

Influences on the relevance judgment process in academic search systems

Christiane Behnert
Hamburg University of Applied Sciences
Department of Information
Hamburg, Germany
christiane.behnert@haw-hamburg.de

Relevance is still considered the core concept of information science today, although there does not exist a definition or theory of relevance in information science (Saracevic, 2016a, 2016b; White, 2009). Decades of research on relevance has formed our understanding of this very complex, multidimensional and dynamic, human concept (Borlund, 2003; Mizzaro, 1997; Saracevic, 2006, 2016b). It is involved in human information seeking and search behaviour as well as it is the dominating concept behind (interactive) information retrieval (IIR) systems evaluations.

In retrieval studies, relevance judgments are obtained by human jurors to measure a system's retrieval effectiveness to improve the system which in turn will better serve the system's users in their information seeking interactions. However, relevance judgments suffer from inconsistencies with regard to inter-rater and intra-rater reliability, i.e., judgments differ across judges and also at different times (Buckley & Voorhees, 2005), emphasising the highly subjective and context-dependent nature of relevance. Thus, there are many influences involved in the process of performing relevance judgments. As the relevance judgement process can be "defined as the sequential use of relevance criteria as delimited by interactions" (Beresi, Kim, Song, Ruthven, & Baillie, 2010, p. 199), it is essential to understand what these criteria are exactly and how they are applied. An extensive literature review of studies on relevance criteria revealed gaps at content and a methodological level:

(A) There seem to be no standardised definitions of the concepts of clues, criteria, factors that allow for clear distinctions of the concepts. The three terms are very often treated differently in the literature, which makes it difficult to identify actual criteria.

(B) Famous studies on relevance criteria had been undertaken in the 1990s (e.g., Bateman, 1998; Schamber, 1991). Since then, result presentation has changed. For example, today's academic search systems integrate additional elements into their search results, such as the number of times a work had been cited or downloaded. This kind of data can be referred to as popularity data since they serve as a factor of popularity which is assumed to imply a certain degree of quality of the work. The results study participants had judged in past studies on relevance criteria did not include such additional data.

(C) To learn about actual human behaviour, behavioural scientists or psychologists conduct experiments. IIR research would also benefit from employing experimental designs to gain a deeper understanding of users' actual behaviour when interacting with the system (Kelly & Cresenzi, 2016).

Only very few studies that investigate relevance criteria employ an experimental design that meets the requirements for experiments from a social science or behavioural research perspective, i.e., manipulating variables that are assumed to cause an observable effect, and control for possible confounding variables mainly through randomisation (Kelly, 2009; Sedlmeier & Renkewitz, 2018).

To address the first-mentioned gap, I developed a user model on predictive relevance judgments in academic search systems (Behnert, 2019b). In this model, clues within a surrogate (e.g., publication date) are defined as operationalised relevance criteria (e.g., currency) along with user-based, system-based and situation-based relevance factors as influencing variables in the judgment process. This model offers a structured, holistic view of clues to relevance, criteria and factors, and, at the same time, it suggests an approach to achieve a more explicit definition of the terms clues, criteria and factors.

To address the second and third gaps, I developed an online experiment to investigate the effects of popularity data on predictive relevance judgments in academic search systems (Behnert, 2019a). During the experiment, participants are asked to perform judgments of surrogates that include, for example, manipulated citations and download counts. At the time of preparing this submission, a pretest is being conducted, while the actual online experiment will take place in summer 2019.

References

- Bateman, J. (1998). Changes in Relevance Criteria: A Longitudinal Study. In R. Larson, K. Petersen, & C. M. Preston (Eds.), *Proceedings of the 61st ASIS Annual Meeting* (pp. 23–32). Information Today.
- Behnert, C. (2019a). Investigating the effects of popularity data on predictive relevance judgments in academic search systems. In *Proceedings of ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR'19) March 10–14, 2019, Glasgow, United Kingdom*. New York, NY, USA: ACM. <https://doi.org/10.1145/3295750.3298978>
- Behnert, C. (2019b). Kriterien und Einflussfaktoren bei der Relevanzbewertung von Surrogaten in akademischen Informationssystemen. *Information - Wissenschaft & Praxis*, 70(1), 24–32. <https://doi.org/10.1515/iwp-2019-0002>
- Beresi, U. C., Kim, Y., Song, D., Ruthven, I., & Baillie, M. (2010). Relevance in Technicolor. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6273 LNCS, pp. 196–207). https://doi.org/10.1007/978-3-642-15464-5_21
- Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10), 913–925. <https://doi.org/10.1002/asi.10286>

Buckley, C., & Voorhees, E. M. (2005). Retrieval System Evaluation. In E. M. Voorhees & D. K. Harman (Eds.), *TREC: Experiment and Evaluation in Information Retrieval* (pp. 53–75). Cambridge, MA ; London, UK: MIT Press.

Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval*, 3(1—2). <https://doi.org/10.1561/15000000012>

Kelly, D., & Cresenzi, A. (2016). From design to analysis: Conducting controlled laboratory experiments with users. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '16* (pp. 1207–1210). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2911451.2914809>

Mizzaro, S. (1997). Relevance: The whole history. *Journal of the American Society for Information Science*, 48(9), 810–832. [https://doi.org/10.1002/\(SICI\)1097-4571\(199709\)48:9<810::AID-ASI6>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-4571(199709)48:9<810::AID-ASI6>3.0.CO;2-U)

Saracevic, T. (2006). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II. *Advances in Librarianship*, 30(October), 3–72. [https://doi.org/10.1016/S0065-2830\(06\)30001-3](https://doi.org/10.1016/S0065-2830(06)30001-3)

Saracevic, T. (2016a). Relevance: In search of a theoretical foundation. In D. Sonnenwald (Ed.), *Theory development in the information sciences* (pp. 141–163). University of Texas Press.

Saracevic, T. (2016b). *The Notion of relevance in information science: Everybody knows what relevance is. But, what is it really?* (G. Marchionini, Ed.), *Synthesis Lectures on Information Concepts, Retrieval, and Services* (Vol. 8). Morgan & Claypool. <https://doi.org/10.2200/S00723ED1V01Y201607ICR050>

Schamber, L. (1991). User's criteria for evaluation in a multimedia environment. In *Proceedings of the 54th ASIS Annual Meeting* (pp. 126–133).

Sedlmeier, P., & Renkewitz, F. (2018). *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler* (3. Aufl.). Hallbergmoos: Pearson Deutschland.

White, H. D. (2009). Relevance in theory. In *Encyclopedia of Library and Information Sciences, Third Edition* (3rd ed., pp. 4498–4511). CRC Press. <https://doi.org/10.1081/E-ELIS3-120043266>

Citation suggestion

Behnert, C. (2019). Influences on the relevance judgment process in academic search systems. *Contribution at ASIS&T European Chapter Seminar on Information Science Trends: Search Engines and Information Retrieval*. April 26, 2019, Hamburg, Germany. doi: 10.5281/zenodo.3265731