Relevance as a Subjective and Situational Multidimensional Concept

Carsten Eickhoff
Delft University of Technology
Delft, Netherlands
c.eickhoff@tudelft.nl

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: [Search Process]; H.1.2 [**Models and Principles**]: [Human Factors]

Keywords

Relevance Models, Personalisation, Score Combination

Abstract

Relevance is the central concept of information retrieval. Although its important role is unanimously accepted among researchers, numerous different definitions of the term have emerged over the years (e.g., [1, 2]). Considerable effort has been put into creating consistent and universally applicable descriptions of relevance in the form of relevance frameworks. Examples of such frameworks are given by [2, 4]. Across these various formal systems of relevance, a wide range of relevance criteria has been identified. The probably most frequently used single criterion, that in some applications even becomes a synonym for relevance, is *topicality*. It expresses a document's topical overlap with the user's information need. For textual resources, it is often estimated based on term co-occurrences between query and document. There is, however, a significant number of further noteworthy relevance criteria. Prominent specimen as for example compiled by [7, 5] are: (**Currency**) determines how recent and up to date the document is. Outdated information may have become invalid over time. (**Availability**) expresses how easy it is to obtain the document. Users might not want to invest more than a threshold amount of resources (e.g., disk space, downloading time or money) to get the document. (**Readability**) describes the document's readability and understandability. A document with a high topical relevance towards a given information need can become irrelevant if the user is not able to extract the desired information from it. (**Credibility**) contains criteria such as the document author's expertise, the publication's reputation and the document's general trustworthiness. (**Novelty**) describes the document's contribution to satisfying an information need with respect to the user's context. E.g., previous search results or general knowledge about the domain.

It is evident that these criteria can have very different scopes. Some of them are static characteristics of the document or the author, others depend on the concrete informa-

tion need at hand or even the user's search context. Currently, state-of-the-art retrieval models often treat relevance (regardless which interpretation of the term was chosen) as an atomic concept that can be expressed through topical overlap between document and query or a plain linear combination of multiple scores [6]. Considering the broad audiences a web search engine has to serve, such a method does not seem optimal as the concrete composition of relevance will vary from person to person depending on social and educational context. Furthermore, each individual can be expected to have situational preference for certain combinations of relevance facets depending on the information need at hand. We investigate combination schemes which respect the dimension-specific relevance distributions. In particular, we developed a risk-aware method of combining relevance criteria inspired by the economic Portfolio theory. As a first stage, we applied this method for result set diversification across dimensions.

To evaluate our methods, we will work with 2 large-scale standard corpora: (1) The Blogs'08 corpus [3], which contains blog posts and related user comments. At the example of this collection, we will investigate criteria such as topicality and currency for satisfying informational queries. (2) The upcoming ClueWeb'12 corpus that is currently being compiled is envisioned to be annotated with additional labels and scores such as the well-known readability measures. Combining readability scores and topicality, we will investigate the use case of inexperienced readers who prefer easy-to-read resources while depending on topical relevance. The evaluation will be based on the notion of usefulness of shown results for users. Starting off from simulated audiences with given preference patterns, we will finally conduct a real-life study with different participant types (e.g., adults and children).

1. REFERENCES

[1] M.J. Bates. Information search tactics. *JASIS*, 30(4), 1979.
[2] P. Borlund. The concept of relevance in IR. *JASIST*, 2003.
[3] C. Macdonald, I. Ounis, and I. Soboroff. Overview of the trec blog track 2009. In *TREC 2009, Proceedings*, 2009.
[4] S. Mizzaro. Relevance: The whole history. *JASIS*, 1997.
[5] T.K. Park. The nature of relevance in information retrieval: An empirical study. *The library quarterly*, 63(3), 1993.
[6] S. Robertson, H. Zaragoza, and M. Taylor. Simple BM25 extension to multiple weighted fields. In *CIKM 2004*.
[7] L. Schamber and J. Bateman. User criteria in relevance evaluation: Toward development of a measurement scale. In *American Society for Information Science*, 1996.