Theoretical Benchmarks of XML Retrieval

Tobias Blanke
Queen Mary College, University of London
London, United Kingdom
tobias@dcs.qmul.ac.uk

Mounia Lalmas
Queen Mary College, University of London
London, United Kingdom
mounia@dcs.qmul.ac.uk

ABSTRACT

This poster investigates the use of theoretical benchmarks to describe the matching functions of XML retrieval systems and the properties of specificity and exhaustivity in XML retrieval. Theoretical benchmarks concern the formal representation of qualitative properties of IR models. To this end, a Situation Theory framework for the meta-evaluation of XML retrieval is presented.

Categories and Subject Descriptors: H.3.3 Information Search and Retrieval

General Terms: Theory, Measurement

Keywords: Meta-evaluation, XML retrieval

1. INTRODUCTION

The aim of XML retrieval is to retrieve not only relevant document components, but those at the right level of granularity, i.e. document components that specifically answer a query. To evaluate how effective XML retrieval approaches are, it is therefore necessary to consider whether the 'right' level is correctly identified. In 2004, INEX, the evaluation initiative for XML retrieval, used two four-graded dimensions of relevance: (1) exhaustivity reflects to which extent the document component satisfies the information need, and (2) specificity refers to the extent to which all the information in the document component is about the information need. A scale from 0 to 3 is used to measure how specific or exhaustive a document component is in relation to an information need, where 0 means that the effect is not measurable and 3 means that the component is highly specific/exhaustive. E.g., a (3,3) result designates a highly exhaustive and specific answer.

Van Rijsbergen [6] suggested, that given the increasing complexity of the retrieval task due to more complex information units like XML elements, an experimental approach to information retrieval (IR) should be complemented with a theoretical evaluation technique. Therefore, this poster will investigate the use of theoretical benchmarks to describe the matching functions of XML retrieval systems and the properties of specificity and exhaustivity.

2. THEORETICAL EVALUATION IN IR

Theoretical evaluation can be complementary to an experimental evaluation if it helps to clarify the assumptions of retrieval models and if it can identify the characteristics leading to a particular experimental behaviour. INEX has used various relevance scales; for demonstration purposes we use the INEX 2004 scale [5].

A theoretical evaluation can be done through the use of a meta-theory, as proposed in previous work based on the logical approach to IR [4]. In 1971, Cooper coined the term 'logical relevance' for an objective view on relevance [3]. Van Rijsbergen and others have expressed the logical relevance in terms of the implication $d \rightarrow q$ [6]. Following Huibers' formalism and approach [4], we call such an implication between query and document 'aboutness'. With aboutness, we aim to theoretically capture the benchmarks of an IR model in general and an XML retrieval model in particular.

Theoretical benchmarks [7] concern the formal representation of qualitative properties of IR models. The properties are described in terms of supported logical axioms and postulates. We use Situation Theory (ST), developed by Barwise and Perry [1], as our logic-based model for XML retrieval. ST offers a logic of information rather than truth assignments and is therefore closer to real-world applications. ST is a mathematical theory of meaning and information with situations as primitives. Situations are partial descriptions of the world and are composed of information items formalised as infons [4]. For IR modelling, queries and documents are modelled as situations, while infons represent a model's information items like keywords or phrases.

Theoretical benchmarks need formalisms, powerful enough to mark the fundamental properties of retrieval models. In this poster we only present the fundamentals of the formalism of a ST-based aboutness language for XML retrieval. If we consider XML elements to be XML situations, then $S \,\Box\!\!\rightsquigarrow\, T$ means that the XML situation S is about T. In order to describe an XML retrieval model, this could express that document component situation S is about the information need in query T. Likewise $S \,\Box\!\!\not\rightsquigarrow\, T$ symbolises that S is not about T. With $\odot$ we formalise the composition of situations. Preclusion, symbolised by $\perp$, expresses that information clashes and leads to anti-aboutness, for which the symbol $\boxtimes\!\!\rightsquigarrow$ is used. $\equiv$ states that two situations are equivalent, e.g. two document components containing the same information.

3. THEORETICAL BENCHMARKS OF XML RETRIEVAL

This section describes how we can apply a theoretical benchmark, based on ST, to show how XML approaches provide exhaustive or specific answers to queries. Firstly,

SIGIR'06, August 6–10, 2006, Seattle, Washington, USA.
ACM 1-59593-369-7/06/0008.

Table 1: INEX exhaustivity and specificity situations

Scale	*Exhaustivity* $D \Box\rightsquigarrow Q$	*Specificity* $Q \Box\rightsquigarrow D$
0	$\frac{D_1 \Box\not\rightsquigarrow Q, ..., D_n \Box\not\rightsquigarrow Q}{D \Box\not\rightsquigarrow Q}$	$\frac{Q_1 \Box\not\rightsquigarrow D, ..., Q_m \Box\not\rightsquigarrow D}{Q \Box\not\rightsquigarrow D}$
1	$\frac{D_1 \boxtimes\not\rightsquigarrow Q}{D \boxtimes\not\rightsquigarrow Q}$, ... , $\frac{D_i \Box\rightsquigarrow Q}{D \Box\rightsquigarrow Q}$, ... , $\frac{D_n \boxtimes\not\rightsquigarrow Q}{D \boxtimes\not\rightsquigarrow Q}$	$\frac{Q_1 \boxtimes\not\rightsquigarrow D}{Q \boxtimes\not\rightsquigarrow D}$, ... , $\frac{Q_i \Box\rightsquigarrow D}{Q \Box\rightsquigarrow D}$, ... , $\frac{Q_m \boxtimes\not\rightsquigarrow D}{Q \boxtimes\not\rightsquigarrow D}$
2	$\frac{D_1 \boxtimes\not\rightsquigarrow Q, ..., D_i \Box\rightsquigarrow Q, ..., D_n \boxtimes\not\rightsquigarrow Q}{D \Box\rightsquigarrow Q}$	$\frac{Q_1 \boxtimes\not\rightsquigarrow D, ..., Q_i \Box\rightsquigarrow D, ..., Q_m \boxtimes\not\rightsquigarrow D}{Q \Box\rightsquigarrow D}$
3	$\frac{D_1 \Box\rightsquigarrow Q, ..., D_n \Box\rightsquigarrow Q}{D \Box\rightsquigarrow Q}$	$\frac{Q_1 \Box\rightsquigarrow D, ..., Q_n \Box\rightsquigarrow D}{Q \Box\rightsquigarrow D}$

we look at the properties of XML aboutness systems that result in either higher exhaustivity or specificity. Secondly, we describe with our formalism the two dimensions of the XML retrieval relevance assessment: exhaustivity and specificity.

Some logical properties of an XML retrieval model support exhaustivity, while others support specificity. The monotonicity postulates are an example of logical properties. They claim that aboutness is preserved under composition. E.g. Left Monotonic Union (LMU) states that if situation S is about T, then $S \odot U$ is also about T.

$$\frac{S \Box\rightsquigarrow T}{S \odot U \Box\rightsquigarrow T}$$

In an aboutness model unconditionally supporting LMU, a query situation containing 'house' does not only lead to document component situations having 'house', but equally valid answers are components with 'house' and 'garden'. However, the right level of granularity is missed. A naive vector space model based on simple overlap supports both left and right monotonic union [4] and cannot lead to the retrieval of highly specific answers. Being able to provide specific answers is only possible from models supporting LMU only conditionally, as for example the vector space models with trained parameters or probabilistic models do [7]. Such models regularly achieve the best results at INEX. Yet, models not supporting LMU at all, neglect exhaustivity.

Using our ST-based framework, we can formally represent the two relevance dimensions used in INEX and their scale. In earlier work, Chiaramella [2] demonstrated that to capture the relevance of structured documents, two implications like those above from Rijsbergen should be used: $d \rightarrow q$ modelling exhaustivity and $q \rightarrow d$ modelling specificity. Table 1 shows how we can express these implications with our ST framework. D stands for the document component situation, and Q for the query situation. Then, $D \Box\rightsquigarrow Q$ models exhaustivity and $Q \Box\rightsquigarrow D$ models specificity.

So far, we have discussed how a theoretical evaluation could help to understand experimental results for XML retrieval. We continue by modelling how a user perceives how exhaustive and specific a component is. We argue that a user will assess the relevance of a component according to the information contained in both Q and D. In Table 1, the document component and query situations are divided into sub-situations, with $D \equiv D_1 \odot ... \odot D_n$ and $Q \equiv Q_1 \odot ... \odot Q_m$. Scale 1 states that only some subsituations are about the query but none involves anti-aboutness. With this, we can e.g. formalise what is meant by a marginally exhaustive document component (1): the topic is only mentioned in passing, leading to very small indications about the document component's relevance. Table 1 shows for scale 1 multiple conclusions demonstrating undecidedness about the component's relevance. For scale 2 the overall conclusion can be derived that $D \Box\rightsquigarrow Q$ or $Q \Box\rightsquigarrow D$. For specificity, the topic is a major theme of the document component and $Q \Box\rightsquigarrow D$ can be concluded. Scale 0 indicates that no D_i is about the query making the whole document component not about the query. The highest satisfaction is achieved with scale 3. For exhaustivity, all subsituations of the component are about the query, while for specificity all subsituations of the query are about the document component. Looking at combined exhaustivity and specificity assessments in Table 1, users focussed on specificity want results like (1,3), (2,3) or (3,3). They want to be able to conclude $Q \Box\rightsquigarrow D$, but care less about how many document component subsituations are about the query situation. A (3,3) result is achieved if all subsituations of the query are about the document component situation and vice versa. Therefore, (3,3) describes a perfect match.

4. CONCLUSIONS AND FUTURE WORK

A ST-based meta-evaluation of XML retrieval can firstly demonstrate benchmarks of XML retrieval models that make them appropriate to particular tasks, and secondly formally represent the exhaustivity and specificity of document components. In the future, we would like to continue our work by elaborating our benchmarks and going into more detail regarding the different INEX metrics. A theoretical evaluation could assist in the difficult INEX discussion about the correct metric [5] and help understand better the implications of the two measures of exhaustivity and specificity.

5. REFERENCES

[1] J. Barwise and J. Perry. *Situations and Attitudes*. MIT Press, Cambridge, MA, 1982.

[2] Y. Chiaramella. Information retrieval and structured documents. In *Lectures on information retrieval*, pages 286–309. Springer-Verlag, New York, 2001.

[3] W. Cooper. A definition of relevance for information retrieval. *Information Storage and Retrieval*, 7:19–37, 1971.

[4] T. W. Huibers. *An Axiomatic Theory for Information Retrieval*. PhD-Thesis, Universiteit Utrecht, Utrecht, 1996.

[5] G. Kazai and M. Lalmas. Notes on what to measure in INEX. In *INEX 2005 Workshop on Element Retrieval Methodology*, Glasgow, July 2005.

[6] C. J. v. Rijsbergen. Towards an information logic, 1989. 77-86.

[7] D. Song, K.-F. Wong, P. Bruza, and C.-H. Cheng. Application of aboutness to functional benchmarking in information retrieval. *ACM Trans. Inf. Syst.*, 19(4):337–370, 2001.