MEDINFO 2001
V. Patel et al. (Eds)
Amsterdam: IOS Press

A New Approach to the Concept of "Relevance" in Information Retrieval (IR)

Yuri Kagolovsky, Jochen R. Möhr

School of Health Information Science
University of Victoria, Victoria, B.C., Canada

Abstract

The concept of "relevance" is the fundamental concept of information science in general and information retrieval, in particular. Although "relevance" is extensively used in evaluation of information retrieval, there are considerable problems associated with reaching an agreement on its definition, meaning, evaluation, and application in information retrieval. There are a number of different views on "relevance" and its use for evaluation. Based on a review of the literature the main problems associated with the concept of "relevance" in information retrieval are identified. The authors argue that the proposal for the solution of the problems can be based on the conceptual IR framework built using a systems analytic approach to IR. Using this framework different kinds of "relevance" relationships in the IR process are identified, and a methodology for evaluation of "relevance" based on methods of semantics capturing and comparison is proposed.

Keywords:

Information Retrieval; Relevance; Evaluation Studies

Introduction

The concept of "relevance" plays an important role in information science in general and information retrieval (IR) in particular. Saracevic [1] argues that the concept of "relevance" is the main reason for the emergence of information science as an independent area of research. Park [2] concurs with this view and describes "relevance" as "a fundamental concept of information science" and "a key problem in IR research."

At the same time, despite the numerous publications that have discussed and used the concept of "relevance" since the beginning of information science as a distinct discipline, there exists "little agreement as to the exact nature of relevance and even less that it could be operationalized in systems or for the evaluation of systems" [3, p. 124]. Schamber, Eisenberg, and Nilan argue that "an enormous body of information science literature is based on work that uses relevance, without thoroughly understanding what it means" [4, p. 756]. Park supports their view: "The idea of relevance has played a major role in the evaluation of information retrieval, but without a consensus on the meaning of this concept" [2, p. 135]. Schamber [5] thinks that the majority of problems are related to the use of different views of relevance, as well as "loose and inconsistent" terminology by different researchers. Froehlich [3] has identified the following main factors contributing to "relevance"-related problems in information science: the inability to define relevance; the inadequacy of topicality as the basis of relevance judgements; the diversity of non-topical, user-centered criteria that affect relevance judgements; the dynamic and fluid character of information seeking behavior; the need for appropriate methodologies; and the need for more complex, robust models for system design and evaluation.

Review of the literature

The classical research paradigm for IR evaluation is the Cranfield paradigm [6]. In this paradigm, experiments are conducted in the artificial conditions of experimental settings, and real users do not participate in the experiments. In the Cranfield paradigm, relevance means "on the topic," "on the subject," "aboutness" [2]. Some researchers [4] refer to this view as "topical relevance," meaning that to be judged "relevant," the topic of the document has to match the topic of the query. This approach implies a fixed and unchanging relationship between a query and a document. This view of relevance has been intensively criticized since the beginning of IR experiments. "Topical relevance" does not take into account circumstances of a search, including the individual's information need. It does not consider a user's individual characteristics, as well as the dynamic nature of a user's state of knowledge and information needs during a search. Many researchers have pointed to the necessity of studying the nature of relevance from the perspective of real users.

Different aspects of user-centered relevance have been investigated. A concept of "utility" was introduced quite early in IR research [7]. This concept includes quality, novelty, importance, and credibility of information from the

point of view of a user, as opposed to just topical relatedness. Wilson [8] discusses a "situational relevance" that implies multiple concepts and involves the relation between information and a particular individual's view and situation. Swanson [9] defines "subjective relevance" as the mental experience of an individual person who has an information need. As another important result of the discussion of user-related relevance, the difference between "relevance" and "pertinence" has been pointed out [10]. "Pertinence" requires that, to be judged "relevant," a particular document has a relationship to a user's specific information problem, and that it changes a user's state of knowledge about this problem [11].

Some attempts to summarize different views of "relevance" and clarify the usage of relevance-related terminology have been made. Schamber [5] presents three views of relevance: a system view, an information view, and a situation view. The system view is objective and refers to a match between a query and a document. The information view is subjective and refers to the relation between a request for information and a document. The situation view is subjective and refers to the relationship between a user's information need and a document.

Schamber [5] shows that there is an overlap in the terminology used in each of these views. This overlap is a result of "conceptual gray areas" in our understanding of the concept of "relevance." One of these areas is the role of topicality in relevance judgements. Even researchers that argue for the subjective character of relevance agree that topicality is a necessary, but not sufficient, condition in the relevance judgements of real users [2]. Therefore, there is a connection between objective and subjective views of relevance. While Froehlich argues that "the prototypical core for relevance judgements or the nuclear sense of relevance is topicality" [3, p. 129], Schamber [4] proposes that the use of topicality in evaluation is a logical step for information systems development.

Some authors believe that evaluation has to involve real users making relevance judgements [2, 12]. Others argue that any person who is knowledgeable about the subject of a search can make relevance judgements [6, 13]. This argument addresses the issue of "relevance judgement." The authors supporting the involvement of real users in evaluation consider a relevance judgement to be a representation of the value of the document for a particular user in a particular situation (time, place, context, etc). The contra-argument is based on the view of a relevance judgement as a decision about whether or not a retrieved document answers a query.

A variety of methods for capturing relevance decisions have been proposed [5, 13, 14]. Some of them use binary ("yes" or "no") or category scales. However, the use of these scales for relevance decisions often implies a direct topical relation between a query and a document based on a fixed and unchanging relationship. The majority of researchers agree that the traditional experimental approach (objectivist approach) is not an adequate methodology for understanding the nature of psychological relevance. The large number of variables that can effect relevance judgements [5] emphasize the necessity of a new methodology for evaluating relevance. As an alternative to the traditional experimental approach, different methods of qualitative research are used for studying user-based relevance. The main advantage of the qualitative research approach is that it takes into account the dynamics of an individual's cognitive state and situational factors. While qualitative research is complex, it can, nevertheless, produce data that is systematic and measurable. All these factors explain the growing popularity of the qualitative research in evaluating relevance in IR [2-4, 15].

A Proposal for Solution

Conceptual IR Framework

We have proposed and applied a theoretical framework based on the systems analytic approach to information retrieval [16, 17]. This framework has permitted to identify components of the IR process, their structural and functional characteristics, and interactions with other components of this process. We proposed a new definition of IR, distinguishing between an IR process and a technical IR system (search engine). Using the proposed framework, we provided a critical analysis of the relevance-based measures of recall and precision, and outlined some alternatives to evaluation of search engines based on users' information needs satisfaction.

We define **Information Retrieval (IR)** as a process of human interaction with a technical IR system (search engine) in a specific setting with the goal to find information sources relevant to a specific information need.

A generalized model of IR process may be stated as follows (Figure 1). Humans initiate the IR process. A user operates in an environment, which is characterized by such variables as setting (laboratory or operational), organizational policy, or type of research. To be able to solve a problem in an environment, the user accesses and evaluates her state of knowledge (SK) about the task. If a user's SK is not sufficient for decision-making, the user has an anomalous state of knowledge (ASK) [18]. This state of knowledge can also be called an implicit information need. Based on the ASK, the user can make a statement about her information need explicitly using a natural language.

The next step is to make a decision about a strategy to resolve the ASK. This decision is based on a user's previous knowledge about the subject, formulation of information need(s), availability of information sources, a user's familiarity with these sources, and other factors. Some of the possibilities include looking for a person who has the needed information, or finding this information in a printed form. A search for printed resources can lead to library and/or electronic data sources. The human interacts with the technical IR system through the interface. The interface transforms user's queries into system specific

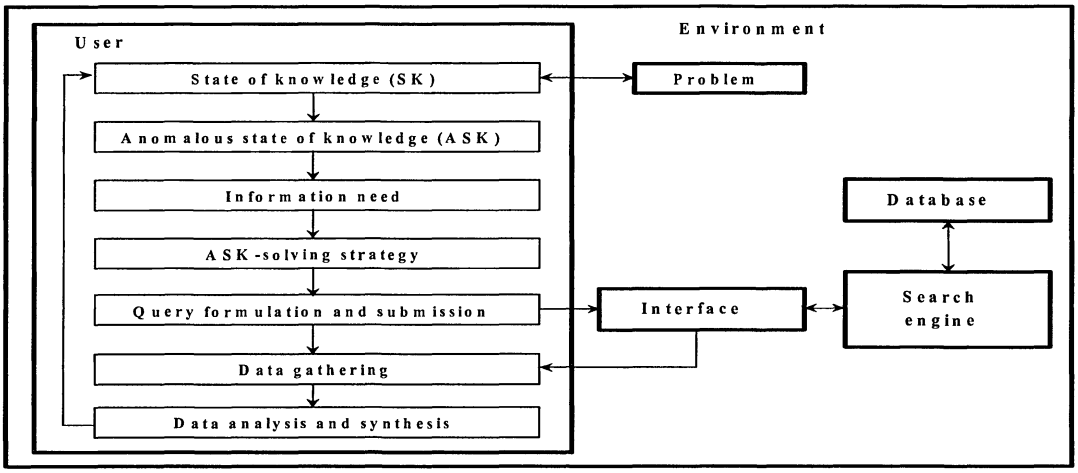

Figure 1 - Process model of IR

commands and presents the retrieved set of documents. This bi-directional function explains the relation between the processes of the user's query formulation and the querying mechanisms of the search engine. These mechanisms work in connection with other functions of the technical IR system. Users have to be able to transform their information needs properly into a search statement (the verbalized statement of information needs) and build a query through the use of the interface. The query consists of the commands stated in syntax permitted by the querying component of the technical IR system. It has a structure of search elements (terms, codes, etc.) that depends on the model of data representation in the document set built by the technical IR system.

The retrieved set of documents is presented to the human component of the IR process through the interface, which serves as transport medium for presentation on a screen or as a printed copy. This set of documents can be read and evaluated by a user who extracts information from the documents and attempts to incorporate it into her cognitive model. Based on this cognitive model, the user can either decide to stop her search or to adjust her information needs. In this last case the cycle outlined above can be continued.

Discussion

Identifying "relevance" relationships

Using the process model of IR (Fig. 1) the following "relevance" relationships can be identified and considered for evaluation:

1. How is a user's state of knowledge relevant to a problem?
2. How is a formulated information need relevant to a user's state of knowledge?
3. How is a user's information seeking strategy (ASK-solving strategy) relevant to
   a. a user's explicit information need?
   b. a user's implicit information need (ASK)?
   c. a problem?
4. How is a formulated query relevant to a user's
   a. state of knowledge (implicit information need)?
   b. explicit information need?
5. How are retrieved documents relevant to
   a. a submitted query?
   b. a user's explicit information need?
   c. a user's ASK?
   d. a user's problem solving?

This list of "relevance" relationships includes the most typically used ones. For example, it includes the most often considered views of "relevance" as the "subject relatedness" of a record to a query and the "value" or "utility" of a record to a user. In addition, it includes those relationships identified by other researchers [19], such as how retrieved documents are relevant to a problem or a user's ASK. To these this list further adds new types of "relevance" relationships and introduces new perspectives on the concept of "relevance" in IR. For example, the following "relevance" relationships have not been identified by IR researchers: how a formulated information need is relevant to a user's state of knowledge; how a user's information seeking strategy is relevant to a user's explicit information need, user's implicit information need (ASK), and a problem. Although further "relevance" relations might exist, we consider the proposed list as likely complete for now.

Although, some authors describe the "subject relatedness" as corresponding to type 5a in our scheme of "relevance" relationships [13], others identify it as corresponding to type 5b [20]. The same is true for the interpretation of the "value" and "utility" aspects of "relevance" that might correspond to types 5c or 5d, depending on the preferences of researchers. In contrast, our proposed classification of "relevance" relationships permits us to identify more precisely the types of relationships under consideration. Therefore, our structured model of IR built using a systems analytic approach makes the process of "relevance" analysis and evaluation more comprehensive and precise. Moreover, it will be shown that the same model provides a basis for choosing appropriate approaches to evaluating these "relevance" relationships in IR.

A proposal for evaluating "relevance" relationships

We propose the following steps to evaluate "relevance": 1) choose a "relevance" relationship for evaluation; 2) characterize this relationship; 3) identify an experimental design; for example: identify subjects, how the experiment will be conducted, and other issues; 4) capture the semantics of each part of the "relevance" relationship; for example: a query, a document, a user's state of knowledge, a problem description, a user's action, and others; 5) compare the semantics of both parts involved in the "relevance" relationship; 6) make suggestions.

One example could be an evaluation of a "relevance" relationship between a problem description and a statement of users' information needs. This relationship reflects how well users are able to express their understanding of the problem by means of a natural language[1]. Evaluating this type of relationship is very important, as it is well known that people usually cannot sufficiently express their needs for data and information by means of a natural language [5]. Thus, librarians often complain that users usually have difficulties formulating what they are looking for during a pre-search interview.

The following is a possible experimental design. Users are presented with a description of a problem[2]. They are asked to "think aloud" as they analyze the problem and identify their knowledge about this problem, trying to formulate a statement of their information needs. A problem description and users' information needs statements are analyzed using, for example, propositional analysis [21]. Their captured meanings are compared and a conclusion about this "relevance" relationship is made.

[1] As we are interested in a user's ability to translate mental constructs into a natural language, we do not have to investigate a user's original level of knowledge about the problem.

[2] Although a problem can be also presented through a story-telling or a movie, a textual form is easier to use for our purposes. However, in the future it could be interesting to investigate different ways of presenting a problem.

The proposed experimental design would also permit analysis of the dynamics of working with cognitive structures that characterize users' states of knowledge about a problem. This can be done using "think aloud" protocol analysis [22]. The results of this analysis can help to identify problems and cognitive blocks, and to plan steps for improving a process of formulating users' information needs statements, for example, users' education.

Possible approaches to the capturing and comparison of semantics

To choose appropriate methods to be used in semantics capturing and comparison for evaluating "relevance" relationships in IR is not a simple task. We are currently comparing existing methods and testing them on textual data. Based on our experience and a literature review the following problems can be identified: the methods have to be domain independent, handle different types of data, and have to be robust and reproducible. For example, to be useful for IR evaluation a required method of semantics capturing has to be able to work with a text (queries and documents), transcripts of "think aloud" protocol representing users' cognitive processes, and users' actions.

Methods of semantics capturing have a goal to represent a meaning formally. Two main areas working on semantics capturing are artificial intelligence (AI) and cognitive psychology. The most often used methods in AI are frames, semantic networks, and conceptual graphs [23]. In cognitive psychology, researchers use methods of propositional analysis [21], prose analysis [24], and Frederiksen's grammar [25]. Kintsch [26] has offered a detailed comparison of different methods of semantics capturing (features systems, associative networks, semantic networks, schemes, frames, scripts, and propositional analysis). This comparison shows that propositional analysis is a robust, flexible method that is well suited for analysis of cognitive processes. Based on this comparison, and on the fact that in IR we deal with analysis of cognitive processes either directly (a user) or indirectly (text), we have chosen Kintsch's propositional analysis as a method for future research.

Semantics comparison is even a more complicated issue that has to be researched further. However, as results of propositional analysis can be represented as a network of propositions, it is possible to explore graph-theoretic methods [27], overlay analysis [28], and other strategies.

Conclusion

The proposed conceptual IR framework built using the systems analytic approach to the IR process provides a basis for a methodology to address long-standing problems and controversies associated with the concept of "relevance." This methodology can help in identifying "relevance" relationships in the IR process and choosing methods for their evaluation. Methods of semantics capturing and comparison can be used to create better approaches to

evaluating "relevance" relationships. Additional research is required to create, test, and implement these new approaches.

Acknowledgments

This research was supported in part by a grant from HEALNet, a Canadian Network of Centres of Excellence in Health Research, supported by the Medical Research Council and the Social Sciences and Humanities Research Council of Canada, and the Province of British Columbia, and by a grant from the Natural Science and Engineering Research Council (NSERC) of Canada.

References

[1] Saracevic T. Relevance: A review of and a framework for the thinking on the notion in information science. *JASIS* 1975;26:321-343.

[2] Park TK. Toward a Theory of User-Based Relevance: A Call for a New Paradigm of Inquiry. *JASIS* 1994;45(3):135-141.

[3] Froehlich TJ. Relevance Reconsidered - Towards an Agenda for the 21st Century: Introduction to Special Topic Issue on Relevance Research. *JASIS* 1994;45(3):124-134.

[4] .Schamber L, Eisenberg MB, Nilan MS. A re-examination of relevance: Toward a dynamic, situational definition. *IPM* 1990;26:755-776.

[5] Schamber L. Relevance and information behavior. In: *ARIST*; 1994. p. 3-48.

[6] Cleverdon CW. The significance of the Cranfield tests on index languages. In: *ACM SIGIR'91*; 1991; 1991. p. 3-12.

[7] Cooper WS. On Selecting a Measure of Retrieval Effectiveness. *JASIS* 1973;24:87-100.

[8] Wilson P. Situational relevance. *ISR* 1973;9:457-471.

[9] Swanson DR. Subjective versus objective relevance in bibliographic retrieval systems. *LQ* 1986;56:389-398.

[10] Foskett DJ. A Note on the Concept of "Relevance". *ISR* 1972;8(2):77-78.

[11] Howard DL. Pertinence as Reflected in Personal Constructs. *JASIS* 1994;45(3):172-185.

[12] Harter SP. Variations in Relevance Asessments and the Measurement of Retrieval Effectiveness. *JASIS* 1996;47(1):37-49.

[13] Salton G, McGill MJ. *Introduction to Modern Information Retrieval*. New York: McGraw-Hill; 1983.

[14] Harter SP, Hert CA. Evaluation of Information Retrieval Systems: Approaches, Issues, and Methods. In: Williams ME, editor. *ARIST*. Medford, NJ: Information Today, Inc.; 1997. p. 3-94.

[15] Fidel R. Qualitative methods in information retrieval research. *LISR* 1993;15:219-247.

[16] Kagolovsky Y, Moehr JR. A structured model for evaluation of information retrieval. In: Cesnik B, McCray AT, Scherrer J-R, editors. *MedInfo '98: 9th World Congress on Medical Informatics*; Seoul, Korea: IOS Press; 1998. p. 171-175.

[17] Kagolovsky Y, Moehr JR. Evaluation of information retrieval: old problems and new perspectives. In: *8th International Congress on Medical Librarianship*; 2000 Online: http://www.icml.org/tuesday/ir/kagalovosy.htm

[18] Belkin NJ. Cognitive models and information retrieval. *Soc Sci Inf Stud* 1984;4:111-129.

[19] Mizzaro S. Relevance: The Whole History. *JASIS* 1997;48(9):810-832.

[20] Meadow CT. *Text Information Retrieval Systems*. 1st ed. San Diego: Academic Press, Inc.; 1992.

[21] Kintsch W. Text processing: A psychological model. In: *Handbook of Discourse Analysis: Dimensions of Discourse*. London: Academic Press; 1985. p. 231-243.

[22] Kushniruk AW, Patel VL. Cognitive evaluation of decision making processes and assessment of information technology in medicine. *Int. J. Med. Inform.* 1998;51:83-90.

[23] Brachman RJ, Levesque HJ, editors. *Readings in knowledge representation*. Los Altos, CA: Morgan Kaufmann Publishers, Inc.; 1985.

[24] Meyer BJF. Prose Analysis: Purposes, Procedures, and Problems. In: Britton BK, Black JB, editors. *Understanding expository text*. Hillsdale, NJ: Erlbaum; 1985. p. 11-64.

[25] Frederiksen CH. Representing logical and semantic structure on knowledge acquired from discourse. *Cog Psyc* 1975;7:371-458.

[26] Kintsch W. The representation of knowledge and the use of knowledge in discourse comprehension. In: Dietrich R, Graumann CF, editors. *Language Processing in Social Context*: Elsevier Science Publishers B.V. (North-Holland); 1989. p. 185-209.

[27] Sowa JF. *Conceptual Structures: Information Processing in Mind and Machines*. Reading, MA: Addison-Wesley; 1984.

[28] Carr B, Goldstein IP. *Overlays: a theory of modeling for computer-assisted instruction*. Cambridge, MA: Massachusetts Institute of Technology; 1977.

Address for correspondence

Yuri Kagolovsky, MD, MSc Candidate
School of Health Information Science
University of Victoria, PO Box 3050
Victoria, BC, CANADA V8W 3P5
E-mail: ykagolov@uvic.ca