



HAL
open science

Specificity

Jovan Pehcevski, Benjamin Piwowarski

► **To cite this version:**

Jovan Pehcevski, Benjamin Piwowarski. Specificity. Ling Liu and M. Tamer Özsu. Encyclopedia of Database Systems, Springer-Verlag, 2007. inria-00174155

HAL Id: inria-00174155

<https://inria.hal.science/inria-00174155>

Submitted on 21 Sep 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Specificity

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

Benjamin Piwowarski
Yahoo! Research Latin America
bpiwowar@yahoo-inc.com

SYNONYMS

Coverage.

DEFINITION

Specificity is a relevance dimension that describes the extent to which a document part focuses on the topic of request. In the context of semi-structured text (XML) retrieval, a document part corresponds to an XML element. Specificity is defined as the length ratio, typically in number of characters, of contained relevant to irrelevant text in the document part.

Different Specificity values can be associated to a document part. These values are drawn from the Specificity relevance scale, which has evolved from a discrete multi-graded relevance scale to a continuous relevance scale.

MAIN TEXT

The Initiative for the Evaluation of XML Retrieval (INEX) has defined Specificity as a relevance dimension that uses values from its own relevance scale to express the extent to which an XML element focuses on the topic of request. Since 2002, different names and relevance scales were used for Specificity at INEX. It initially evolved because the relevance dimension was not sufficiently well defined, and later because the assessment procedure changed.

In 2002, Specificity was named *coverage* at INEX, which reflected the extent to which an XML element was focussed on aspects of the information need (as represented by the INEX topic). The component coverage used a relevance scale comprising four relevance grades, from “no coverage”, “too large”, “too small”, to “exact coverage”. However, this dimension was used solely in 2002, partly because of the vagueness introduced in the terminology for its name, and partly because it has been subsequently shown that the INEX 2002 assessors did not particularly understand the notion of “too small” [1]. In particular, assessors understood “too small” as a measure of quantity while Specificity is more related to the concentration of relevant information.

In 2003 and 2004, four grades were used for the Specificity relevance dimension at INEX, such that the extent to which an XML element may focus on the topic of request could range from “none” (0), to “marginally” (1), to “fairly” (2), or to “highly” (3) focused. An XML element was considered relevant only if its Specificity value was greater than zero.

From 2005 onwards, a highlighting assessment procedure is used at INEX to gather relevance assessments for the XML retrieval topics. The Specificity of an XML element is automatically computed as the ratio of highlighted to fully contained text, where the relevance values that can be associated to the element are drawn from a continuous relevance scale. These values are in the range between 0 and 1, where the value of 0 corresponds to an element that does not contain any highlighted text, while the value of 1 corresponds to a fully highlighted element.

With the highlighting assessment procedure, assessors are asked to highlight all the relevant information contained by returned XML documents. This results in a reduced cognitive load on the assessor, since in this case there is no need for the assessor to explicitly associate a Specificity value to a judged element. Studies of the level of assessor agreement, which used topics that were double-judged at INEX, have shown that the use of the new highlighting procedure further increases the level of assessor agreement compared to the level of agreement observed among assessors during previous years at INEX [2, 3].

CROSS REFERENCE

Relevance; Evaluation metrics.

RECOMMENDED READING

- [1] 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.
- [2] 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, pages 43–57, 2006.
- [3] A. Trotman. Wanted: Element retrieval users. In *Proceedings of the INEX 2005 Workshop on Element Retrieval Methodology*, pages 63–69, Glasgow, UK, 2005.