrow and partially relevant) as well as un-assessed elements are considered.

Exploiting the gaze behaviour of searchers in this manner is a way of bringing the immediate *context* of the user into the IR process: rather than just relying on the user's query to facilitate the match between information need and documents we can attempt to improve the quality of the interaction not only by explicit feedback, but also by implicit feedback from the searcher. The idea of exploiting eye-tracking data can be put in relation Ingwersen's interpretation of the cognitve viewpoint in IR [23-24] and his model of the cognitive communication system for information science [see e.g., 25, p. 33]: by relying on eye-tracking data we are getting indications of the *perception* that takes place as the searcher attempts to understand the information and put it into the context of her knowledge. If any improvement can indeed be obtained by exploiting such eye-tracking data, it may exactly be because they capture indications of the information processing that takes place in the searcher as she strives to make sense of the document. Admittedly, the current cost of an eye-tracker may prohibit this from being applied in any practical settings for some time yet. However, when cheaper eye-trackers (perhaps based on cameras in laptops or cell phones) become available we may begin to exploit this type of context, and the research results produced now, more widely.

5. Conclusions

Our goal in this study was to explore the potentials of extracting features from eye-tracking data that have the potential to improve performance in implicit relevance feedback. We view this type of data as an example of the searcher' immediate context and as containing useful clues of the interaction between the searcher and the IR system that might improve the quality of search results. In particular, we explored if we could qualitatively identify features have potential to improve performance in implicit relevance feedback, and how such features correlate with document elements assessed as relevant or non-relevant. The results indicate that the feature *through reading* have the potential to identify relevant information as input for implicit relevance feedback – especially when it is related to the total time the searcher has looked an element. Theoretically the use of eye-tracking data as contextual clues can be related to the cognitive viewpoint as put forward by Ingwersen [23-24].

Because of the limited size of the study (6 searchers and 15 tasks) the results of the study are indicative only. The size is not only limited because of the available material, but also because of the chosen method: the qualitative identification of features from the eye-tracking data is time consuming and limits the number of search sessions that can be analysed. Nonetheless this explorative approach has allowed us to investigate the value of a number of features without first having to implement algorithms to automatically identify these features. Future research can then focus on those features that show most potential.

We are now working with an extended dataset with 12 searchers, where we have extracted terms from the documents based on total viewing time and thorough reading. The extracted terms will be used in implicit relevance feedback experiments and the performance compared to explicit relevance feedback based on judged elements. The initial results indicate that the implicit relevance feedback generally performs as well as explicit feedback.

Acknowledgments

We wish to thank the test persons who gave their time to participate in the study. We also thank the anonymous reviewers for constructive feedback.

References

[1] Ingwersen, P. and Järvelin, K. (2004): Information retrieval in contexts. In: Ingwersen, P., van Rijsbergen, K. and Belkin, N. eds. *ACM SIGIR 2004 Workshop on "Information Retrieval in Context"*. Sheffield: [University of Sheffield], p. 6-9. [http://ir.dcs.gla.ac.uk/context/, visited 1-5-2006]

[2] Rocchio, J.J. (1971). Relevance Feedback in Information Retrival. In: Salton, G. (red.), *The SMART Retrival System: Experiments in Automatic Document Processing.* Englewood Cliffs, New Jersey: Prentice Hall, 313-323.

[3] Eftimiadis, E.N. (1992): Interactive Query Expansion and Relevance Feedback for Document Retrieval Systems. London, UK: City University. (PhD thesis)
[4] Hawking, D., Thistlewaite, P, and Craswell, P. (1997): ANU/ACSsys TREC-6 experiments. In: Voorhees, E.M. and Harman, D.K. (eds) Proceedings of TREC-6, 275-290.
[5] Beaulieu, M. & Jones, S. (1998). Interactive Searching and Interface Issues in the Okapi Best Match Probabilistic Retrieval System. *Interacting with Computers*, 10, 237-248.
[6] Ruthven, I., Tombros, A. & Jose, J.M. (2001). A Study on the Use of Summaries and Summary-based Query Expansion for a Question-answering Task. In: Thiel, U. & Fuhr, N. (red.), *Proceedings of the 23rd BCS-IRSG European Colloquium on Information Retrieval Research*. Berlin: Springer, 1-14.
[7] Kelly, D. & Belkin, N. (2004). Display Time as Implicit Feedback: Understanding Task Effects. In: Järvelin, K., Allan, J., Bruza, P. & Sanderson, M. (red.), *Proceedings of Sheffield SIGIR 2004: The 27th Annual ACM SIGIR International Conference on Research and Development in Information Retrieval*. Sheffield, UK: ACM Press, 377-384.
[8] Kelly, D. (2005). Implicit Feedback: Using Behavior to Infer Relevance. In: Spink, A. & Cole, C. (red.), *New Directions in Cognitive Information Retrieval*. Dordrecht: Springer, 169-186.
[9] Joachims, T., Granka, L., Pan, B., Hembrooke, H. & Gay, G. (2005). Accurately Interpreting Clickthrough Data as Implicit Feedback. In: Marchionini, G., Moffat, A., Tait, J., Baeza-Yates, R. & Ziviani, N. (red.), *SIGIR 2005: Proceedings of the 28th Annual ACM International Conference on Research and Development in Information Retrieval*. New York: ACM Press, 154-161.
[10] Pfeiffer, M., Saffari A.A., A.R. & Juffinger, A. (2005). Predicting Text Relevance from Sequential Reading Behavior. In: Poulamäki, K. & Kaski, S. (red.). *Proceedings of the NIPS 2005 Workshop on Machine Learning for Implicit Feedback and User Modeling: Workshop at NIPS 2005, in Whistler, BC, Canada, on December 10, 2005.* Helsinki: Helsinki University of Technology, 25-30.
[11] Salojärvi, J., Kojo, I., Simola, J. & Kaski, S. (2003). Can Relevance be Inferred from Eye Movements in Information Retrieval? In: *Proceedings of WSOM `03, Workshop on Self-Organising Maps, September 2003*. Kitakyushu, Japan: Kyushu Institute of Technology, 261-266.
[12] Brooks, P. Phang, K.Y., Bradley, R., Oard, D., White, R. & Guimbretiere, F. (2006). *Measuring the Utility of Gaze Detection for Task Modeling: A Preliminary Study*. Paper presented at Workshop on Intelligent User Interfaces for Intelligence Analysis. Sydney, Australia.
[13] Larsen, B., Malik, S. & Tombros A. (2006). The Interactive Track at INEX 2006. In: Fuhr, N., Lalmas, M., & Trotman, A. (red.), *INEX 2006 Workshop Pre-Proceedings.* Dagstuhl, Germany, 250-261.
[14] Lalmas, M. and Kazai, G. (2006): Report on the ad-hoc track of the INEX 2005 workshop. *ACM Sigir Forum*, 40(1) (To appear)
[15] Denoyer, L. and Gallinari, P. (2006): The Wikipedia XML Corpus. *SIGIR Forum*, 40(1), 64-69.
[16] Malik, S., Klas, C.-P., Fuhr, N., Larsen, B. and Tombros, A. (2006): Designing a user interface for interactive retrieval of structured documents : lessons learned from the INEX interactive track? In: Gonzalo, J., Thanos, C., Verdejo, M. F. and Carrasco, R. C. eds. Research and Advanced Technology for Digital Libraries, 10th European Conference, ECDL 2006, Alicante, Spain, September 17-22, 2006, Proceedings. Berlin: Springer, (Lecture Notes in Computer Science ; 4172)

[17] Morita, M. & Shinoda, Y. (1994). Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. In: Croft, W.B. & van Rijsbergen, C.J. (red.), *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: Springer-Verlag, 272-281.

[18] Kim, J., Oard, D.W. & Romanik, K. (2000). *Using Implicit Feedback for User Modeling in Internet and Intranet Searching*. University of Maryland CLIS Technical Report 00-01.

[19] Puolamäki, K., Salojärvi, J., Savia, E., Simola, J. & Kaski, S. (2005). Combining Eye Movements and Collaborative Filtering for Proactive Information Retrieval. In: Marchionini, G. et al. (red.), *SIGIR 2005: Proceedings of the 28th Annual ACM International Conference on Research and Development in Information Retrieval.* New York: ACM Press, 146-153.

[20] Toms, E., Fruend, L. & Jordan, C. (2006). Evaluating Tasks by Type and Form. In: Fuhr, N., Lalmas, M., og Trotman, A. (red.), *INEX 2006 Workshop Pre-Proceedings.* Dagstuhl, Tyskland, 264-266.

[21] Campbell, C.S. & Maglio, P.P. (2001). A Robust Algorithm for Reading Detection. In: *Proceedings of the 2001 workshop on Perceptive user interfaces PUI '01*. New York: ACM Press, 1-7.

[22] Larsen, B., Tombros, A. and Malik, S. (2005): Obtrusiveness and relevance assessment in interactive XML IR experiments. In: Trotman, A., Lalmas, M. and Fuhr, N. eds. *Proceedings of the INEX 2005 Workshop on Element Retrieval Methodology, held at the University of Glasgow, 30 July 2005. Second Edition.* Dunedin (New Zealand): Department of Computer Science, University of Otago, p. 39-42.

[23] Ingwersen, P. (1992): *Information retrieval interaction*. London: Taylor Graham. x, 246 p. ISBN: 0-947568-54-9. [http://www.db.dk/pi/iri/, visited 05.07.2007]

[24] Ingwersen, P. (1996): Cognitive perspectives of information retrieval interaction : elements of a cognitive IR theory. *Journal of Documentation*, 52(1), 3-50.

[25] Ingwersen, P. and Järvelin, K. (2005): *The turn : integration of information seeking and retrieval in context*. Springer. xiv, 448 p. (The Information Retrieval Series ; 18) ISBN: 1-4020-3850-X.