Designing a User Interface for Interactive Retrieval of Structured Documents — Lessons Learned from the INEX Interactive Track*

Saadia Malik[1], Claus-Peter Klas[1], Norbert Fuhr[1],
Birger Larsen[2], and Anastasios Tombros[3]

[1] University of Duisburg-Essen, Duisburg, Germany
{malik, klas, fuhr}@is.informatik.uni-duisburg.de
[2] Royal School of Library & Information Science, Copenhagen, Denmark
blar@db.dk
[3] Queen Mary, University of London, United Kingdom
tassos@dcs.qmul.ac.uk

Abstract. The interactive track of the Initiative for the Evaluation of XML retrieval (INEX) aims at collecting empirical data about user interaction behaviour and to build methods and algorithms for supporting interactive retrieval in digital library systems containing structured documents. In this paper we discuss and compare the usability aspects of the web-based user interface used in 2004 with the application based user interface implemented with the Daffodil framework in 2005. The results include a validation of the element retrieval approach, successful implementation of the berrypicking model, and that additional clues for facilitating interactive retrieval (e.g. table of contents, indication of entry points, related terms, etc.) are appreciated by users.

1 Introduction

Many of today's DL systems still treat documents as atomic units, providing little support for searching or navigating along the logical structure of documents. With the steadily increasing use of the eXtensible Markup Language (XML), we have a widely adopted standard format for structured documents. Thus, there is now an opportunity for providing better support for structured documents in digital libraries (DLs). Besides supporting navigation, the logical structure of XML has the potential to assist the DL systems in providing more specific results to users by pointing to document elements rather than to whole documents.

Since 2002, the Initiative for the Evaluation of XML Retrieval (INEX) has organised annual evaluation campaigns for researchers in this field. However, little research has been carried out to study user behaviour and to investigate methods supporting interaction in the context of retrieval systems that take advantage of the additional features offered by XML documents.

* This work was funded by the DELOS Network of Excellence on Digital Libraries (EU 6. FP IST, G038-507618).

J. Gonzalo et al. (Eds.): ECDL 2006, LNCS 4172, pp. 291–302, 2006.

In order to address these issues, an interactive track (iTrack) was added to INEX in 2004. In this paper, we report on the usability issues addressed in the interactive XML retrieval systems that formed the baseline in these tracks in 2004 and 2005 (hereafter called iTrack 04 and iTrack 05). We show how the findings from the first year led to the development of an improved system in 2005, and we report on the user reactions to both systems.

In iTrack 04, the main goal was to study user behaviour with an XML retrieval system and to validate the element retrieval approach. For this, the user interface design was kept simple in order to give a clear picture of element retrieval systems. During iTrack 04 many usability issues arose, and these led to formulating the main hypotheses for iTrack 05. In addition, more elaborate design principles and the berrypicking paradigm [1] were followed for the iTrack 05 interface design.

This paper is structured as follows: Section 2 gives a brief overview of related work. Section 3 describes the evaluation methodology, the user interface and findings of iTrack 04. The description of the iTrack 05 user interface follows in section 4 including the necessary adaptions derived from iTrack 04, the evaluation and findings. The last section presents a comparison of both evaluations and an outlook.

2 Related Work

Classical information retrieval (IR) research has focused on a system-oriented view and taken a simplified view of user behaviour: the user submits a query and then looks through the ranked items one by one. Thus the goal of the system is to rank relevant items at the top of the list. A broader perspective has been taken in interactive IR research, as represented by the TREC interactive tracks [2]. Quite surprisingly, results of these evaluations showed that differences in system performance identified in laboratory experiments are hard to recreate in interactive retrieval. As described in [3], this result is due to users being able to easily identify relevant entries in a list of documents. Thus, cognitive factors should be considered, as well as richer interaction functions, that can enhance user interaction with the system.

Whereas the standard IR model assumes that the user's information need does not change throughout the search process, empirical studies (e.g. [4]) have shown that interactive retrieval consists of a sequence of related queries targeting different aspects of an ever changing information need. For coping with this problem, Bates et al. has proposed the berrypicking model of information seeking, which assumes that the user's need changes while looking at the retrieved documents, thus leading into new unanticipated directions [1]. During the search, users collect relevant items retrieved by different queries (berrypicking).

So far, there has been little work on interactive XML retrieval. Finesilver and Reid describe the setup of a small collection from Shakespeare's plays in XML, followed by a study of end user interaction with the collection [5]. Two interfaces were used: one highlighting the best entry points and the other highlighting the relevant objects.

Some recent efforts have been made within the INEX interactive track [6, 7]. In addition to the baseline systems which are the topic of this paper, Kamps et al. tested a web-based interface that used a hierarchal result presentation with summarisation and visualisation[8], and van Zwol, Spruit and Baas worked with graphical XML query formulation and different result presentation techniques also in a web-based interface [9]. Besides these systems, various techniques for visualisation of structured documents have been proposed in [10] and [11, 7].

3 iTrack 04

3.1 Evaluation Methodology

Document Corpus. The document corpus used was the 500 MB corpus of 12,107 articles from the IEEE Computer Society's journals covering articles from 1995-2002 [12].

Topics. We used content only (CO) topics that refer to document contents. In order to make the tasks comprehensible by other people besides the topic author, it was required to add why and in what context the information need had arisen. Thus the INEX topics are in effect simulated work task situations as developed by Borlund [13]. Four of the 2004 CO topics were used in the study.

Participating sites. The minimum requirement for sites to participate in the iTrack 04 was to provide runs using 8 searchers on the baseline version of the web-based XML retrieval system provided. 10 sites participated in this experiment, with 88 users altogether.

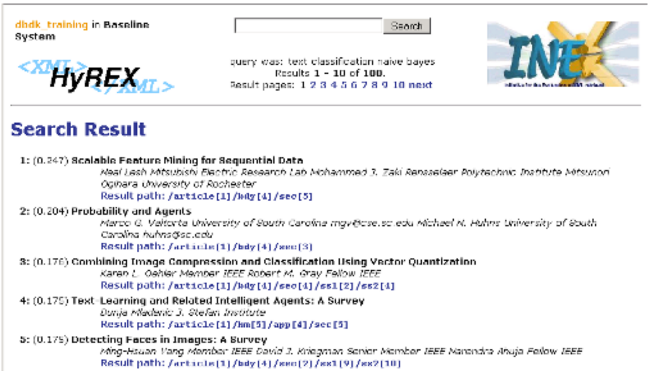
Experimental protocol & data collection. Each searcher worked on one task from each task category. The task was chosen by the searcher and the order of task categories was permuted. The goal for each searcher was to locate sufficient information towards completing a task, in a maximum timeframe of 30 minutes per task.

Searchers had to fill in questionnaires at various points in the study: before the start of the experiment, before each task, after each task, and at the end of the experiment. An informal interview and debriefing of the subjects concluded the experiment. The collected data comprised questionnaires completed by searchers, the logs of searcher interaction with the system, the notes experimenters kept during the sessions and the informal feedback provided by searchers at the end of the sessions.

3.2 User Interface

The user interface in iTrack 04 was a browser-based frontend connecting to the HyREX retrieval engine [14, 15].

In response to a user query, the system presented a ranked list of XML elements including title and author of the document in which the element occurred.



dbdk_training in Baseline System

query was: text classification naive bayes
Results 1 - 10 of 100.
Result pages: 1 2 3 4 5 6 7 8 9 10 next

Search Result

1: (0.247) Scalable Feature Mining for Sequential Data
Neal Lesh Mitsubishi Electric Research Lab Mohammed J. Zaki Rensselaer Polytechnic Institute Mitsunori Ogihara University of Rochester
Result path: /article[1]/bdy[4]/sec[5]

2: (0.204) Probability and Agents
Marco G. Valtorta University of South Carolina mgv@cse.sc.edu Michael N. Huhns University of South Carolina huhns@sc.edu
Result path: /article[1]/bdy[4]/sec[3]

3: (0.176) Combining Image Compression and Classification Using Vector Quantization
Karen L. Oehler Member IEEE Robert M. Gray Fellow IEEE
Result path: /article[1]/bdy[4]/sec[4]/ss1[2]/ss2[4]

4: (0.175) Text-Learning and Related Intelligent Agents: A Survey
Dunja Mladenic J. Stefan Institute
Result path: /article[1]/bm[5]/app[4]/sec[5]

5: (0.175) Detecting Faces in Images: A Survey
Ming-Hsuan Yang Member IEEE David J. Kriegman Senior Member IEEE Narendra Ahuja Fellow IEEE
Result path: /article[1]/bdy[4]/sec[2]/ss1[9]/ss2[10]

Fig. 1. iTrack 04: Query form and resultlist

In addition, a retrieval score expressing the similarity of the element to the query and the path to the element was shown in form of an result path expression (see Figure 1). The searcher could scroll through the resultlist and access element details by clicking on the result path. This would open a new window displaying this element.

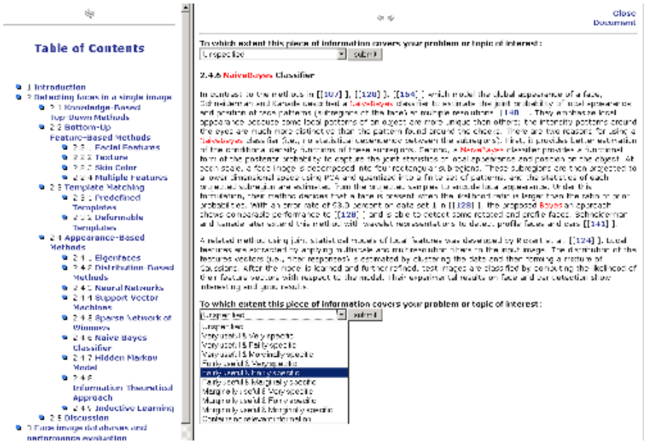


Table of Contents

- 1 Introduction
- 2 Detecting faces in a single image
  - 2.1 Knowledge-Based Top-Down Methods
  - 2.2 Bottom-Up Feature-Based Methods
    - 2.2.1 Facial Features
    - 2.2.2 Texture
    - 2.2.3 Skin Color
    - 2.2.4 Multiple Features
  - 2.3 Template Matching
    - 2.3.1 Predefined Templates
    - 2.3.2 Deformable Templates
  - 2.4 Appearance-Based Methods
    - 2.4.1 Eigenfaces
    - 2.4.2 Distribution-Based Methods
    - 2.4.3 Neural Networks
    - 2.4.4 Support Vector Machines
    - 2.4.5 Sparse Network of Winnows
    - 2.4.6 Naive Bayes Classifier
    - 2.4.7 Hidden Markov Model
    - 2.4.8 Information Theoretical Approach
    - 2.4.9 Inductive Learning
  - 2.5 Discussion
- 3 Face image databases and performance evaluation

To which extent this piece of information covers your problem or topic of interest:
Unspecified [submit]

2.4.6 NaiveBayes Classifier

In contrast to the methods in [[107]], [[128]], [[154]] which model the global appearance of a face, Schneiderman and Kanade described a NaiveBayes classifier to estimate the joint probability of local appearance and position of face patterns (subregions of the face) at multiple resolutions [[140]]. They emphasize local appearance because some local patterns of an object are more unique than others; the intensity patterns around the eyes are much more distinctive than the pattern found around the cheeks. There are two reasons for using a NaiveBayes classifier (i.e., no statistical dependency between the subregions). First, it provides better estimation of the conditional density functions of these subregions. Second, a NaiveBayes classifier provides a functional form of the posterior probability to capture the joint statistics of local appearance and position on the object. At each scale, a face image is decomposed into four rectangular subregions. These subregions are then projected to a lower dimensional space using PCA and quantized into a finite set of patterns, and the statistics of each projected subregion are estimated from the projected samples to encode local appearance. Under this formulation, their method decides that a face is present when the likelihood ratio is larger than the ratio of prior probabilities. With an error rate of 93.0 percent on data set 1 in [[128]], the proposed Bayesian approach shows comparable performance to [[128]] and is able to detect some rotated and profile faces. Schneiderman and Kanade later extend this method with wavelet representations to detect profile faces and cars [[141]].

A related method using joint statistical models of local features was developed by Rickert et al. [[124]]. Local features are extracted by applying multiscale and multiresolution filters to the input image. The distribution of the features vectors (i.e., filter responses) is estimated by clustering the data and then forming a mixture of Gaussians. After the model is learned and further refined, test images are classified by computing the likelihood of their feature vectors with respect to the model. Their experimental results on face and car detection show interesting and good results.

To which extent this piece of information covers your problem or topic of interest:
Unspecified [submit]

- Unspecified
- Very useful & Very specific
- Very useful & Fairly specific
- Very useful & Marginally specific
- Fairly useful & Very specific
- Fairly useful & Fairly specific
- Fairly useful & Marginally specific
- Marginally useful & Very specific
- Marginally useful & Fairly specific
- Marginally useful & Marginally specific
- Contains no relevant information

Fig. 2. iTrack 04: Detail view of an element

The detailed element view is depicted in Figure 2. The content of the selected element was presented on the right hand side. The left hand part of the view showed the table of contents (TOC) of the whole document. Searchers could access other elements within the same document, either by clicking on entries in the TOC, or by using the Next and Previous buttons (top of right hand part). A relevance assessment for each viewed element could be given on two dimensions of relevance: how useful and how specific the element was in relation to the task. These dimensions corresponded to the relevance dimensions of the main ad-hoc track of INEX in an attempt to ensure comparability of the results of the two tracks. Each dimension had three grades of relevance, and ten possible combinations of these dimensions could be given in a drop down list as shown in Figure 2.

3.3 Findings

The main findings based on the log and questionnaires are reported in [16]. Here, only the findings related to the usability of the baseline system are discussed. We analysed the questionnaire and interview data to investigate these issues. Most questionnaire questions were answered on a 5-point scale, which we have analysed statistically.

The overall opinion of the participants about the baseline system was recorded in the final questionnaire which they filled after the completion of both tasks. Users were asked to rate the different features of the system on the scale of 1 to 5, where 1 stood for 'Not at all', 3 'Somewhat' and 5 for 'Extremely'. The results are summarised in Table 3.

In addition to these ratings, users were asked to comment on the different aspects of the system after the completion of each task and after the completion of the experiment. Example questions were:

- *In what ways (if any) did you find the system interface useful in the task?*
- *In what ways (if any) did you find the system interface not useful in the task?*
- *What did you like about the search system? What did you dislike about the system?* and
- *Do you have any general comments?*

The analysis of the most frequent comments are presented in the following sections. Table 1 summarises the positive and Table 2 the negative results.

Element overlap. One of the critical issues of element retrieval is the possible retrieval of overlapping result elements, i.e. components from the same document where one includes the other (due to the hierarchic structure of XML documents). Typically these elements are shown at non-adjacent ranks in the hit list. In our case, the HyREX retrieval engine did not take care of overlapping elements and thus searchers frequently ended up accessing elements of the same document at different points in time and at different result ranks.

Data from both the system logs and the questionnaires showed that searchers found the presence of overlapping elements distracting. By recognising that they had accessed the same document already through a different retrieved element, searchers typically would return to the resultlist and access to another element instead of browsing again within a document visited before. 31 users commented negatively on the element overlap.

Document structure provides context. The presence of the logical structure of the documents alongside the contents of the accessed elements was a

Table 1. Positive responses on system usefulness (iTrack 04, 88 searchers)

System Features	Response Count
Table of contents	66
Keyword highlighting	36
Simple/easy	34
Good results	13
Fast	8
Simple querying	6

Table 2. Negative responses on system usefulness (iTrack 04, 88 searchers)

System Features	Response Count
Overlapping elements	31
Insufficient summary	30
Distinction b/w visited & unvisited	24
Limited query language	22
Poor results	10
Limited collection	9
Slow	9

feature that searchers commented positively on. The table of contents of each document (see Figure 2) seemed to provide sufficient context to searchers in order to decide on the usefulness of the document. 66 users found the TOC of the whole article very useful because it provided easy browsing, navigation, less scrolling or gave a quick overview of which elements might be relevant and which might not be.

Element summaries. The resultlist presentation in the iTrack 04 system did not include any element summarisation. Only the title and authors of the document were displayed in addition to the result path expression of the element and its similarity to the query. As a consequence searchers had little clues available to decide on the usefulness of retrieved elements at this point. 30 users commented on these insufficient clues.

Keyword highlighting. Within the detail presentation of an element, all query terms were highlighted. This feature was very much appreciated, and several users suggested to provide this feature not only at the resultlist level, but also at the table of contents level. 36 users gave positive comments on this feature.

Distinction between visited and unvisited elements. There was no distinction between visited and unvisited elements at the resultlist and detail levels. Thus, a number of times users visited the same elements/documents more than once. 24 users commented negatively on this.

Limited query language. The system did not support sophisticated queries and there was no possibility to use phrases, boolean queries, or to set the preference for terms. 22 users found this an obstacle.

General issues. There are also some more general issues that were commented on. These stated that the multiple windows of the web-interface were somewhat confusing and that the "Result path" shown in the resultlist was mostly meaningless, and with the square brackets, it had a very technical appearance.

iTrack 04 was the first attempt to set up an interactive track for XML retrieval, and there was very little knowledge on which we could build upon when designing the iTrack 04 interface. In contrast, the design of the iTrack 05 interface was based on the expereinces from the previous year. In designing the interface, we aimed at overcoming the main weaknesses of the 2004 interface.

4 iTrack 05

4.1 Evaluation Methodology

The evaluation methodology used in iTrack 05 was similar to the one used in iTrack 04. An extended version of the INEX IEEE document collection was used (now comprising 16819 documents).

This time six topics were selected from the INEX 2005 ad-hoc topics, and modified into simulated work tasks. In addition, searchers were asked to supply two examples of their own information needs. Depending on the coverage in the collection, one of these tasks was selected by the experimenter for the experiment. In total, each searcher performed three tasks. With a total of 11 participating organisations, 76 searchers performed 228 tasks in iTrack 05.

4.2 Desktop-Based System

For iTrack 05 the Daffodil framework was used and extended to meet the functionality of XML retrieval. Daffodil is a front-end to federated, heterogeneous digital libraries. It is aimed at providing strategic support (see [17]) during the information search process and already supports interactive retrieval through integrated high-level search and browse services.

The Daffodil framework consists of two parts, the graphical user interface client and the agent-based backend services (see [18, 19]). The user interface client, implemented in Java, is based on a tool metaphor, where each service is presented by a tool and the tools are integrated among each other.

The interface for iTrack 05 was designed by taking into account the findings of the iTrack 04, the berrypicking model described in section 2 and iconic visulisation techniques for better recall and immediate recognition.

Additions to the Architecture. The base system had to be extended for INEX in order to deal with the highly structured XML data. These extensions affected both the user interface and the corresponding backend services, e.g. connecting the XML search engine.

Query formulation. The problem of limited query language expressiveness was resolved by allowing Boolean queries, in combination with proactive query formulation support [20]. The latter feature recognises syntactic errors and spelling mistakes, and marks these. Besides full-text search, the system now also allowed for searching on metadata fields such as authors, title, year.

For further support during query formulation we added a Daffodil service for suggesting related query terms (based on statistical analysis of a different corpus). While the user specifies her query, a list of possible alternative terms are presented to her. This service follows the berrypicking model because the newly discovered related terms can change the search direction of the user. For easy query reformulation, the drag&drop feature of DAFFODIL could be used to add new query terms from documents or the related term list.

Resultlist presentation. In order to resolve the issues of *overlapping elements* and *element summarisation* identified in iTrack 04, results in the resultlist were now grouped document-wise and hits within documents were presented as possible entry points within the hierarchical document structure. The document metadata information is shown as the top level element, as depicted in Figure 3.

In addition, whenever some element within a document is retrieved, the title of that element is presented as a document entry point, depicted as a clickable folder icon. This change reflected user preference for the TOC view, where titles of elements are displayed.

We also took into account the comments about the retrieval score and the result path expression from iTrack 04. The retrieval score of each retrieved element was now shown in pictorial (as opposed to numerical) form, and result path expressions of elements were removed from the resultlist. The whole resultlist entry was made clickable.

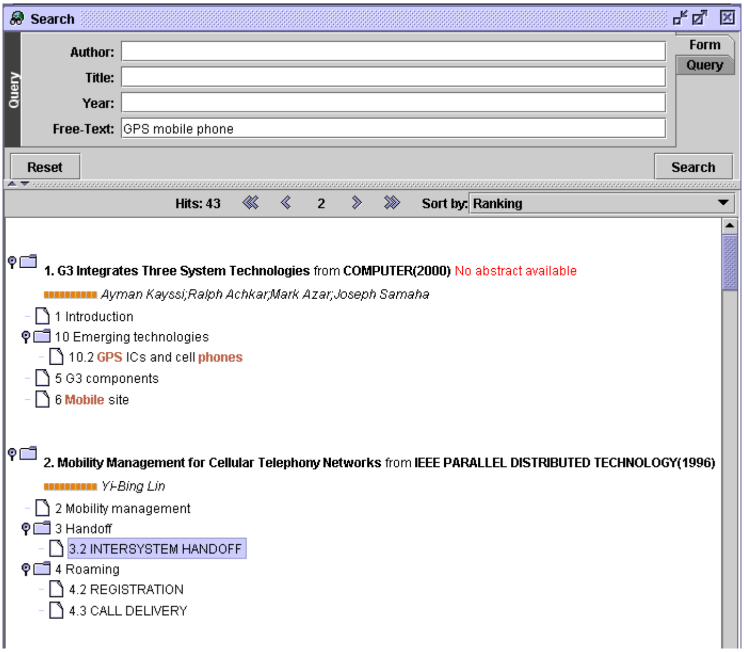
The comments on the distinction between visited and unvisited elements were considered by using an iconic visualisation technique. An eye icon is shown with any resultlist entry that has been visited before. The analogy with the berrypicking model is given here as marking the paths where a user walked to pick only unknown berries, to avoid looking twice at the same information. We also adopted query keyword highlighting at the resultlist level, since searchers appreciated this feature at the detail view level.

Detail view. The main layout of the detail level was kept the same as in iTrack 04, as seen in Figure 4. Some additions were made for supporting document browsing. First, the entry points from the resultlist level are now also highlighted in the detail view. Second, elements already visited are indicated with an iconised eye in the table of contents.

Many participants in iTrack 04 felt that the two-dimensional relevance scale used in these experiments was too complex [21]. For this reason, we moved to a simple 3-point scale, measuring only the usefulness of an element in relation to the searcher's perception of the task: 2 (Relevant), 1 (Partially Relevant), and 0 (Not Relevant). This three grade relevance scale was visualised as shown in Figure 4 (top left hand). The same icons were added to the viewed element when a relevance value was assigned by the user. Here again one more aspect of the berrypicking model analogy was implemented successfully: the user puts the 'good' beeries into her basket, and also can see which berries she has picked before.

4.3 Findings

The analysis was made along the same lines as for iTrack 04. The overall opinion of the participants about the system was recorded in the final questionnaire



Author:

Title:

Year:

Free-Text: GPS mobile phone

Hits: 43 2 Sort by: Ranking

1. G3 Integrates Three System Technologies from COMPUTER(2000) No abstract available
Ayman Kayssi;Ralph Achkar;Mark Azar;Joseph Samaha
- 1 Introduction
- 10 Emerging technologies
  - 10.2 GPS ICs and cell phones
- 5 G3 components
- 6 Mobile site

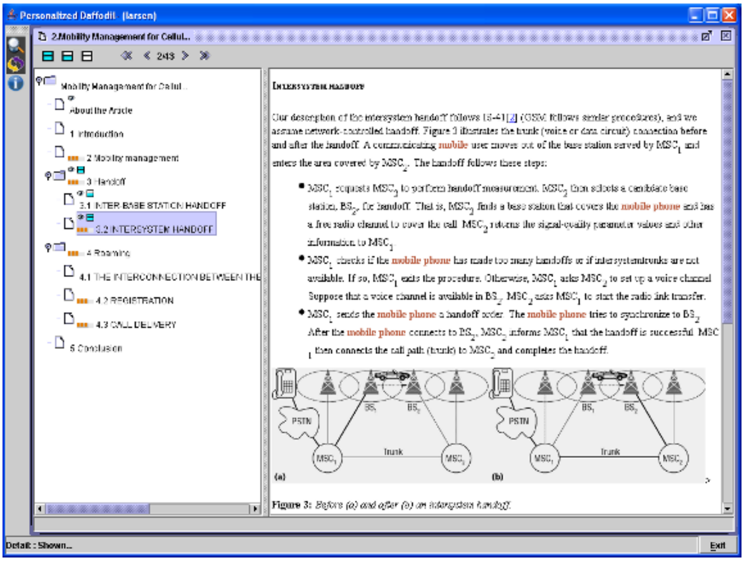
2. Mobility Management for Cellular Telephony Networks from IEEE PARALLEL DISTRIBUTED TECHNOLOGY(1996)
Yi-Bing Lin
- 2 Mobility management
- 3 Handoff
  - 3.2 INTERSYSTEM HANDOFF
- 4 Roaming
  - 4.2 REGISTRATION
  - 4.3 CALL DELIVERY

Fig. 3. iTrack 05: Query form and resultlist

Table 3. Overall opinion about the system on the scale of 1 (Not at all) to 5 (Extremely) in iTrack 04 (88 searchers) & iTrack 05 (76 searchers)

System Features	iTrack 04		iTrack 05	
	μ	σ^2	μ	σ^2
How easy was it to learn to use the system?	4.17	0.6	3.40	0.9
How easy was it to use the system?	3.95	0.7	3.96	0.9
How well did you understand how to use the system?	3.94	0.5	3.84	0.9
How well did the system support you in this task?	-	-	3.13	0.9
How relevant to the task was the information presented to you?	-	-	2.97	1.13
Did you in general find the presentation in the resultlist useful?	-	-	3.35	0.8
Did you find the table of contents in the detail view useful?	-	-	3.72	1.0

that they filled after the completion of all tasks. New questions enquiring about the distinct aspects of the system used in 2005 were added. The results are summarised in Table 3. As can be seen users were in general positive on both systems, and the major difference between the two years was the better learnability of the 2005 system. In addition, there were many informal comments in response to the questions mentioned in section 3.3. We analyse the data in the following paragraphs.



INTERSYSTEM HANDOFF

Our description of the intersystem handoff follows IS-41[2] (GSM follows similar procedures), and we assume network-controlled handoff. Figure 3 illustrates the trunk (voice or data circuit) connection before and after the handoff. A communicating mobile user moves out of the base station served by MSC_1 and enters the area covered by MSC_2. The handoff follows these steps:

- MSC_1 requests MSC_2 to perform handoff measurement. MSC_2 then selects a candidate base station, BS_2, for handoff. That is, MSC_2 finds a base station that covers the mobile phone and has a free radio channel to cover the call. MSC_2 returns the signal-quality parameter values and other information to MSC_1.
- MSC_1 checks if the mobile phone has made too many handoffs or if intersystemtrunks are not available. If so, MSC_1 exits the procedure. Otherwise, MSC_1 asks MSC_2 to set up a voice channel. Suppose that a voice channel is available in BS_2. MSC_2 asks MSC_1 to start the radio link transfer.
- MSC_1 sends the mobile phone a handoff order. The mobile phone tries to synchronize to BS_2. After the mobile phone connects to BS_2, MSC_2 informs MSC_1 that the handoff is successful. MSC_1 then connects the call path (trunk) to MSC_2 and completes the handoff.

Figure 3: Before (a) and after (b) an intersystem handoff.

Fig. 4. iTrack 05: Detail view

Resultlist presentation. Presentation of results in a hierarchy is generally found useful. 43 users commented positively on it, whereas 3 users found the information presented insufficient for deciding about relevance or irrelevance. 2 users commented on the inconsistency of the result presentation. This situation occurred when a whole article was retrieved as a hit, with no further elements within this article, 3 users disliked scrolling at the result list level.

Table of contents and query term highlighting. As in iTrack 04, the TOC is found to be extremely useful and 32 users commented positively on it. Query term highlighting in the resultlist and the detail view were also appreciated (22 positive comments).

Related terms. The new functionality of suggesting related query terms was found highly helpful: 29 users found this function useful in their performance of search tasks. There were some cases when the suggested terms either retrieved no documents, or there was no obvious semantic relationship to the query terms. These situations led to negative remarks by 11 searchers.

Awareness in the detail view. The document entry points shown in the resultlist were also displayed in the detail view, 14 users commented positively on it. In addition, icons indicating visited elements and their relevance assessments are shown in the TOC: 3 users found this useful. In addition, 15 users also wanted to have the relevance assessment information in the resultlist.

Retrieval quality. Although the underlying retrieval engine had shown good retrieval results in previous INEX rounds, it produced poor answers for some queries, so 25 users commented negatively on this. A possible reason could be the limited material on the choosen topic of search.

Other Issues. 4 users remarked positively on the interface usefulness and 3 liked the query form. The response time of the system was encountered as being too high, so 35 users comments negatively on it.

Overall, user responses show that the main weaknesses of the iTrack 04 interface have been resolved. In addition, the new features supporting the berrypicking paradigm were appreciated by the users.

5 Conclusion and Outlook

In this article we presented the lessons learned from INEX iTrack 04 to iTrack 05. The analysis of iTrack 04 showed several negative responses to the used web-based interface. The main issues were the overlapping elements presented in a linear resultlist, insufficient summaries to indicate the relevance of an item, the lack of distinction between visited and unvisited items and a limited query language. Also some positive comments were made, e.g., the document structure (TOC) provided sufficient context and was a quick way of locating the interesting information. Keyword highlighting was also found to be helpful in 'catching' information parts that may be relevant to the existing query terms.

These findings were used to shift to an application-based interface. The analysis of iTrack 05 showed that the overlapping elements presentation in a hierarchy can provide sufficient summerisation and context for the decision of relevance or irrelevance. The second major improvement was the addition of design elements based on the berrypicking model [1], which received substantial appreciation. These desgin elements included keyword highlighing, iconic visualisation and provision of related terms.

The most problematic issue with the iTrack 05 system was the responsiveness of the system. This was due to the underlying search engine and inefficiencies within the DAFFODIL message flow. These issues will be worked on for iTrack 06.

Overall, the evaluations showed that interface design adaptation based on the 2004 findings were taken as an improvement. The shift to an application based framework proved to be the right step, as we gained more flexibilty in features besides a web-based framework. In iTrack 06 a major focus will be the efficiency, by replacing the underlying search engine and a tighter integration with the DAFFODIL framework to lower response times.

References

1. Bates, M.J.: The design of browsing and berrypicking techniques for the online search interface. Online Review **13** (1989) 407–424
2. Voorhees, E., Harman, D.: Overview of the eighth Text REtrieval Conference (TREC-8). In: The Eighth Text REtrieval Conference (TREC-8). NIST, Gaithersburg, MD, USA (2000) 1–24

3. Turpin, A.H., Hersh, W.: Why batch and user evaluations do not give the same results. In: Proc. of SIGIR, ACM Press (2001) 225–231
4. O'Day, V.L., Jeffries, R.: Orienting in an information landscape: How information seekers get from here to there. In: Proc. of the INTERCHI '93, IOS Press (1993) 438–445
5. Finesilver, K., Reid, J.: User behaviour in the context of structured documents. In: Proc. of ECIR. (2003) 104–119
6. Larsen, B., Malik, S., Tombros, A.: The interactive track at inex 2005. In: Advances in XML Information Retrieval and Evaluation: Springer, p. 398-410. (Lecture Notes in Computer Science vol. 3977). (2006)
7. Tombros, A., Larsen, B., Malik, S.: The interactive track at inex 2004. In: Advances in XML Information Retrieval: Springer, p. 410-423. (Lecture Notes in Computer Science vol. 3493). (2004)
8. Kamps, J., de Rijke, M., Sigurbjörnsson, B.: University of amsterdam at inex 2005. (In: Advances in XML Information Retrieval and Evaluation: Springer, p. 398-410. (Lecture Notes in Computer Science vol. 3977))
9. van Zwol, R., Spruit, S., Baas, J.: B^3-sdr@inetractive track: User interface design issues. (In: INEX 2005 Workshop Pre-Proceedings)
10. Crestani, F., Vegas, J., de la Fuente, P.: A graphical user interface for the retrieval of hierchically structured documents. Information Processing and Management **40** (2004) 269–289
11. Großjohann, K., Fuhr, N., Effing, D., Kriewel, S.: Query formulation and result visualization for XML retrieval. In: Proceedings ACM SIGIR 2002 Workshop on XML and Information Retrieval. (2002)
12. Gövert, N., Kazai, G.: Overview of the INitiative for the Evaluation of XML retrieval. In: Proc. of INEX workshop. (2003) 1–17
13. Borlund, P.: Evaluation of interactive information retrieval systems. (2000) 276 PhD dissertation.
14. Fuhr, N., Gövert, N., Großjohann, K.: HyREX: Hyper-media retrieval engine for XML. In: Proceedings of the 25th Annual International Conference on Research and Development in Information Retrieval. (2002) 449 Demonstration.
15. Gövert, N., Fuhr, N., Abolhassani, M., Großjohann, K.: Content-oriented XML retrieval with HyREX. In: Proc. of INEX workshop. (2003) 26–32
16. Tombros, A., Malik, S., Larsen, B.: Report on the INEX 2004 interactive track. SIGIR Forum **39** (2005)
17. Klas, C.P., Fuhr, N., Schaefer, A.: Evaluating strategic support for information access in the DAFFODIL system. In: Proc. of 8th ECDL. (2004)
18. Fuhr, N., Klas, C.P., Schaefer, A., Mutschke, P.: Daffodil: An integrated desktop for supporting high-level search activities in federated digital libraries. In: Proc. of 6th ECDL, Springer (2002) 597–612
19. Fuhr, N., Gövert, N., Klas, C.P.: An agent-based architecture for supporting high-level search activities in federated digital libraries. In: Proc. of ICADL, Taejon, Korea, KAIST (2000) 247–254
20. Schaefer, A., Jordan, M., Klas, C.P., Fuhr, N.: Active support for query formulation in virtual digital libraries: A case study with DAFFODIL. In: Proc. of 7th ECDL. (2005)
21. Pehcevski, J., Thom, J.A., Vercoustre, A.: Users and assessors in the context of inex: Are relevance dimensions relevant? In: Proc. of INEX Workshop on Element Retrieval Methodology. (2005)