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 or work 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.
The different relevance definitions for XML retrieval have mainly been investigated by the INitiative for the Evaluation of XML Retrieval (INEX). 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. 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.
From 2005 onwards, a highlighting assessment procedure is used at INEX to gather relevance assessments for the XML retrieval topics. As a result, the INEX relevance definition was simplified 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. From 2006 onwards, relevance in INEX is defined only according to the notion of Specificity, where the 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. The interactive track was run again at INEX in 2005 and 2006, comprising various tasks and different XML document collections [2].
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 [2, 3]. 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 [3]. 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 [3], which allows for a better understanding of the definition of relevance in XML retrieval.

CROSS REFERENCE

Evaluation metrics; Specificity.

REFERENCES

[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] 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.

[3] 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.