Relevance Clues

Developing an Experimental Design to Examine the Criteria Behind Relevance Judgments

Christiane Behnert

Hamburg University of Applied Sciences, Germany
christiane.behnert@haw-hamburg.de

Abstract

When users of information systems judge the relevance of search results, diverse criteria beyond topical relevance come into play. In this paper, we introduce the doctoral project *Relevance Clues*, which, through an experimental design, seeks to gain significant knowledge on the criteria by which users make relevance judgments.

Keywords: relevance; information behavior; interactive information retrieval; experiments; user criteria

1 Introduction

In modern library information systems (LIS), data as indicators for popularity (e.g., the number of clicks on a document) are available due to search engine technology. Web search engines consider such data as indicators for the relevance of a search result, i.e., they include these data in their relevance ranking algorithms. The project *LibRank – New Approaches to Relevance Ran-*

In: M. Gäde/V. Trkulja/V. Petras (Eds.): Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science (ISI 2017), Berlin, 13th–15th March 2017. Glückstadt: Verlag Werner Hülsbusch, pp. 330–335.

ing in Library Information Systems[1] researched how such relevance factors can be adopted by libraries. Possible ranking factors suitable for LIS were identified (Behnert & Lewandowski, 2015) and systematically evaluated following standard procedures from Information Retrieval (IR) evaluation based on the methodological framework proposed by Tague-Sutcliffe (1992), which also, in the past, had been utilized for Web search evaluation using human relevance assessments (Lewandowski, 2012). Data as indicators for popularity were obtained from internal sources[2] (e.g., circulation data, the number of copies) and external sources (e.g., citation data), and were implemented in the test rankings in order to perform a total of three evaluation runs (Plassmeier et al., 2015). However, due to the research design and based on the assessment data, we do not know the criteria by which the jurors judged the documents.

At this point, the doctoral research project *Relevance Clues: Development and empirical examination of a model for relevance decisions on search results based on individual user criteria* (working title) sets in. The research goal is to gain significant knowledge on the criteria according to which users of academic search engines and LIS judge an information object (document surrogate) to be relevant to their individual information needs. In order to achieve this goal, an experimental research design will be applied.

2 Research questions

The project aims to answer the following research questions (RQs):

(I) What clues within a surrogate do users use to judge on its relevance?

(II) Which clues affect the relevance decision to what extent?

(III) What influence does the use situation (e.g., the user's location, time pressure) have on the relevance judgment?

(IV) What relevance criteria can be determined by the answers to RQ I–III, and how can they be weighed against each other?

1 www.librank.info

2 The data were obtained from *EconBiz*, an information portal for economics by the German National Library of Economics (www.econbiz.de).

The RQs are pursued in two steps: Based on an extensive literature review of current studies, a user model of relevance criteria will be developed and empirically examined through the conduction of a series of online experiments with human test persons. After reflecting the results, the user model will be altered accordingly.

3 Related research

3.1 Relevance criteria

In the 1990's, a shift from a system-oriented view towards a user-oriented view on relevance in Information Retrieval (IR) occurred, which is mirrored by studies on relevance criteria that were undertaken at this time (Mizzaro, 1997). In a recent publication, Saracevic (2016) synthesizes what we have learned from decades of relevance research. He provides an overview of 21 "observational, empirical, or experimental" studies on relevance clues, including the important works by Barry and Schamber (1998). They had analyzed criteria according to which users judge the relevance of a document. Among other studies, their results show that diverse criteria beyond topical relevance are involved in relevance judgments, for example, *validity*, *recency*, *availability*, and *credibility* of the information source.

With respect to the dynamic nature of the Web and its exponential growth, credibility and quality are both very important factors in order to filter and judge information retrieved by Web search engines (Rieh & Belkin, 1998). Credibility, in particular, can be considered in terms of *cognitive authority* (Wilson, 1983), which is highly subjective: A person is not only an expert but a cognitive authority as well, when his or her statements of knowledge are accepted by others as truth – the information is trustworthy – while he or she also influences other people's thoughts. Assessments of an author's cognitive authority are based on his or her present reputation and accomplishments (Rieh, 2009). Thus, in an academic context, information about an author's impact is helpful for users making relevance judgments. While general information about a document is presented as a surrogate, academic search systems integrate additional data into search results presenta-

tion. For example, Google Scholar's results include the number of citations or other versions of the particular work.

3.2 Document representations as objects of investigation

Documents have been represented by LIS using metadata ever since. Due to integrating electronic materials into modern systems and (1) full-text indexing as well as (2) enhancement with external data, for example, tables of contents and abstracts, users today are provided with a large amount of information about a document's content that they can utilize to judge its relevance. Further, Web 2.0 functionalities enable user comments and recommendations that can also be involved in the relevance decision process, as they indicate popularity. In academic information-seeking, popularity data would also include citation information, as mentioned above. So far, no studies on relevance assessments of surrogates including such popularity data and, at the same time, considering the user perspective have been published.

4 Methods

In order to examine which clues of a surrogate affect users' relevance judgments, a series of online experiments will be conducted. The essential characteristics of an experiment, as common in the field of psychology, are *manipulation* and *control*, which are the basic requirements for testing whether a causal relationship exists between a stimulus (independent variable) and an effect (dependent variable) (Sedlmeier & Renkewitz, 2007: 124–127). In relevance research so far, experimental research designs have not been applied very often, but this method follows the relatively recent trend in Interactive IR (IIR) that puts the user interacting with the system and the information objects in the center of retrieval evaluation (Kelly, 2009). Experiments are not only used for IIR evaluation but also to gain knowledge on user information behavior.

Within this project, the experiments will be conducted with students or academic staff using an online software tool. Subjects will be presented with a sequence of search results pages containing ten surrogates to a search query or information need in a randomized order. There will be two levels of poten-

tial relevance clues (of each independent variable) that will be manipulated (e.g., a low number of citations versus a large number of citations) to measure the effect on the relevance judgments (dependent variable). The clues to be tested will be selected based on the results of the literature review and a pre-test. Since it is assumed that there are diverse clues that affect relevance judgments, a multifactorial within-subjects design has to be developed. This requires a relatively large sample size. The goal is to recruit at least 400 subjects.

Acknowledgements

The research project presented in this paper is funded by a three-year Ph.D. scholarship from the Hamburg University of Applied Sciences, Germany. The doctoral candidate acknowledges her supervisors Prof. Dr. Ulrike Spree, Hamburg University of Applied Sciences, Germany, and Prof. Dr. Joachim Griesbaum, University of Hildesheim, Germany.

References

Barry, C. L. & Schamber, L. (1998): Users' criteria for relevance evaluation: A cross-situational comparison. In: *Information Processing & Management*, 34 (2–3), 219–236. https://doi.org/10.1016/S0306-4573(97)00078-2

Behnert, C. & Lewandowski, D. (2015): Ranking search results in library information systems – Considering ranking approaches adapted from web search engines. In: *The Journal of Academic Librarianship*, 41 (6), 725–735. https://doi.org/10.1016/j.acalib.2015.07.010

Kelly, D. (2009): Methods for evaluating interactive information retrieval systems with users. In: *Foundations and Trends® in Information Retrieval*, 3 (1–2). https://doi.org/10.1561/1500000012

Lewandowski, D. (2012): A framework for evaluating the retrieval effectiveness of search engines. In: C. Jouis, I. Biskri, J.-G. Ganascia, & M. Roux (Eds.): *Next Generation Search Engines* (pp. 456–479). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-4666-0330-1.ch020

Mizzaro, S. (1997): Relevance: The whole history. In: *Journal of the American Society for Information Science*, 48 (9), 810–832. https://doi.org/10.1002/(SICI)1097-4571(199709)48:9%3c810::AID-ASI6%3e3.0.CO;2-U

Plassmeier, K., Borst, T., Behnert, C. & Lewandowski, D. (2015): Evaluating popularity data for relevance ranking in library information systems. In: *Proceedings of the 78th ASIS&T Annual Meeting*. https://www.asist.org/files/meetings/am15/proceedings/submissions/posters/270poster.pdf

Rieh, S. Y. (2009): Credibility and cognitive authority of information. In: *Encyclopedia of Library and Information Sciences* (3rd ed.) (pp. 1337–1344). CRC Press. https://doi.org/10.1081/E-ELIS3-120044103

Rieh, S. Y. & Belkin, N. J. (1998): Understanding judgment of information quality and cognitive authority in the WWW. In: *Proceedings of the 61st ASIS Annual Meeting* (pp. 279–289).

Saracevic, T. (2016): *The notion of relevance in Information Science: Everybody knows what relevance is. But, what is it really?* Morgan & Claypool. https://doi.org/10.2200/S00723ED1V01Y201607ICR050

Sedlmeier, P. & Renkewitz, F. (2007): *Forschungsmethoden und Statistik in der Psychologie*. München, Boston et al.: Pearson Studium.

Tague-Sutcliffe, J. (1992): The pragmatics of information retrieval experimentation, revisited. In: *Information Processing & Management*, 28 (4), 467–490. https://doi.org/10.1016/0306-4573(92)90005-K

Wilson, P. (1983): *Second-hand knowledge: An inquiry into cognitive authority*. Westport, Conn.; London: Greenwood Press.