Relevance

Jovan Pehcevski, Birger Larsen

▶ **To cite this version:**

Jovan Pehcevski, Birger Larsen. Relevance. Ling Liu and M. Tamer Özsu. Encyclopedia of Database Systems, Springer-Verlag, 2007. inria-00174160

HAL Id: inria-00174160

https://hal.inria.fr/inria-00174160

Submitted on 21 Sep 2007

Relevance

Jovan Pehcevski
INRIA Rocquencourt, France
jovan.pehcevski@inria.fr

Birger Larsen
Royal School of Library and Information Science, Denmark
blar@db.dk

SYNONYMS

Exhaustivity; Specificity; Topical-Hierarchical Relevance.

DEFINITION

Relevance is the extent to which some information is pertinent, connected, or applicable to the matter at hand. It represents a key concept in the fields of documentation, information science, and information retrieval.

In information retrieval, the notion of relevance is used in three main contexts. Firstly, an algorithmic relevance score is assigned to a search result (usually a whole document) representing an estimated likelihood of relevance of the search result to a topic of request. This relevance score often determines the order in which search results are presented to the user. Secondly, when the performance of information retrieval systems is tested experimentally, the retrieved documents are assessed for their actual relevance to the topic of request by human assessors (topic experts). A binary relevance scale is typically used to assess the relevance of the search result, where the relevance is restricted to be either zero (when the result is not relevant to the user request) or one (when the result is relevant). Thirdly, in experiments involving users (or in operational settings) a broader notion of relevance is often used, with the aim of expressing the degree to which the retrieved documents are perceived as useful in solving the user's search task.

In semi-structured text (XML) retrieval, the search result is typically an XML element, and the relevance score assigned by an XML retrieval system again represents an estimated likelihood of relevance of the search result to the topic of request. However, when the results are subsequently assessed for relevance, the binary relevance scale is not sufficient, primarily due to the hierarchical relationships that exist among the elements in an XML document. Accordingly, in XML retrieval one or more relevance dimensions (each with a multi-graded relevance scale) have been used to assess the relevance of the search result.

MAIN TEXT

In traditional information retrieval experiments where whole documents are retrieved, a fairly simple notion of relevance may suffice for most purposes [1]. The challenge in XML retrieval is that the relevance assessments must capture not only whether the retrieved elements are relevant, but also how they relate to one another. Specifically, the hierarchical relationships between overlapping elements necessitates a more complex relevance definition. For instance, if a subsection of a document contains sufficient information to solve the search task (i.e. it is highly relevant), the containing section must also be relevant, but perhaps less so because it may also contain other information that is not much useful for solving the search task.

The different relevance definitions for XML retrieval have mainly been investigated by the INitiative for the Evaluation of XML Retrieval (INEX). From 2002 until 2005, two relevance dimensions were used as part of the INEX relevance definition, whereas from 2006 onwards only one INEX relevance dimension is used. Overall, the development has been a move from using multi-graded discrete relevance scales to the use of continuous scales.

Where the early definitions required a large intellectual effort, with the later definitions the assessment process was made less demanding by the possibility of a software assessment support. In addition to this use of relevance by expert assessors, a number of interactive user experiments have also been carried out in INEX providing a user perspective on relevance in XML retrieval.

INEX 2002 relevance definition

In 2002, the INEX relevance definition comprised two relevance dimensions named *topical relevance* and *component coverage*. This relevance definition has not been used by INEX since then, partly because of the vague terminology used for the names of the two relevance dimensions, and partly because it has been subsequently shown that the INEX 2002 assessors did not fully comprehend component coverage [2].

INEX 2003 and 2004 relevance definition

In 2003 and 2004, the two INEX relevance dimensions were named Exhaustivity and Specificity, which respectively reflect the extent that an element *covers* and is *focussed on* aspects of an information need represented by the topic of request. The two INEX relevance dimensions used four grades to assess the relevance of an element (either its exhaustiveness or its specificity): "none", "marginally", "fairly", and "highly". The grades from each dimension were then combined into a single 10-point relevance scale.

Studies have shown that the INEX relevance definition used in 2003 and 2004, comprising two dimensions based on topical relevance, was too hard for users to understand and relate to [9, 10]. A related study also revealed that, although the assessors did agree (to a certain extent) on the relevant documents found for an INEX topic, they did not seem to agree on the relevant elements within these documents, mostly because of the various combinations of relevance grades that can be assigned to relevant elements [12]. It has also been observed that a complex relevance scale, such as the INEX 10-point relevance scale, could lead to an increased level of obtrusiveness in interactive user environments [4].

INEX 2005 relevance definition

From 2005 onwards, a highlighting assessment procedure is used at INEX to gather relevance assessments for the XML retrieval topics. The highlighting assessment procedure had two main steps: the assessor was first required to highlight the relevant content in each returned document, and after the assessment tool automatically identified the elements that enclosed the highlighted content, the assessor was then asked to judge the Exhaustivity of these elements, and of all their ancestors and descendants. As a result, the INEX relevance definition was simplified in 2005, such that only three Exhaustivity values were assigned to a relevant element, while the Specificity of the relevant element was measured on a continuous relevance scale and computed automatically as the ratio of highlighted to fully contained text.

Mostly due to the highlighting assessment procedure applied at INEX in 2005, the common assessor understanding of the INEX 2005 relevance definition was different than that of the two previous INEX relevance definitions. Specifically, an analysis of the level of assessor agreement on five topics double-judged at INEX 2005 suggested that there is a good reason for dropping the Exhaustivity relevance dimension, since it appeared to be easier for the assessor to be consistent when highlighting relevant content than when choosing one of the three Exhaustivity values [8]. Dropping the Exhaustivity dimension did not only result in a reduced cognitive load on the assessor, but more importantly, this also did not significantly affect the measured evaluation outcome [6].

Current INEX relevance definitions

The above findings indicated that a much simpler relevance scale, and therefore, a much simpler relevance definition, would be a preferable choice for INEX and the field of XML retrieval. As a result, from 2006 onwards relevance in INEX is defined only according to the notion of Specificity, where a continuous relevance scale is used to assess the relevance of retrieved elements.

The experience of both assessors and users is important when defining relevance in XML retrieval. An interactive track was established at INEX in 2004 to investigate the behaviour of users when elements of XML documents (rather than whole documents) are presented as answers [11]. The interactive track was run again at INEX 2005 and 2006, comprising various tasks and different XML document collections [3, 5]. A *topical-hierarchical* relevance

definition was used by the INEX Interactive tracks in 2005 and 2006, which comprises two relevance dimensions and a five-point nominal relevance scale [5, 7]. The first *topical* relevance dimension determines the extent to which an XML element contains relevant information for the search task. It can take one of the following three values: highly relevant, relevant, or not relevant. The second *hierarchical* relevance dimension determines the extent to which an XML element needs the context of its containing XML document to make full sense as an answer. It can take one of the following three values: just right, too large, or too small. As a result, the final assessment score of an XML element can take one of the following five nominal values: "exact", if the XML element is just right and highly relevant; "partial", if the XML element is just right and relevant; "broad", if the XML element is too large and relevant; "narrow", if the XML element is too small and highly relevant; and "not relevant", if the XML element does not contain relevant information.

An analysis of the feedback gathered from users participating in the INEX 2005 Interactive track showed that users did not find the five-point scale to be very hard to understand [7]. Furthermore, a mapping between the five-point relevance scale (used by users) and the continuous Specificity scale (used by the expert assessors) can easily be established [7], which allows for a better understanding of the definition of relevance in XML retrieval.

CROSS REFERENCE
Evaluation metrics; Specificity.

RECOMMENDED READING

[1] P. Borlund. The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10):913–925, 2003.

[2] G. Kazai, S. Masood, and M. Lalmas. A study of the assessment of relevance for the INEX 2002 test collection. In *Proceedings of the 26th European Conference on IR Research (ECIR)*, pages 296–310, Sunderland, UK, 2004.

[3] B. Larsen, S. Malik, and A. Tombros. The Interactive track at INEX 2005. In *Advances in XML Information Retrieval and Evaluation: Fourth Workshop of the INitiative for the Evaluation of XML Retrieval, INEX 2005*, volume 3977 of *Lecture Notes in Computer Science*, pages 398–410, 2006.

[4] B. Larsen, A. Tombros, and S. Malik. Obtrusiveness and relevance assessment in interactive XML IR experiments. In *Proceedings of the INEX 2005 Workshop on Element Retrieval Methodology*, pages 39–42, Glasgow, UK, 2005.

[5] S. Malik, A. Tombros, and B. Larsen. The Interactive track at INEX 2006. In *Comparative Evaluation of XML Information Retrieval Systems: Fifth Workshop of the INitiative for the Evaluation of XML Retrieval, INEX 2006*, volume 4518 of *Lecture Notes in Computer Science*, pages 387–399, 2007.

[6] P. Ogilvie and M. Lalmas. Investigating the Exhaustivity dimension in content-oriented XML element retrieval evaluation. In *Proceedings of the Fifteenth ACM Conference on Information and Knowledge Management (CIKM '06)*, pages 84–93, Arlington, USA, 2006.

[7] J. Pehcevski. Relevance in XML retrieval: The user perspective. In *Proceedings of the SIGIR 2006 Workshop on XML Element Retrieval Methodology*, pages 35–42, Seattle, USA, 2006.

[8] J. Pehcevski and J. A. Thom. HiXEval: Highlighting XML retrieval evaluation. In *Advances in XML Information Retrieval and Evaluation: Fourth Workshop of the INitiative for the Evaluation of XML Retrieval, INEX 2005*, volume 3977 of *Lecture Notes in Computer Science*, pages 43–57, 2006.

[9] J. Pehcevski, J. A. Thom, and A-M. Vercoustre. Users and assessors in the context of INEX: Are relevance dimensions relevant? In *Proceedings of the INEX 2005 Workshop on Element Retrieval Methodology*, pages 47–62, Glasgow, UK, 2005.

[10] N. Pharo and R. Nordlie. Context matters - an analysis of assessments of XML documents. In *Proceedings of the Fifth Conference on Conceptions of Library and Information Science (COLIS)*, pages 238–248, Glasgow, UK, 2005.

[11] A. Tombros, B. Larsen, and S. Malik. The Interactive track at INEX 2004. In *Advances in XML Information Retrieval: Third International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2004*, volume 3493 of *Lecture Notes in Computer Science*, pages 410–423, 2005.

[12] A. Trotman. Wanted: Element retrieval users. In *Proceedings of the INEX 2005 Workshop on Element Retrieval Methodology*, pages 63–69, Glasgow, UK, 2005.