

User Needs and Strategies in Structured Information Retrieval

Georgina Ramírez

CWI, Amsterdam
georgina@cwi.nl

Abstract. Structured information retrieval studies the combination of the content and the structure information of documents to perform different IR tasks. Different approaches make use of the structural information of documents to improve information retrieval effectiveness. However, most of these studies do not take the user into account and they use the same strategy to perform all types of queries. This work aims to identify the relationship between user task and strategies on the usage of structural information. The theoretical part of this research consists of three main phases designed to acquire a good understanding of (1) the nature of user tasks on structured documents, (2) the types of structural information and its role in retrieval strategies, and (3) the formalization of a model to correlate both, user tasks and strategies. Two different experimental studies are planned. The first one investigates the combination of evidence from different strategies for the defined user tasks, and the second one investigates how to use relevance feedback techniques to refine the structural information for a given user need.

1 Introduction

The growing amount of structured information available, e.g., web pages and XML documents, poses interesting new challenges to different information seeking research communities like digital libraries or information retrieval. On one hand, the structure of the documents provide a new source of information that retrieval systems may exploit to improve their search effectiveness. On the other hand, the appearance of new query languages that work on structure provides the users with a more powerful tool to express complex and specific needs. This last aspect can be an important factor in the retrieval process because the knowledge that users may have of the structure of the documents, could lead to different interpretations of the structural constraints of the query.

In a similar way as done in other areas of information retrieval, like web search, this work studies the relationship between different user tasks and different strategies on the usage of structural information.

The main hypothesis of the present work is that the different types of information that can be extracted from the structural components of documents can be treated as multiple sources of information. The best combination of evidence from these sources will be determined by the different types of user tasks and intentions.

2 Research plan

The research presented is divided into five steps:

1. **Definition of a taxonomy of user tasks.** Based on the hypothesis that the type of search tasks performed by users on structured collections are the same as in those on plain text documents, we plan to investigate if we can apply existing categorizations of user tasks to structured information retrieval. A possible extension might be needed in order to include the different degrees of knowledge users have of the structure of the documents.
2. **Definition of a taxonomy of types and uses of structure.** The hypothesis here is that the different types of structure and their potential use in search can be clustered into different views (dimensions), e.g., semantic or organizational. We investigate if we can define a fixed set of independent dimensions, and use this set to classify the existing (and maybe new) retrieval strategies.
3. **Formalization of a model to correlate user tasks and strategies.** This part aims to formalize a prescriptive model that correlates, based on their properties, different user needs with different retrieval strategies for each possible dimension. The hypothesis here is that different user needs will require the use of different structural features.
4. **Experimentation on the combination of evidence from different dimensions.** Different experiments need to be done in order to estimate the best way to combine the evidence from the different dimensions used in the different user tasks. A hypothesis in this part of the research is that the problem we address of combination of evidence is similar to the one of combining evidence from different classifiers [1].
5. **Experimentation on the use of structural relevance feedback.** This experimental part of the research, intends to study if relevant structural information can be used to update the parameters of the combination of evidence during a relevance feedback process, in the same way as content is refined when applying traditional relevance feedback techniques.

3 Conclusions

The work presented aims to study and develop an experimental approach to correlate user needs and strategies for structured information retrieval. In particular, it investigates the use of structural features for effective information retrieval, and defines a model to link user tasks to different retrieval strategies.

References

1. W. B. Croft. *Combining approaches to information retrieval*, chapter 1, pages 1–36. Kluwer Academic Publishers, 2000.