# Relevant Contextual Features in XML Retrieval

Georgina Ramírez and Arjen P. de Vries

Centre for Mathematics and Computer Science,
P.O. Box 94079, 1090GB Amsterdam, The Netherlands
{georgina,arjen}@cwi.nl

**Abstract.** We present and analyze results of a collaborative user study carried out by the *Interactive Track* at INEX 2005 [4]. The overall goal of the Interactive Track is to investigate the behavior of users when interacting with components of XML documents. In this paper we use the data collected from this user study to investigate dependencies between several contextual features and the structural characteristics of the relevant components (e.g. type and number of relevant elements). The three contextual features analyzed are the user's familiarity with the topic, the request's type, and the user's motivation to perform the search. The dependencies found indicate that XML retrieval systems should also benefit from the use of contextual information.

## 1 Introduction

The work presented in this paper is part of a research project that studies and develops an experimental approach to correlate search tasks and IR strategies on the use of structural information. Its goal is to investigate the use of structural features for effective information retrieval and to define a model to link search tasks to specific retrieval strategies.

To achieve our goal and to be able to define the best retrieval strategy for each of the tasks, we need to understand the nature of the different search tasks as well as the different contextual factors that influence the search.

In this paper, we present results of an interactive user study and analyze the effects on relevance judgments for three different contextual factors: (1) user's familiarity with the topic, (2) request's type, and (3) user's motivation (intention) to perform the task.

In particular, we investigate if there are correlations between these contextual features and the structural characteristics of the relevant XML components (e.g. type and number of relevant elements). Our main research question can be expressed as follows:

- Can we identify a measurable dependency between a topic's contextual factors and the structural aspects of the topic's relevant components?

These dependencies would indicate that XML retrieval systems can use contextual information to restrict their search space or adapt their search results to the specific user by exploiting the structural nature of the XML documents.

The paper is organized as follows. We start giving some background in Section 2. In Section 3 we present the user study from the Interactive Track at INEX. The different contextual and structural features analyzed are described in Section 4 and our findings presented in Section 5. We conclude and discuss future work in Section 6.

## 2 Background

Many user studies have been performed to investigate user's seeking behavior in different domains (e.g., [10],[2],[5]). In the area of XML retrieval, the Interactive Track at INEX [11] started in 2004 to fulfill the need of understanding user behavior in this setting. However, as far as we know, most of the efforts have concentrated on collecting evidence and few analysis have been made available.

On the other hand, many information retrieval systems try to incorporate the use of contextual information to improve effectiveness. Thus, understanding and modeling contextual information is becoming an important issue (see, for instance, [8, 7]).

The analysis presented in this paper is related to both fields. It tries to understand the importance of some contextual features by analyzing the relevance data of an interactive experiment. In this way, a more *realistic* contextual behavior can be incorporated into information retrieval systems.

## 3 The Interactive User Study at INEX

The data analyzed in this paper is part of a collaborative effort carried out by the *Interactive Track* at INEX 2005 [4]. Each of the eleven participant institutions performed a minimum of six user experiments following a common methodology. The overall goal of these experiments was to investigate the behavior of users when interacting with components of XML documents. We describe below the general aspects of the methodology used (designed by the organizers of the track) and refer the reader to [11] for more detailed information on the specific setup of the experiments or the track.

### 3.1 Participants

A total of 73 test persons from 20 different nationalities performed the experiment. Their ages ranged from 19 to 52 and the average age was 28. 29% of the participants were female. 60% of the participants were students, 12% were Ph.D. students, 18% had another academia related profession (e.g. researcher, post-doc, assistant professor), and 10% had other occupations (e.g. designer, librarian, system administrator).

### 3.2 Collection

The collection searched by the test persons was the collection used at INEX [3]; a subset of IEEE Computer Society publications, consisting of 16.819 scientific articles from 24 different journals (ranging within the years 1995-2004).

### 3.3 Tasks

Each of the test persons performed two simulated search tasks and one search task from an information need of their own. For each of the simulated tasks, the searcher could choose one out of three possibilities. An example of a simulated search task is given in Table 1. To create their own search task, the searchers were given a description of the collection's content and an example of a search task. By filling a questionnaire, searchers had to specify (1) *what are they looking for*, (2) *what is the motivation of the topic* (i.e. *why* are they searching the information, what *problem* can be solved with the information and in what *context* did the problem arise), and (3) what would an ideal answer look like. In order to guarantee that topics were covered by the collection, the test persons were asked to present two different information needs before the experiment. The experimenter could then perform a preliminary search and get an idea of the collection's coverage on the topics. If both topics had good or little coverage, the searcher could choose the preferred one. In case one of them was covered and the other not, the experimenter would advise the searcher to use the most covered one.

**Table 1.** Example of simulated search task.

Your department has produced a Linux-program and it is being discussed whether to release it under a public license such as GNU or GPL (General Public License). Therefore, you have been asked to find information about the implications of releasing the code under a public license as an open source program. Find, for instance, information that discusses different licensing schemes or articles about the impact of open source programs.

### 3.4 Procedure

The experiments started with an explanation of the procedure, a description of the system and a training session with an example topic. After that, searchers filled in a general entry questionnaire and performed the three tasks. For each task they filled a pre-task and a post-task questionnaires. Searchers had a maximum of 20 minutes to perform each task. After all tasks were performed, the searches filled in a final questionnaire and had a short interview with the experimenter. To neutralize learning effects, the order in which task categories were

performed was permuted. Thus, for each 6 searchers, no order for performing the tasks was repeated. Participants were also asked to assess the relevance, while performing the task, of the components and documents they were seeing. However, this was not enforced by the system. Users could judge the documents/components seen as *Relevant*, *Partially Relevant* or *Not Relevant*.

### 3.5 System, interface and logs

The organizers of the track provided a common system that all participant sites used to perform their experiments. The XML elements (components) considered by the system were limited to: articles, article's metadata (fm), sections (sec), subsections (ss1) and sub-subsections (ss2). The metadata component contained the title, author, journal, year, and abstract of the article.

In response to a searcher's query, the system presented a result list with the title of the highest scored elements in the collection grouped by their containing articles (represented by the title, author, journal and year). When the user clicked on any of these elements, the system presented the table of contents of that article and the text of the clicked component. Users could then move within the article by clicking at the components in the table of contents. Note that to access the full text of an article, users had to first click on any of the elements of the result list and then click on the title of the article in the table of contents. In this second view, searchers could assess the relevance of the component shown. The system recorded the *click* data as well as the relevance judgments done by the searchers.

## 4 Data preparation

The main goal of our analysis is to investigate which contextual features are relevant in an XML retrieval setting. In particular, we want to find out if the structural characteristics of relevant components differ when the search is influenced by different contextual factors. In this section we describe the different contextual and structural features used in the analysis and explain how the data was classified into the different categories. Note that, since the experiment was mainly designed to investigate the behavior of users when interacting with XML documents, not many aspects of the context of the search were recorded. However, when creating their own information needs users were asked about several issues regarding the context of the search. We used the descriptions given by the test persons to extract different contextual information.

### 4.1 Contextual Features

For our analysis, we chose the following three different contextual features that we considered to be possibly relevant in an XML retrieval setting:

**Searcher's familiarity with the topic** It has already been shown that behavior of searchers differs between those that have different degrees of familiarity with the searched topic (e.g., [10]). In this paper, we investigate if, in a similar manner, structural characteristics of relevant elements also differ between users that have different degrees of familiarity with the topic. The information of the familiarity of the searcher was recorded in a 1-5 scale in the pre-task questionnaire. From this information, we classified the users into three categories: the users that are *Not* familiar with the topic (1-2), the users that are *Somehow* familiar with the topic (3), and the users that are (*Yes*) familiar with the topic (4-5).

**Request's type** One of the characteristics of XML retrieval is that users can perform very focused searches and ask only for a specific type of information (references, experimental results, etc.). This is one of the reasons why several query languages and interfaces have been designed – to allow users to explicitly express more complex needs. However, these tools are not always available and users often specify in their *keyword* queries not only what they are looking for but also the type and the specificity of the information they are searching for. We hypothesize that this type of contextual information can help an XML retrieval system to decide which type of elements the user would like to see and thus return the most appropriate element types for each of the requests.

Since users were not explicitly asked about this aspect of the information need, we analyzed the descriptions they wrote about what they are searching for and manually classified their tasks using two different dimensions that can be used to classify standard IR requests [6]: The *specificity* and the *complexity* of the request.

In the *specificity* dimension, we classified requests into *Narrow* (N) and *Broad* (B) (also seen in the literature as *Specific* and *Generic*, e.g. [6]). In our case, *Narrow* topics are those which specify any type of constraint on the expressed information need, both, topically (i.e. focusing on a specific aspect of the topic) and structurally (i.e. asking for a specific type of information such as experiments or references). On the contrary, *Broad* topics are those that simply ask for information about a topic, in a general way, without any type of constraint.

In the *complexity* dimension, two categories were used: *Simple* (S) and *Compound* (C). *Simple* requests are those that ask for information about just one topic. While *Compound* requests are those that ask for information about several topics (or aspects of the topic) or want information about the relationship between two topics (e.g. technique A in the field of B or information about A for B).

*Compound* topics might be Broad or Narrow or both (B + N). The latter includes those search tasks where general information about a topic is requested but the user also mention some specific point of interest.

We classified the information needs given by the users into these five categories. The number of topics for each class and an example of each of them is given in Table 2.

**Table 2.** Number and example of search tasks belonging to each of the request type categories.

| Complexity | Specificity | Num. | Example |
|---|---|---|---|
| **Simple** (S) | **Broad** (B) | 12 | I search information about web services. |
| | **Narrow** (N) | 10 | I am looking for introductions to Data Mining. |
| **Compound** (C) | **Broad** (B) | 20 | Papers about 'named entity recognition' and 'clause boundary recognition'. |
| | **Narrow** (N) | 12 | Decidability and complexity results of (bounded/live) Petri Nets. |
| | **Broad and Narrow** (B + N) | 14 | I want information about web standards and W3Cs role in implementing these in various web browsers. |

**User's intention** Why do the searchers want the information and which problem this information might be able to solve could be an important contextual factor that might help to improve retrieval effectiveness. In web search, several works have shown that retrieval effectiveness can be improved when knowing user intentions [9, 1].

We analyzed the descriptions given by the searchers of their information need and classified their search task according to what searchers intend to do with the information found. The different *intentions* found can be classified into 5 general categories:

**Decide** The information is searched for making a decision. In most of the cases, the user wants to compare possibilities and then decide or draw some conclusions. Work tasks include reviewing a paper or business decisions.

**Apply** The information is searched for using it in a practical way. Searchers have a specific design problem and search for information to solve it. The underlying work tasks are rather practical: programming, developing a software, implementing, etc.

**Explain** The information is searched for *knowledge transfer*. The motivating work tasks are writing (articles, reports, etc.) and teaching (preparing lectures).

**Study** The information is search for learning, studying. Searchers want to know and understand more about a topic. Work tasks behind the search are related to following courses or participating in some research project, but also for business or job interest.

**Personal Interest** The information is searched for general and personal interest or curiosity. No specific work task motivates the search.

Unfortunately, many searchers did not give a proper description of the purpose of their search. Table 3 shows for each category the number of topics that could be classified and an example topic.

**Table 3.** Number and example of description belonging to each of the intention categories.

| Class | Num. | Example |
|---|---|---|
| **Apply** (A) | 9 | My computer was upgraded by a friend. I have the state of the art anti virus. Yet the worms keep coming. I want to know what to do. |
| **Decide** (CD) | 6 | The department is trying to decide whether to release a produced Linux-program under a public license such as GNU or GPL. |
| **Explain** (E) | 10 | I am writing an article about the history of information systems and the projections and expectations made by experts when they were introduced. |
| **Study** (S) | 13 | I am taking a course in networks, and want to know more. The literature we used didn't give the right information. |
| **Personal Interest** (PI) | 6 | Just out of general interest. I would like to know, for instance, when spamming was first acknowledged as a problem. |

## 4.2 Structural features

XML retrieval systems try to exploit the structural characteristics of the documents to effectively retrieve XML components from XML documents. Thus XML retrieval systems differ from standard document retrieval systems in that they have the extra job of deciding which type of components are the most appropriate to fulfill each of the information needs. However, XML retrieval systems have also the benefit of this extra source of information not available to plain document retrieval systems: the document structure. Structural information may be useful to identify the relevant part of the documents and thus, produce a more focussed ranking. In this respect XML retrieval systems could benefit from contextual information by using it as a way to find the structurally relevant information for a specific context and thus reducing the search space. We want to investigate if structural differences exist between different search contexts; with that purpose in mind, we analyze the following types of structural information:

**Number and type of relevant elements** Since XML retrieval systems can decide which components to retrieve for a specific request, to know how many and which types are desirable for each of the contextual factors can significantly improve effectiveness. In the experiment, only 5 types of elements were shown to the users: articles, metadata (fm), sections (sec), subsections(ss1) and subsubsections(ss2). We analyze which of these elements were classified as *Relevant* for each of the contexts defined above.

**Number of different articles/journals** In the INEX collection, articles are grouped by journals. To know how many different journals or articles contain the information desired for each task is an important clue for the information systems. If systems can find which are the important articles and journals for a task (e.g. during an interactive session), the search space can be reduced and a more specific search can be performed. We analyze also this type of information. Note that articles that contain relevant information might have not been assessed. Thus the information of the number of articles containing relevant information might differ from the information of the articles (as element type) assessed relevant.

## 5 Results and Findings

The interactive user study provided 219 search tasks. Of these, we excluded 11 because of logging problems or lack of relevance judgments. The remaining 208 search tasks consist of 68 user formulated tasks and 140 simulated tasks. We present only the analysis from the user formulated tasks because, for the type of information we analyse, the fact that many users performed the same simulated task could create a bias in our results. The user formulated tasks are all unique and independent from each other.

We focus our analysis on the structural characteristics of what users assessed as relevant during the experiments. Since users were not forced to asses everything they viewed, relevance judgments are not complete. However, by analyzing what they assessed, we can get an estimation of what is relevant and what is not for each of the tasks.

We first present a general overview of what was assessed and present the general statistics of the relevant structural characteristics. In the rest of the section, we analyze the three contextual features presented and investigate if there are differences in what is assessed relevant in each of the cases.

### 5.1 Relevance Overview

During the 68 search tasks, 956 elements were assessed; an average of 14.1 elements per task. From those, 31% were assessed as relevant, 34% as partially relevant and 35% as not relevant. On average, 9.1 elements were found relevant per task (including partially relevant). The average number of articles containing relevant information was 4.0 and relevant information appeared, on average, in 2.6 different journals. The distribution of the element types that were assessed during the experiments is shown in Figure 1.

In absolute numbers, sections and subsections were the most relevant and partially relevant elements found by the test persons. This means that users considered small parts of documents relevant: a good indication that focused retrieval is useful for these types of tasks. In relative numbers however, if we look at what users considered most relevant (not partially relevant), we find that 33% of the articles assessed were considered relevant, almost the same as with
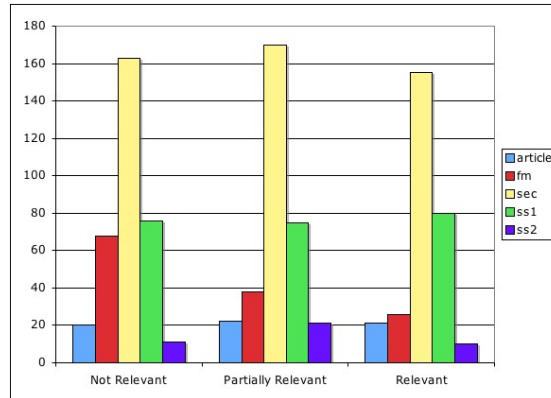
**Fig. 1.** Histogram of assessed element types for each of the relevance values.

the sections (32%), and the subsections (35%). Sub-subsections and metadata were found less useful (24% and 20% respectively). This might indicate that too small elements do not contain enough information to be relevant on their own and therefore they are not desirable by users.

### 5.2 Searcher's familiarity with the topic

We first analyze if the different degrees of knowledge the users had on the topic lead to different categories of tasks, in terms of specificity, complexity or intention. In Figure 2, we can see which type of requests users with different degrees of familiarity with the topic performed.

As expected, the more familiar the user is with the topic, the more *compound* tasks are performed. It is also not surprising that users without much knowledge on the topic performed *broader* (B) tasks than those knowing it well. Another interesting result is that the intentions of the less informed users were only distributed among three of the intention categories: personal interest (PI), apply (A), and decide (CD), while only users with knowledge on the topic performed study (S) and explain (E) tasks. This type of information is important because it can be used by retrieval systems to predict users intentions and types of request and adapt their search accordingly.

Regarding the effects of this contextual feature to the amount of relevant information found, we can see in Figure 3 that, on average, the users that did not know much about the topic found less relevant information. The most knowledgeable users are the ones that found more relevant articles and journals. Unfortunately, none of these differences are statistically significant.

Figure 4 shows the element types that were found relevant per user category. Note that partially relevant elements are not considered – we are interested in learning where users found the most useful information. All users found the most useful information at a section and subsection level. However, some difference can
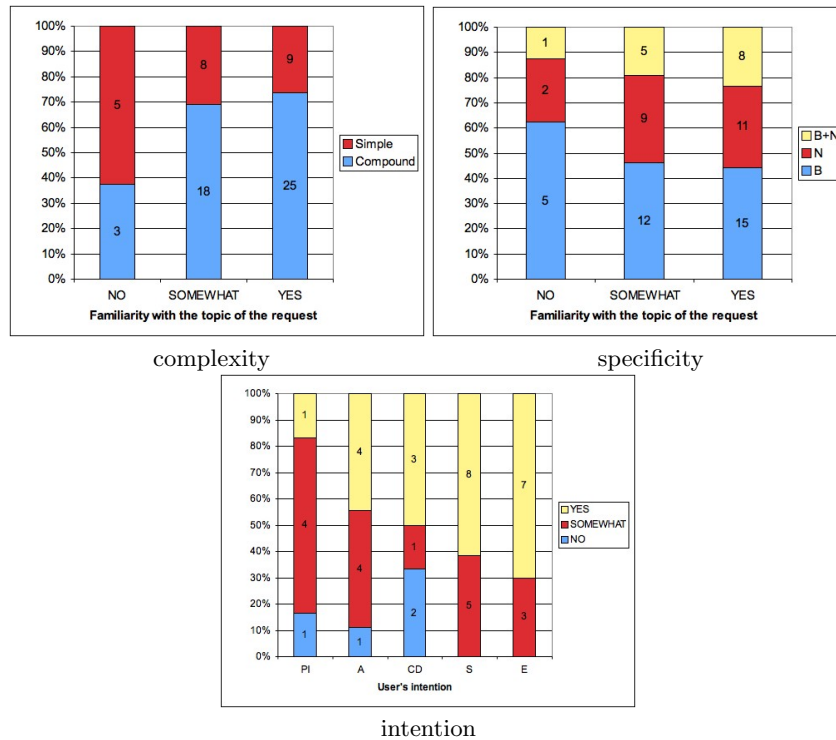
complexity · specificity · intention

**Fig. 2.** Number and types of search tasks performed by the test persons

be observed. While users without knowledge seem to prefer around three times more the sections than the subsections, for knowledgeable users this difference is only around the double. The less knowledgeable users did not find the metadata information useful.

### 5.3 Type and specificity of the request

Figure 5 presents the number of elements, articles and journals found relevant for each of the categories described in Section 4.1. On average, almost three more elements were assessed relevant in *Narrow* (NS&NC) tasks (11.5) than in *Broad* (BS&BC) tasks (8.7). Users with *Simple* (BS&NS) tasks found more relevant elements (10.2) than those with *Compound* (BC&NC&B+N) tasks (8.5). However, users with *Compound* tasks found more relevant articles (4.1) that those with *Simple* tasks (3.8). That is an indication that *Compound* tasks might require of several pieces of information distributed among different articles while the information to answer *Simple* tasks is gathered in the same few articles.

Although some of these differences are not large, the graphs do indicate some tendencies that retrieval systems might be able to exploit. For instance,
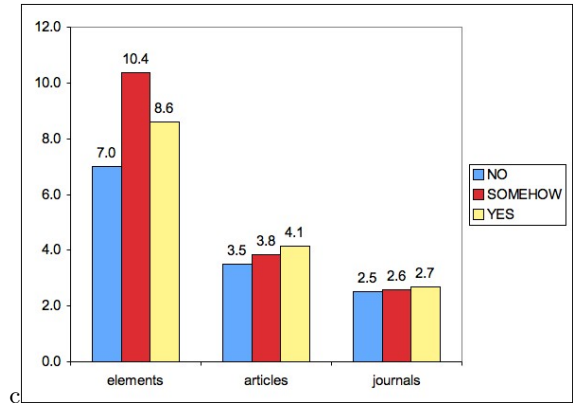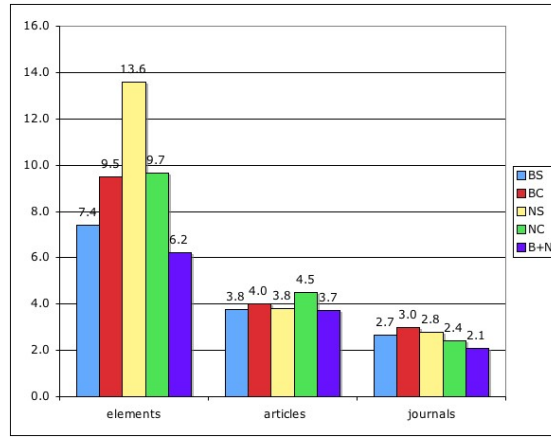
**Fig. 3.** Average number of elements, articles and journals containing relevant information per task.
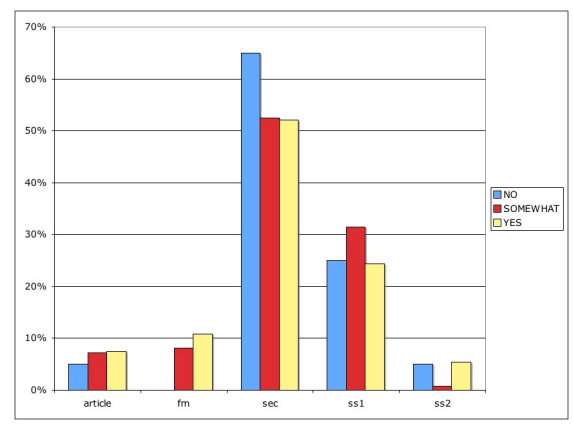


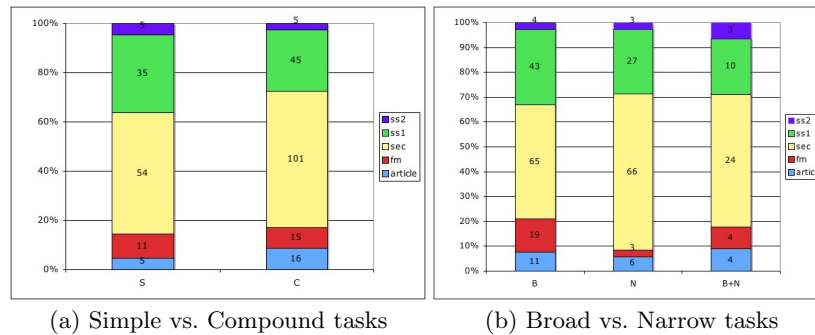**Fig. 4.** Types of elements assessed relevant (not partially relevant) per group of users.

they confirm our intuition that users with *Broad* search tasks are pleased with information from a wider variety of journals (2.9), as opposed to the information requested from users performing *Narrow* tasks (2.6) or *Broad and Narrow* tasks (2.1). It is more surprising that users performing specific tasks are the ones finding more relevant elements.

Regarding the type of elements found most useful by the test persons (assessed as relevant during the experiments), we see in Figure 6(a) that users of *Simple* tasks are less happy with the article components than users of the *Compound* tasks. They also liked more very small elements such as subsections and sub-subsections than the users performing *Compound* search tasks. This might be because complex tasks require information contained in different elements.

**Fig. 5.** Average number of elements, articles and journals containing relevant information per request type.

Thus, bigger units that include several parts might be needed to fulfill this type of information needs.



(a) Simple vs. Compound tasks

(b) Broad vs. Narrow tasks

**Fig. 6.** Types of elements assessed relevant for each of the request types categories. The numbers in the bars indicate the number of elements of that type assessed relevant.

In Figure 6(b) we see the same information for the specificity dimension. Here, we find that *Broad* requests users assessed more articles and metadata as useful for their search than users with *Narrow* tasks. More surprising is that these users also assessed very small elements as relevant (more than 30% of the assessed elements were subsections and sub-subsections). For both types of tasks, sections are the most useful element. However, that was in 63% of the cases for elements assessed in the *Narrow* tasks and only 46% for the ones assessed by users performing *Broad* tasks.
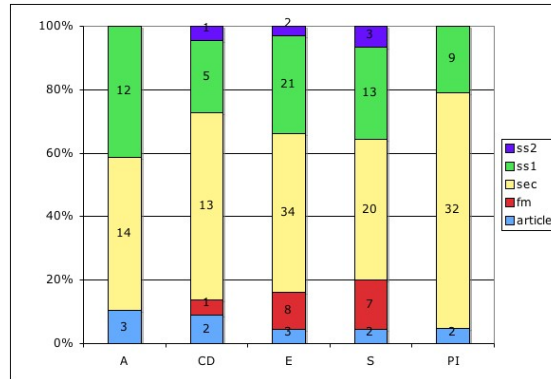
### 5.4 User's intention

Figure 7 shows the number of relevant elements found for each of the *intention* categories described in Subsection 4.1. It is somehow surprising that the users that found more relevant information are those that needed the information for teaching or for writing. A more expected result is the fact that users that were just searching for personal interest found many relevant elements as well. The users that needed the information to apply it in their own environments are the ones that found less relevant information. Notice, however, that to find more relevant elements does not necessarily imply that more information was contained in the collection. If we look at the number of articles containing relevant information, the users of the search tasks classified as *Explain* are those that found less articles. So, they must have assessed many parts of these articles as relevant. That suggests that this type of users might want to see complete articles or large parts of them, instead of smaller fractions. While the users that search for personal interest might be happy to see a wider range of information contained in several articles.
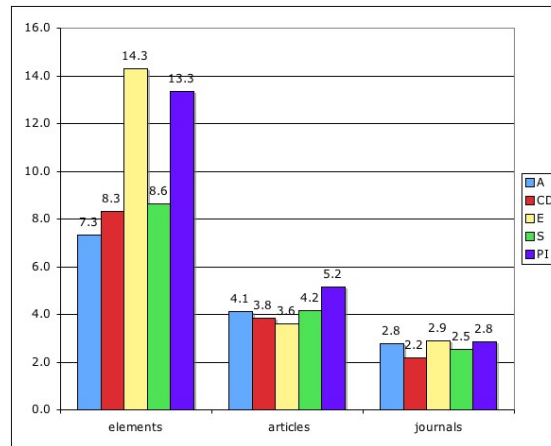


**Fig. 7.** Average number of elements, articles and journals containing relevant information per categories of intention.

We summarize the number of elements of each type that were found relevant for the different intentions in Figure 8. The users that needed to *Apply* and the users that searched for personal interest were the ones finding the representations of articles (articles and metadata) the least useful, and the section and subsection levels the most useful. The users searching for *Study* or to *Explain* are the ones that found the metadata descriptions useful. That might be because users with these work tasks are often trying to find references or works related to what they study. A reference describing it might already be of interest.

**Fig. 8.** Types of elements assessed relevant (not partially relevant) according to intentions. The numbers in the bars indicate the number of elements of that type assessed relevant.

## 6 Conclusions

We presented the results of an interactive experiment where users performed searches on a collection of XML documents. We investigated if there are structural differences between elements that were assessed relevant for the different contextual features: user's familiarity with the topic, request's type, and user's intentions.

Answering our research question, we can say that several dependencies between the different topic's contextual factors and the structural aspects of the topic's relevant components were found. These differences indicate what users want in each of the context situations, and we expect that retrieval systems could make use of this information to tune their results to the preferences of that specific type of request or user.

The presented analysis demonstrates that the familiarity of the user with the topic of the search task is an important factor to estimate the type of task the user is performing (e.g., for automatically classifying tasks or queries). The other two contextual factors give indications of which type of elements the user is searching for. For instance, although on average users had a preference for section level results, users with *Compound* and *Broad* tasks prefer longer elements while users trying to find information for a practical purpose prefer the shorter ones.

We have also shown general behavior of users searching XML documents. On the one hand, users assessed large numbers of sections and subsections as relevant. Thus, focused information is highly relevant. This means that for many tasks, users are happy with small elements, and systems that perform focused retrieval might be wh at they need. On the other hand, we showed that the number of relevant articles and journals containing relevant information is, on average, very small compared to the number of articles (16.819) and journals (24) that exist in the collection. We argue that this is another important contextual factor.

If XML retrieval systems can find out what are the articles or journals where the relevant information for that task is contained (during, e.g., an interactive retrieval session), they could automatically reduce the search space and concentrate on finding the relevant parts of those. Since any collection of documents is structured in one way or another, we believe that the organization of the collection (in a similar way as the journals for INEX) can be a good contextual factor to consider in other scenarios too.

Unfortunately, most of the differences presented are not statistically significant and more evidence need to be collected in order to decide whether these contextual information can be used by IR systems to adapt their search strategy. We are also aware that the study presented has several phases and any of them is a potential confounding factor to be considered when analysing the results. Further analysis need to be done in order to establish the effects on the results of the various phases and variables of the study (such as the classification of tasks and intentions). Doing some preliminary work on this direction, we have seen that when analyzing the information from the user formulated topics together with the simulated ones almost all the same differences and tendencies appear again. This is a good indication that the trends shown in this paper are already a good estimation of the effects of contextual information on user's judgements.

As future work, we plan to continue our analysis on this data and apply the acquired knowledge to our XML retrieval system and see if retrieval effectiveness can be improved by the use of this type of contextual information.

## 7 Acknowledgements

## References

1. M. Bomhoff, T. Huibers, and P. van der Vet. User Intentions in Information Retrieval. In *Proceedings of the Fifth Dutch-Belgian Information Retrieval Workshop*, 2005.
2. A. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.
3. N. Fuhr, N. Gövert, G. Kazai, and M. Lalmas. INEX: INitiative for the Evaluation of XML Retrieval. In *Proceedings of the SIGIR 2002 Workshop on XML and Information Retrieval*, 2002.
4. N. Fuhr, M. Lalmas, S. Malik, and G. Kazai, editors. *Advances in XML Information Retrieval. Fourth Workshop of the INitiative for the Evaluation of XML Retrieval (INEX 2005)*, volume 3977 of *Lecture Notes in Computer Science*. Springer-Verlag, 2006.
5. C. Häscher and G. Strube. Web search behavior of Internet experts and newbies. In *Proceedings of the 9th international World Wide Web conference on Computer networks : the international journal of computer and telecommunications netowrking*, pages 337–346, Amsterdam, The Netherlands, The Netherlands, 2000. North-Holland Publishing Co.

6. P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*, volume 18 of *The Information Retrieval Series*. Springer, 2005.

7. P. Ingwersen, K. Järvelin, N. Belkin, and B. Larsen, editors. *ACM SIGIR 2005 Workshop on Information Retrieval in Context (IRiX)*, 2005. `http://irix.umiacs.umd.edu/ACM-SIGIR2005-IRiX-proceedings.pdf`.

8. P. Ingwersen, K. van Rijsbergen, N. Belkin, and B. Larsen, editors. *IRiX: ACM SIGIR 2004 Workshop on Information Retrieval in Context*, 2004. `http://ir.dcs.gla.ac.uk/context/IRinContext_WorkshopNotes_SIGIR2004.pdf%`.

9. I. Kang and G. Kim. Query type classification for web document retrieval. In *SIGIR 2003*, pages 64–71, 2003.

10. D. Kelly and C. Cool. Effects of topic familiarity on information search behavior. In *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2002*, pages 74–75, 2002.

11. B. Larsen, S. Malik, and A. Tombros. The Interactive Track at INEX 2005. In N. Fuhr, M. Lalmas, S. Malik, and G. Kazai, editors, *INEX 2005 Workshop Proceedings*, Dagstuhl, Germany, 2005.