The Concept of Relevance in IR

Pia Borlund
Department of Information Studies, Royal School of Library and Information Science, Aalborg Branch, Sohngaardsholmsvej 2, DK-9000 Aalborg, Denmark. E-mail: pb@db.dk

This article introduces the concept of relevance as viewed and applied in the context of IR evaluation, by presenting an overview of the multidimensional *and* dynamic nature of the concept. The literature on relevance reveals how the relevance concept, especially in regard to the multidimensionality of relevance, is many faceted, and does not just refer to the various relevance criteria users may apply in the process of judging relevance of retrieved information objects. From our point of view, the multidimensionality of relevance explains why some will argue that no consensus has been reached on the relevance concept. Thus, the objective of this article is to present an overview of the many different views and ways by which the concept of relevance is used—leading to a consistent and compatible understanding of the concept. In addition, special attention is paid to the type of situational relevance. Many researchers perceive situational relevance as the most realistic type of user relevance, and therefore situational relevance is discussed with reference to its potential dynamic nature, and as a requirement for interactive information retrieval (IIR) evaluation.

Introduction

The main objective of information retrieval (IR) is the retrieval of *relevant* information, or as van Rijsbergen (1979, p. 6) formulates it ". . . to retrieve all the *relevant* documents [and] at the same time retrieving as few of the *non-relevant* as possible." Thus, the concept of relevance is acknowledged as a fundamental issue and of a central concern to the functioning and evaluation of IR systems. Since the ASTIA (Gull, 1956) and Cranfield Uniterm tests (Cleverdon, 1960; Thorne, 1955) the debate concerning the concept of relevance has formed an important part of the discussions in the field of information science. The importance of the relevance concept is, for instance, documented with the comprehensive relevance studies by Cuadra and Katter (1967a; 1967b) and by Rees and Schultz (1967), which contribute to an improved understanding of relevance. The analytical review of relevance by Saracevic (1975) is a key paper of how relevance is understood and employed in the past. During the 1990s, the relevance discussion has intensified (e.g., see Froehlich, 1994; Green, 1995; Harter, 1992; Janes, 1994; Mizzaro, 1998; Park, 1994; Saracevic, 1996; Schamber, Eisenberg, & Nilan, 1990). One may say it is Eisenberg and Schamber (1988), and Schamber et al. (1990) who reopen the discussion by reintroducing the concept of situational relevance. Schamber et al. (1990, p. 774) draw three central conclusions from the nature of relevance and its role in information behavior after having reviewed the literature and presented various views of relevance:

- relevance is a multidimensional cognitive concept whose meaning is largely dependent on users' perceptions of information and their own information need situations;
- relevance is a dynamic concept that depends on users' judgements of quality of the relationship between information and information need at a certain point in time;
- relevance is a complex but systematic and measurable concept if approached conceptually and operationally from the user's perspective.

In summary, the three conclusions define what Schamber et al. (1990), as well as we in the present article, refer to as situational relevance. Harter (1992) suggests a related concept: psychological relevance, whereas Janes (1994) empirically tests and compares different types of relevance assessments. Froehlich (1994), Park (1994), Green (1995), and Mizzaro (1998) describe, review, and analyze the relevance concept in terms of a variety of aspects and views as well as the research on the issue. Saracevic (1996) proposes an enhanced model for the understanding of relevance, and applies the relevance concept in relation to IR and information seeking.

It is often said that no consensus exists on the relevance concept (e.g., Mizzaro, 1998, p. 305). It is, however, our opinion that the recent years of renewed debate of the concept has resulted in an improved understanding of the concept. Thus, the aim of this article is to introduce the

Received July 12, 2002; revised October 15, 2002; accepted February 3, 2003

concept of relevance as viewed and applied in the context of IR evaluation, and hereby demonstrate that a consistent and compatible understanding of the relevance concept has been reached. This is done via reviews of the relevance literature from a multidimensional and dynamic perspective, respectively—inspired by the first and second conclusions of Schamber et al. (1990). The demonstrated level of a consistent and compatible understanding of relevance outlines at the same time a relevance framework that enables the concept of relevance to be approached conceptually and operationally from the user's perspective—the third conclusion of Schamber et al. (1990). As a consequence of the three conclusions special attention is paid to the type of situational relevance. Situational relevance is perceived by many researchers (e.g., Harter, 1992; Saracevic, 1996; Schamber et al., 1990) as the most realistic type of user relevance, and therefore, is recommended for interactive information retrieval (IIR) evaluation (e.g., Borlund, 2000a, 2000b; Borlund & Ingwersen, 1997).

Reviews of the relevance literature reveal that the relevance concept can be divided into various classes and types of relevance, that the concept is applied with reference to relevance criteria, degrees of relevance, and at different levels. The classes, types, criteria, degrees, and levels of relevance refer to the multidimensionality of relevance. However, as pointed out by Schamber et al. (1990), relevance is also dynamic. Basically, this division corresponds to the subsections of the next section. Then, situational relevance is defined and discussed in relation to its characteristics, including the relationship between situational relevance and information need development, and evaluation of IR systems. Then we close the article with an overall discussion and summary.

Multidimensional *and* Dynamic Relevance

At a general level multidimensional relevance refers to how relevance can be perceived and assessed differently by different users. Dynamic relevance refers to how this perception can change over time for the same user. The relevance concept may be perceived differently by different users, but is also used and defined differently by different scientists.

The literature on relevance reveals how the relevance concept, especially in regard to the multidimensionality of relevance, is many faceted, and does not just refer to the various relevance criteria users may apply in the process of judging relevance of retrieved information objects. From our point of view, the multidimensionality of relevance explains why some will argue that no consensus has been reached on the relevance concept.

Classes and Types of Relevance

Basically, the concept of relevance can be divided into two main classes of relevance as, for instance, done by Saracevic (1975), Swanson (1986), and Harter (1992). The two classes are: (1) objective *or* system-based relevance; and (2) subjective *or* human (user)-based relevance.

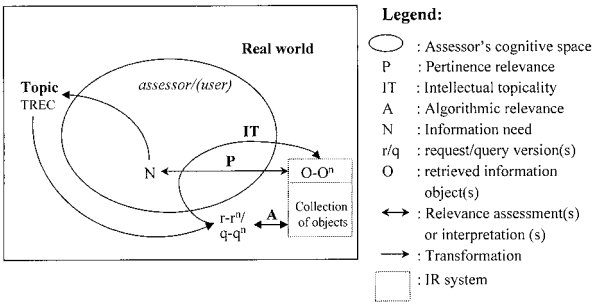

FIG. 1. Illustration of the traditional view of relevance types involved in a noninteractive IR situation during an IR session (Borlund, 2000a, p. 29).

The two main classes of relevance are quite different in nature and by default imply different degrees of intellectual involvement. Each of the two main classes of relevance corresponds to the understanding of relevance employed by each of the two main approaches to IR research and evaluation: the system-driven and the cognitive user-oriented approaches. The system-driven approach treats relevance as a static and objective concept as opposed to the cognitive user-oriented approach that considers relevance to be a subjective individualized mental experience that involves cognitive restructuring (Swanson, 1986, pp. 390–391).

Based on the two main *classes* different *types* of relevance are identified. These types refer to the different kinds of relations, which can be expressed between (retrieved) information objects and query, request, information need, or the underlying situation that triggers the need for information (Saracevic, 1996). Types of relevance are sometimes also defined or characteriszd by relevance criteria (e.g., Cosijn & Ingwersen, 2000; Saracevic, 1996), but to our knowledge, no empirical evidence supports this approach, although Vakkari (2000) does presents some indications.

Saracevic (1996, p. 214) distinguishes between five basic types of relevance, or what he refers to as manifestations of relevance. These are: (1) *System* or *algorithmic relevance*, which describes the relation between the query (terms) and the collection of information objects expressed by the retrieved information object(s); (2) a *topical-like* type, associated with aboutness; (3) *pertinence* or *cognitive relevance*, related to the information need as perceived by the user; (4) *situational relevance*, depending on the task interpretation; and (5) *motivational* and *affective,* which is goal-oriented. The first type of relevance outlined by Saracevic (1996) exists within the *class* of objective relevance. Algorithmic relevance is also commonly known as "logical" relevance (Cooper, 1971) or as "topicality" relevance, and can be defined as ". . . how well the topic of the information retrieved matches the topic of the request. A document is objectively relevant to a request if it deals with the topic of the request" (Harter, 1992, p. 602). Algorithmic relevance (see Figs. 1 or 2) seems the most common and clearest

definition of relevance, and is the relevance type applied in traditional evaluation of IR systems. But the prerequisite for the definition is to understand the concept of *topic*. The objective-oriented type of algorithmic relevance is restricted to deal only with the degree to which the query representation matches the contents of the retrieved information objects, i.e., topic equals *contents*. This algorithmic and feature-based type of relevance is context free and is, for instance, applied when the relevance of an information object is computed as a function of the number of features in common between the query representation and the information objects, usually resulting in values between 0 and 1. The values are then used to rank the retrieved information objects by listing the most topically relevant information object first. A specification of objective types of algorithmic relevance could be labeled, for example, vector space relevance, probabilistic relevance, or Boolean relevance.

The following four relevance types specified by Saracevic (1996) belong to the class of subjective or user-based relevance. The types of relevance within this class are concerned with the aboutness and appropriateness of a retrieved information object and refers to the various *degrees of intellectual interpretations* carried out by human observers—whether assessors or users. The subjective class of relevance may, as a generic concept, refer to the aboutness, usefulness, usability, or utility of information objects in relation to the fulfilment of goals, interests, work tasks, or problematic situations intrinsic to the user. It is context-dependent.

In Saracevic's subjective topical-like relevance category the concept of topic is understood as *aboutness*, not contents, i.e., an intellectual assessment of how an information object corresponds to the topical area required and described by the request(s) for information. This relevance type is consequently not entirely based on the relationship between a query representation and a retrieved information object. An observer, either an assessor or a user, makes the judgement. It is this kind of subjective relevance assessment we ascribe to individual assessors who participate in common IR experiments like TREC,[1] although their judgements traditionally are intended to be of an objective nature. To distinguish the subjective type of topical oriented relevance form algorithmic relevance we name this type of relevance: *intellectual topicality* (see Fig. 1 or 2). Confusion exits in the literature when the topicality label of relevance is used to refer to either algorithmic relevance or intellectual topicality, and sometimes to even both of them at the same time, as in the case of Schamber et al. (1990).

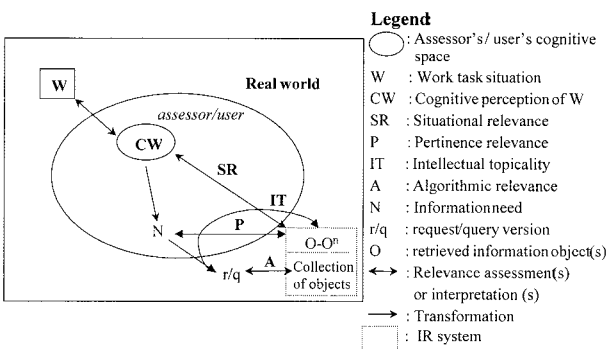

FIG. 2. Illustration of main types of relevance relationships, including situational relevance, involved in an IR session for a given instance (Borlund, 2000a, p. 31).

Pertinence represents the intellectual relation between the intrinsic human information need and the information objects as currently interpreted or perceived by the cognitive state of an assessor or user. This allows for the existence of a *dynamic* information need. Like for human test persons or real users we are aware that pertinence may indeed be involved in assessors' relevance judgements and, depending on the experimental scenario, strongly related to the more narrow intellectual topicality type (see Figs. 1 or 2).

Situational relevance is understood as the utility or *usefulness* of the viewed and assessed information object(s) by pointing to the relationship between such retrieved object(s) and the work task at hand underlying the information need as perceived by the user. Situational relevance is a highly context dependent as well as a potentially dynamic type of relevance. However, in practice, it may not be possible to distinguish between the types of pertinence and situational relevance.

The fifth type of relevance listed by Saracevic (1996), motivational or affective relevance, is in our opinion not an independent type of relevance. To us, motivational/affective relevance is a characteristic of all of the subjective types of relevance. According to Saracevic (1996), motivational/affective relevance describes the relation between the intents, goals, and motivations of the user and the information object(s). In short, the intent, goals, and motivation is what make users search for information, carry out information retrieval, and assess relevance of the retrieved information objects. Thus, the "drive" to want information is not an independent, specific type of relevance, but an *inherent characteristic* of relevance behavior in general. Cosijn and Ingwersen (2000) have put forward a similar argumentation. Hence, we consider the three mentioned subjective types of relevance as overall main types of subjective relevance.

Figure 1 illustrates different main types of relevance involved in traditional IR experiments, like Cranfield or TREC, during a session. The objective type of relevance is signified by the topical nearness of algorithmic relevance **(A)** between the collection of information objects and the expressed query **(q)** or request **(r)** resulting in the set of retrieved objects **($O–O^n$)**. In this scenario, because **(N)** is assumed invariable or equal to a TREC topic **($q–q^n$)**; thus,

[1] TREC is the acronym for Text REtrieval Conferences. For an overview of TREC the reader is directed to the TREC homepage at: http://trec.nist.gov/pubs.html, where overview papers can be found.

only signify traditional query transformations. If the concept of "topic" signifies "document contents," algorithmic relevance **(A)** can in reality be assessed in an automatic way by means of various similarity measures by machines. Each of the measures will, however, produce slightly different assessments for the same objects (e.g., Borlund and Ingwersen, 1997). But if the concept means the subject matter or aboutness of the document contents, topicality assessments require *intellectual* interpretations **(IT)**, i.e., they are carried out by human observers/assessors. In that kind of assessment the variation of assessments between different assessors is held to be of a greater magnitude than in the algorithmic type. In other words, the decade-old discrepancy between natural language representation and human subject indexing can be paralleled in the relevance discussion (e.g., Hjørland, 1997). Intellectual topicality **(IT)** expresses the intellectual judgement of the *aboutness correspondence* between each retrieved object **($O–O^n$)** and information request **(r)** put forward to the IR system. Pertinence **(P)** describes the relevance relation between retrieved information objects **($O–O^n$)** and the static information need **(N)** as seen by the observer/assessor otherwise normally the owner of the information need, the user.

Figure 1 thus illustrates the Cranfield/TREC relevance assessment problem. The assessor, who in TREC generates the TREC Topics, must have an idea or reason for the Topic. Hence, when assessing the object(s) retrieved by the system [algorithmic assessments **(A)**], he or she is assumed to do this quite *objectively* by means of **(IT)**. In principle, this objectivity is not possible due to the subjectivity inherent in **(IT)**—a subjectivity that is reinforced by a possible (but real) influence of **(P)** assessments. The degree of **(P)** influence is impossible to measure, but the results by Voorhees (1998, 2000) indicate that indeed there exists an amount of significant subjectivity causing variation of assessments among three assessors that only become statistically insignificant when a huge number of topics are involved.

Figure 2 shows how situational relevance **(SR)** exists as a relationship between the retrieved objects **($O–O^n$)** and the user's perception of a given work task situation **(CW),** in real-life environments over *one instance*. The stated request **(r)** functions as a transformation and indication of the current information need version **(N)** internal to the user. Although situational relevance **(SR)** embodies the subjective relevance types **(IT)** and **(P)**, **(SR)** does not necessarily relate directly to **(A)**. Some objects may be situationally relevant, but not algorithmically relevant at the same time, and vice versa.

The users' relevance assessments are consequently not based on the relationship **(A)** between the query **(q)** and the collection of information objects, but the outcome of **(A)**—the retrieved information objects **($O–O^n$)**. The subjective relevance assessments are made based on the relationship between the request, the information need **(N)**, or work task situation **(W)**, as interpreted by the user **(CW)**, and the retrieved objects **($O–O^n$)**—according to the employed types of subjective relevance **(IT)**, **(P)**, or **(SR)**. In the case of **(SR)** the request is simply the "tool," that the user applies to describe and/or indicate the current information need, based on the perceived work task **(CW)**. The user employs the intellectual interpretation of a given work task situation as a platform **(CW)** for the relevance judgement. Aside from the search engines or interfaces under evaluation, the only unknown variables in the model (Fig. 2), is the **(N)** and **(CW)**. These variables can be standardized by having the same set of *simulated work task situations* searched by several test persons (Borlund, 2000a, 2000b; Borlund & Ingwersen, 1997). All other parameters **(SR, P, IT, A, q/r, O,** and **W)** are or become known during experimentation. The *role* of the assessor is different in the settings depicted in Figures 1 and 2. In the latter, the assessor may not generate the **(W)**, but acts in line with other users getting a **(W)** and performing IIR. Assessors may thus act as the control group. The same condition must apply to the users and assessors due to the subjective nature of the assessments. Otherwise, the assessors become an uncontrollable variable. For a realistic and dynamic relevance setting that takes time into account, we refer to Figure 3 presented later on in the article.

Additional subjective types of relevance are proposed in the relevance literature, for example, "psychological relevance" (Harter, 1992), "ostensive relevance" (Campbell & van Rijsbergen, 1996), and "task relevance" (Mizzaro, 1998; Reid, 1999). We consider the proposals as examples belonging to any of the three listed types, if not yet another characteristic of the subjective types of relevance. For instance, the type of psychological relevance by Harter (1992), adopted from Sperber and Wilson (1986), is in line with the basic and fundamental ideas of the cognitive viewpoint: the change of knowledge structures of the recipient by the act of information processing. Harter (1992, p. 612) describes psychological relevance as a state of effect that exists when the user retrieves information, which suggests new cognitive connections, fruitful analogies, insightful metaphors, or an increase or decrease in the strength of a belief. Thus, psychological relevance is to be seen as the effect of a change of knowledge structures. As such, psychological relevance basically describes the dynamic nature of situational relevance and the corresponding dependency of the cognitive perception and interpretation of the information need and underlying situation. Ostensive relevance, proposed by Campbell and van Rijsbergen (1996, p. 264), is defined as the degree to which evidence from the retrieved information object is representative and indicative of the current information need. The definition of ostensive relevance is no different than the definition of pertinence. However, although Campbell and van Rijsbergen (1996) define the ostensive relevance similar to pertinence, they do work with the concept of a dynamic information need reflected in the probability weighting in the "ostensive model" of IR. This implies a stronger intellectual as well as a dynamic type of relevance than pertinence—perhaps in reality a mixing of pertinence and situational relevance. Nevertheless, the discussion of relevance types may be of no impor-

tance to Campbell and van Rijsbergen (1996), as they base their work on similar ideas as those put forward by Doyle (1963). Doyle (1963, p. 199) argues that the ideal IR system should not be designed around the concept of relevance but should be based on the concept of *exploratory capability*. The reason for this is that the users have problems stating their true information needs in the simple form required by conventional IR systems, therefore ". . . the searcher needs an efficient exploratory system rather than a request implementing system" (Doyle, 1963, p. 199). As a consequence of this, Campbell has developed an IR system that based on *ostensive evidence* indicated by the user then suggests tracks or paths of information objects for the user to trace down and explore. Mizzaro (1998) proposes a four-dimensional relevance model in which relevance is defined as the relationship between: (1) an information resource (surrogate, document, information), and (2) a representation of a user's problem (query, request, real information need, perceived information need) judged according to (3) one or more of the following components: topic, task, and/or context at (4) a given time point. To Mizzaro (1998, p. 311) the ultimate type of relevance is: *rel*[*Information, RIN, t(f),* {*Topic, Task, Context*}]. Which refers to: ". . . the relevance of the information received to the RIN [real information need] at time *t(f)* for all the three components (the relevance the user is interested in)" (Mizzaro, 1998, p. 311). It is, however, not Mizzaro who name this type of relevance, task relevance, but Reid (1999, pp. 105–106). Reid builds upon the work by Mizzaro (1998). According to Reid (1999, p. 105), task relevance acknowledges: (a) the situational nature, as affected by the context, as well as by the task; and (b) the need to measure relevance at the time-point where the real information need is satisfied.

Both Mizzaro and Reid are interested in capturing the user's perceived *usefulness* of the retrieved information objects with reference to the user's (work) task and real information need, this makes task relevance equal to situational relevance. To us, it makes no operational difference if the relevance is judged during the course of the retrieval session or later, for example, as proposed in the framework by Reid (1999), by the time the user has completed the (work) task. The relevance definitions, that of (work) task relevance and situational relevance are identical.

Concurrently, Harter (1992), Campbell and van Rijsbergen (1996), Mizzaro (1998), and Reid (1999) acknowledge as well as require relevance to be of a subjective and dynamic nature, which supports the employment of pertinence and situational relevance.

Relevance Criteria

The concept of relevance is, as mentioned earlier, the primary criterion for the evaluation of IR systems. The process of assessing or judging relevance is, no matter the applied classes and types of relevance, based on various criteria. With reference to the objective class of relevance and the types of relevance within this class, the relevance criterion applied is that of the match between the query representation and the content terms of the retrieved information objects. In regard to the subjective class and types of relevance it becomes much more complex. Users judge relevance of retrieved information objects according to various criteria. Relevance judgement criteria are the parameters by which the users determine the relevance of the retrieved information objects.

The factors that affect relevance judgements have been a research issue for several decades, for example, the studies by Cuadra and Katter (1967a, 1967b) and Rees and Schultz (1967). Each of the studies identifies about 40 possible variables, which may influence on the judgement of relevance. However, these studies are not focusing on the relevance criteria employed by end-users in real information need situations, and this is exactly the type of studies that Schamber et al., (1990) call for. They believe ". . . that an understanding of relevance criteria, or the reasons underlying relevance judgments, as observed from the user's perspective, may contribute to a more complete and useful understanding of the dimensions of relevance" (Schamber et al., 1990, p. 771). In 1994, Schamber (1994, p. 11) published a compiling list of 80 relevance criteria suggested in the literature. Schamber does not regard the list of the 80 relevance criteria a complete list, as she points out that "[the] statement made earlier—that relevance is a multidimensional phenomenon—is, of course, a gross understatement. In fact, so many factors have been suggested as affecting relevance judgements that it is not possible to list them all here. The 80 factors listed . . ., however, represent a reasonable sample" (Schamber, 1994, p. 19). Several studies have been published on user employed relevance criteria by, for example, Park (1993), Barry (1994), Barry and Schamber (1998), and Bateman (1998). Park (1993) studied the criteria employed by 10 academic users who should make selection decisions when presented with lists of bibliographic citations. Park (1993, pp. 330–341) identified the following criteria: (1) interpretation of citation including title, author name, journal name etc.; (2) internal (experience) context including user's previous experience and perceptions, and user's level of expertise in the problem area; (3) external (search) context including perceptions of the search quality, purpose of search, perception about the availability of information, etc.; (4) problem (content) context including the user's motivation underlying the intended use of a citation. The objective of Barry's study ". . . was to describe the criteria mentioned by [academic] users evaluating the information within [printed] documents as it related to the users' information need situations" (Barry, 1994, p. 149). Barry (1994, pp. 153–157) found 23 criteria that she groups into seven broad criteria groups: (1) the information content of documents; (2) user's previous experience and background; "3) user's belief and preferences; (4) other information and sources within the information environment; (5) sources of documents; (6) documents as a physical entity; and (7) the user's situation. Lately, Barry and Schamber (1998) have published a research paper,

which '. . . compares and contrasts the results of two empirical studies in which criteria were elicited directly from individuals who were seeking information to resolve their own information problem" (Barry & Schamber, 1998, p. 219). Barry contributes with the identified criteria from her academic users, and Schamber has her set of criteria from occupational users of weather-related information in a multimedia environment. The comparative study resulted in a considerable overlap of criteria identified and shared by the test persons of the two studies. The criteria common to both studies were categorized into 10 groups of criteria. The few criteria that did not result in an overlap were due to the differences in the situational contexts and the research task requirements (Barry & Schamber, 1998, p. 234). The important finding of the comparative study by Barry and Schamber (1998) is that two different groups of users in different work environments share relevance criteria. This provides evidence for the existence of a finite range of criteria, which are applied across types of users, information problem situations, and information sources. Bateman (1998) reports on a longitudinal study of university students' relevance criteria of highly relevant assessments. The findings show that topics and topic focus did change for many respondents during the study (Bateman, 1998, p. 31). The work by Bateman has recently been confirmed by Vakkeri (2000), who, like Bateman, bases his longitudinal study of students' information-seeking behavior on the framework developed by Kulhthau (e.g., 1993). Vakkari's test persons are in the process of writing research proposals, and the study shows ". . . that different types of information are sought at different stages of the writing process and thus the contribution of the information also differs at the different stages" (Vakkari, 2000, p. 2). The work by Bateman (1998) and Vakkari (2000) emphasizes two interesting issues. Namely, that relevance criteria (may) change as the information need develops and matures, but also that an information need may be composed of several subfoci, which, consequently, are represented by different relevance criteria. The possible existence of subfoci, and priorities of subfoci, may explain the assessment behavior in cases where information objects are assessed as either partially relevant; or only parts of the object are assessed (highly) relevant.

A wide range of relevance criteria are identified in the relevance criteria research of the past; however, if we are to concentrate on the main objective of IR, the retrieval of *relevant* information, we may divide the overall group of criteria into two groups. That is, a group of IR *effectiveness* related criteria, and (2) a group of *efficiency* related criteria. The group of efficiency criteria relates to ". . . the cost (e.g., accessibility and availablility) and quality (e.g., accuracy, dept, and tangibility) of the documents" (Hertzum, Andersen, Andersen, & Hansen, 2002, p. 6). Hertzum and colleagues (2002) explain that these criteria (which we refer to as efficiency criteria) are not directly concerned with the user's information need, and consequently, do not fit into the definition of relevance as a relation between corpus of documents and different (identifiable) aspects of an information need. Hence, the group of effectiveness related criteria should be the attention of future research. That is, to advance our understanding of IR depending criteria, and relevance assessment behavior and consequently improve IR. A starting point may be the verification of possible existence of correlation between relevance criteria and relevance types put forward by Saracevic (1996).

Degrees of Relevance

Degree of relevance refers to the rating and indication of the value of relevance of a given assessed information object. The assignments of degrees of relevance are done independently of the applied classes and types of relevance, as well as the relevance criteria employed. Degrees of relevance are commonly referred to the as the assignment of binary[2] or nonbinary relevance. However, this does not refer to additional types of relevance, but refers to the employment of either binary or nonbinary relevance assignments.

Degrees of relevance could also refer to whether it is the information object as a whole, or just a part of it is considered relevant. However, this distinction is rarely made in the evaluation of IR systems.

In traditional evaluation of IR systems (i.e., according to the Cranfield model) binary relevance assignments, that is, the indication of "relevance" or "nonrelevance" of the retrieved information objects, are carried out with reference to algorithmic relevance (including intellectual topicality, e.g., see Figs. 1 and 2). The assignment of nonbinary relevance assessments are commonly used in connection with user involved experiments and investigations, and always associated with algorithmic relevance assessments in best match ranking systems.

The assignment of nonbinary relevance assessments is also known as the assignment of, for example, partial relevance, scaled relevance, rated relevance, three-valued or tri-partite relevance, with the latter examples indicating the specific scaling employed to relevance assessments. The number of employed rating categories can vary from test to test. In the case of the relevance study by Rees and Schultz (1967), an 11-point scale was employed. Eisenberg and Hu (1987) used a seven-point category scale. Three-category rating scales of relevance have been used by, for example, Saracevic (1969), Lancaster (1969), Saracevic and Su (1989), Pao (1993), Borlund and Ingwersen (1997), and Borlund (2000a). As pointed out by Tang et al. (1999, p. 256), it is not the number of categories alone that describe a rating scale because the scales may differ on how relevance categories are defined and on the style of category anchoring. Rees and Schultz (1967, pp. 87–91) presented their 11-point scale graphically as a line with 11 equally

[2] The assignment of *binary* relevance is in the literature also known as the assignment of *dichotomous* relevance (e.g., Eisenberg & Hu, 1987).

sized intervals, and with the ends of the scale labeled "completely irrelevant" and "extremely relevant." In the cases where only some of the categories are defined, the defined categories stand out as the verbal base for the undefined categories. The three categories of relevance applied by, for example, Saracevic (1969), are labeled relevant, partially relevant, and nonrelevant. Most IR experiments (e.g., Borlund, 2000a; Borlund & Ingwersen, 1997; Pao, 1993; Saracevic, 1969; Sacarevic & Su, 1989) make use of three scaled relevance categories. The initial reason for that might have been an intuitive understanding of the need for more than two categories in user-based relevance assessment in IR. Tang and Solomon (1998) document this intuitive understanding in their analysis based on the observation of one person's search and relevance assessment behavior. In their observation of the test person assessing relevance of retrieved documents, they noticed that "[s]he quickly elaborated on her initial binary classification scheme by inserting *possible relevant* between *relevant* and *not relevant* categories" (Tang & Solomon, 1998, p. 245). However, Cuadra and Katter (1967a, p. 111) as well as Rees and Schultz (1967, pp. 221–225) found that the type of scale employed, that is, category rating or continuous scales, might result in a difference in relevance results. Since then, category rating has been the most commonly employed method in IR testing. Nevertheless, the methods of continuous scale, such as magnitude estimation, have become popular with some researchers (see, e.g., Bruce, 1994; Eisenberg, 1988), who argue that these methods are more reliable than category scales and easier to administer and analyze. Findings by Eisenberg (1988) suggest that the magnitude evaluation method is less influenced by potential biases than category rating scales. However, Eisenberg and Barry (1988) found in later tests that both magnitude estimation as well as the category rating scale is affected by the order documents are presented for relevance judgement.

Attention has also been dedicated to the recommendable number of relevance categories to apply (e.g., Jacoby & Matell, 1971; Tang et al., 1999). Jacoby and Matell (1971, p. 495) explain: "[the] basic question about any given rating instrument is whether or not it has an optimum number of response categories or at least a number beyond which there is no further discrimination between the rated items . . . Too few response categories result in too coarse a scale and loss of much of the rater's discrimination powers. Conversely, too fine a scale may go beyond the rater's limited powers of discrimination." The work carried out by Tang et al. (1999, p. 263) on the identification of the optimal numbers of relevance categories suggests that the optimal numbers of relevance categories are seven. However, because minimal research has been carried out in this respect, and because previous research (Cox, 1980) has indicated that no single optimal scale is appropriate under all conditions, Tang et al. (1999) recommend a replication of their own study.

For further reading on scaling techniques and the measurement of attitudes the reader is referred to the book by Lemon (1973), or for a shorter overview, to the article by Tang et al. (1999).

In a general sense one may propose to view the variety of relevance assessment categories as properties of clarity of information needs and cognitive states. Hence, during the IR process the importance of the properties (subfoci) may shift radically, changing the (overall) focus of the information need as well as the criteria for relevance judgement. The change in focus and in the use of relevance criteria is empirically documented by, for example, Bateman (1998), Spink et al. (1998), and Vakkari and Hakala (2000). The findings by Spink et al. (1998, p. 613) resulted in the proposal of a three-dimensional spatial model of relevance, which provides an integrated view of users' relevance judgements in relation to the applied type of relevance, the category of the relevance judgement (relevant/partially relevant/partially not relevant/not relevant), and time.

The assignment and application of relevance in nonbinary manner has resulted in the call for alternative IR performance measures. The traditional performance measures of recall and precision do not allow for a nonbinary indication of how relevant the relevant information objects are, but allow only for a binary relevance representation. Thus, it is often the case in tests where nonbinary relevance judgements are applied that two or more relevance categories are merged into the binary scale of relevant and nonrelevant to facilitate the calculation of the precision and recall measures (e.g., Su, 1992). According to Schamber (1994, p. 18), the relevance categories get merged because it is assumed that no information is being lost in the merging process. Responses have been made to the call for alternative performance measures with the proposal of, for example, the measure of relative relevance (RR) and the ranked half-life indicator (RHL) (Borlund, 2000a; Borlund & Ingwerssen, 1998) as well as the measures by Järvelin and Kekäläinen (2000) of cumulative gain (CG) and cumulative gain with discount (DCG). Recently, the cumulative gain measures have been validated by Voorhees (2001). Each of the proposed measures are capable of handling nonbinary-based relevance assessments, and are put forward as supplements to the measures of recall and precision—not to substitute recall and precision (Borlund, 2000a).

The following and final section on the multidimensionality of relevance shows how the concept of relevance is dealt with also at different levels.

Levels of Relevance

Reviews of the literature on relevance reveal how the concept of relevance is viewed also according to different levels of relevance (e.g., Mizzaro, 1998; Saracevic, 1996). Saracevic (1996, p. 212) considers IR interaction to occur at several connected levels, stating "[w]e can typify relevance as it manifests itself at different levels, and we can then study its behavior and effects within and between strata." When treating relevance according to various levels or strata Saracevic accomplishes two things: (1) to distinguish

between the applied types of relevance, and (2) to specify and clarify the focus of the given relevance relationship. Saracevic (1996) outlines the following levels, each involving different elements and processes. The levels are: the user level and the computer level, which interact with each other at and via the interface surface level. The user and computer levels correspond to the classes of subjective and objective relevance, respectively. The surface level refers to the presentation of the search functions to use in the search tasks.

Like the work on levels by Saracevic the four-dimensional relevance model by Mizzaro (1998) helps clarifying specific relevance relationships. However, the elements involved differ from the pragmatic relevance levels outlined by Saracevic. The model by Mizzaro considers by means of logic the relevance relationships between information resources; representations of information need; time; and the components of topic, task, and context. The levels of relationships are clearly expressed when the four dimensional model is presented in a matrix format. Mizzaro's model can be viewed as an illustration of how relevance assessments process over time, from the user's perspective, may take place at different levels during an IR session.

Further, levels of relevance may refer to the different stages or levels at which relevance is judged by the user during the IR process. According to Kuhlthau (1993), the user looks for topical relevant documents at earlier stages, while preferring "pertinent" relevant documents at later stages when a focus is gained. One may say that it is the relevance manifestations at these levels Spink et al. (1998) verify in their examination of different regions of relevance. Their findings indicate that partially relevant documents may play an important part in the early stage of the user's IR process (Spink et al., 1998, p. 599).

Apart from helping in specifying the types of relevance and relevance relationships involved, the work on levels also shows how relevance relations are the result of interactive dynamic processes. For instance, the levels of relevance by Saracevic (1996, p. 218) illustrated in his stratified model of IR interaction, depicts the dynamic interplay of the elements involved at the various levels. In addition, Kuhlthau's proposition of *shifts* in the user's judgement from the level of topical relevant documents to the level of "pertinent" relevant documents indicates a developing and dynamic nature in the process of judging relevance.

Dynamic Relevance

As pointed out earlier, dynamic relevance refers to how the same user's perception of relevance may change over session time. The awareness of relevance to be dynamic is not new. Back in 1967, Cuadra and Katter (1967a) reported on the difficulty of controlling the variables, which affect the user's relevance judgement—because they may change over time. The proposition by Kuhlthau (1993) that the user may have different relevance preferences at different stages of the information seeking process points to this dynamic phenomenon of relevance. Swanson (1977) describes IR as a trial–error process. By trial–error, Swanson means that the request that the user states to the IR system represents only an initial guess about the attributes that a desired document is about to have. The response the user receives from the system based on the initial guess is then used to modify and improve the initial guess for another try (Swanson, 1977, p. 129). Swanson implies that relevance judgements evolve in a series of cognitive changes because IR essentially is a trial–error process. As such, the implication by Swanson (1977) is in line with the cognitive view point, and the formation and development of an information need put forward by, for example, Brookes (1975a, 1975b, 1977, 1980) and Belkin (1977, 1978). Similarly, the idea of a *change in the user's cognitive state*, which causes the effect of changes in the relevance judgement, is the underlying theoretical assumption of the concept of psychological relevance as pointed out by Harter (1992, p. 610). Harter (1992, p. 610) proposes that a relevance judgement is a psychological state in which the retrieved information objects serve as stimuli that result in cognitive changes in the user's mental state. To Harter (1992, p. 612) the '. . . relevance judgments are a function of one's mental state at the time a reference is read. They are not fixed; they are dynamic."

Recently, a number of studies have been published that provide empirical evidence for the dynamic nature of relevance (e.g., Bateman, 1998; Bruce, 1994; Park, 1993; Robins, 1997; Spink et al., 1998; Tang & Solomon, 1998; Vakkari, 2000). Park's research (1993) on the relevance criteria employed by users assessing relevance of bibliographic citations supports a view of relevance that takes information problems as dynamic and changing as citations are evaluated. The research by Bruce (1994) is based on the assumption that a user's relevance judgement might change during the IR interaction. To prove it, Bruce develops a cognitive view point-based methodology for the capturing and identification of the relevance changes. Although Bruce (1994) primarily reports on the methodology for the identification of users' possible change of relevance preferences as a result of an alteration of knowledge structure—the latter referred to as cognitive schema by Bruce, evidence is provided that supports the dynamic nature of relevance. Bruce (1994, p. 146) observes ". . . that an alteration to the problem state for subject 1 influenced a change to the cognitive schema for relevance estimation." Also, Robins (1997) provides evidence of the dynamic nature of relevance with the identification of interaction shifts between search intermediaries and users. Robins (1997, p. 126) define interaction shifts as ". . . any change in focus of the conversation between the user and the search intermediary with respect to the user's information problem." Spink et al. (1998) examine the relationship between users' relevance judgements of partial relevance and possible corresponding changes in users' information problem and information seeking processes (Spink et al., 1998, p. 606). The results of Spink and colleagues' study reveal that partially relevant

information objects provided new information that often changed the users' understanding of their information problem and the criteria used to make relevance judgements (Spink et al., 1998, p. 610). Intermediary searchers confirmed the changes independently. Further, partially relevant information objects also provided information related to changes in users' problem definition (Spink et al., 1998, p. 610). Another piece of evidence of the dynamic nature of relevance is reported on by Tang and Solomon (1998). They found evidence of change in the test person's relevance judgements as she interacted with, first, bibliographic representations of documents, and then the actual full text documents (Tang & Solomon, 1998, p. 255). Tang and Solomon conclude that "[i]t is evident from the intensive analysis of one individual here that relevance judgment is a process of meaning construction, learning, and sense making. Thus, an individual's perceptions of what constitutes needed information itself undergoes transformation and thereby may support change in the individual's judgment behavior" (Tang & Solomon, 1998, p. 255). Further, the potential dynamic nature of relevance is verified in the longitudinal studies of university students' information-seeking behavior by Bateman (1998) and Vakkari (2000), who reports on change of relevance criteria applied and types of information sources requested, respectively.

Summary on Multidimensional and Dynamic Relevance: In Relation to Evaluation of (I)IR Systems

In previous works (e.g., Borlund, 2000a, 2000b; Borlund & Ingwersen, 1997) it is proposed to apply multidimensional *and* dynamic relevance to the evaluation of IIR systems—implicitly, a cognitive user-oriented relevance approach is suggested. Thus, it is the subjective user-oriented class of relevance that is referred to. Within this class, it is the type of "situational relevance" that best fulfills the requirements to embrace both the multidimensional and dynamic nature of relevance. Cognitively speaking, situational relevance equals the individual pragmatic state of information processing (De Mey, 1980) in which the retrieved objects provide that context that supplements the work task situation (the task context) as perceived at a given point in time by a user. The "pertinence" type of relevance is cognitively founded and dynamic but does not meet the situational (contextual) dependencies that are necessary for realistic subjective interactive relevance judgements. "Intellectual topicality" alone is indeed too limited for the purpose. However, aside from situational relevance, the other types are also applicable in the experimental setting, both objective as well as the subjective types of relevance. The reason that multidimensional and dynamic relevance is required as a condition for the proposed experimental setting (Borlund, 2000a, 2000b; Borlund & Ingwersen, 1997), is due to the ambition of ensuring an evaluation process and procedure that is as close as possible to end-users' actual information-seeking and retrieval processes. For this purpose, the employment of situational relevance is obvious, as also concurrently acknowledged by Harter (1992), Campbell and van Rijsbergen (1996), Mizzaro (1998), and Reid (1999). For the same reason is the assignment of nonbinary relevance assessments the approach chosen. Studies of relevance criteria used by users to judge relevance of retrieved information objects, for example, the study by Park (1993), reveal that the relevance judgements of citations are dependent of cognitive and situational factors. This further supports the application of situational relevance. So does the dynamic and interactive nature between the levels of the relevance relationships outlined and modelled by Saracevic (1996), as well as the propositions of shifts in the user's relevance preferences during interactive IR by, for example, Kuhlthau (1993) and Bruce (1994). Furthermore, the dynamic nature of relevance is verified with empirical findings by, for example, Bruce (1994), Robins (1997), Spink et al. (1998), and Tang and Solomon (1998).

Stated briefly, it takes a special type of relevance to embody both the multidimensional and dynamic aspects of relevance required for realistic interactive IR behavior. Multidimensional and dynamic relevance aspects are defined as the unification of cognitive elements and situational factors that may result in mutual dynamic processes of relevance judgement, learning, and information need development. The type of relevance that fulfils the requirements is: situational relevance.

Situational Relevance

Originally, Wilson (1973), who was inspired by Cooper (1971), was the first to put forward the concept of situational relevance. Wilson (1973) introduced the idea that relevance judgement varies with the situation or the task. According to Wilson (1973, p. 458) the concept of relevance may concern ". . . the actual uses and actual effects of information: how people do use information, how their views actually change or fail to change consequent on the receipt of information"—referred to by Wilson as "psychological relevance." However, relevance may also be of a logical nature concerning the ". . . uses or effects only insofar as conform to standards of criticism, only as they correspond to what should happen or be done" (Wilson, 1973, p. 458). Wilson (1973, p. 457) concludes that relevance is not a single notion but multiple concepts and proposes the idea of situational relevance. He defines situational relevance as the relation between an information object and the information recipient's individual and personal view of the world and his or her situation in it (Wilson, 1973, p. 458). An information object is situational relevant if it brings about a change in the information recipient's view of his or her situation whether the change [of knowledge structure(s)] comes from the topic, or the potential utility (Wilson, 1973, p. 458).

According to Wilson, it is *information* that is the corresponding dynamic element in the relevance judgement process: ". . . a corresponding dynamic notion, that of significant situationally relevant information, which is, roughly,

information whose receipt requires a significant change in one's view of his situation" (Wilson, 1973, p. 458). This is in line with how *information* is defined as a *process* within the cognitive viewpoint (e.g., Ingwersen, 1996). Despite of the overall identical understanding and definition of situational relevance, that is, of Wilson (1973) and, for example, Schamber et al. (1990), and Saracevic (1996), a difference in opinion exists. Wilson does not perceive situational relevance as a psychological phenomenon. His definition is contrasted by the psychological relevance proposed by Harter (1992) that we use to characterize and describe the nature of situational relevance. Wilson writes ". . . items of information are situationally relevant if they answer, or help to answer, questions of concern. The 'help' is of course 'logical help' rather than 'psychological help'" (Wilson, 1973, p. 463). The difference in understanding of situational relevance, from Wilson's to our view, may be explained by the relevance revolution pointed to by Robertson and Hancock-Beaulieu (1992) that relevance should be judged against the information need (and the work task) rather than the request. With the perception of situational relevance as a nonpsychological phenomenon, Wilson may have tried to avoid the intractability of subjective view point. A solution to the problem of intractability caused by subjectivity in a relevance study is a situation description in linguistic form, for example, a simulated works task situation. Wilson explains that "[w]e shall suppose that whatever views and concerns a man has can be represented, to any desired degree of fidelity, in linguistic form; and so we shall make the natural, if hazardous, move of substituting for a man's view of his situation a hypothetical situation *description*" (Wilson, 1973, p. 461). However, Wilson does not accomplish to propose a detailed methodology of how to gather the situation description to manage it in a research setting, especially including the time dimension. Thus, we see our contributions as a further development of the work of Wilson with the dynamic and interactively founded definition of situational relevance, and with the proposal of the subcomponent of simulated work task situations for the development of information need interpretations and as the pragmatic platform for the use of situational relevance during the evaluation of IIR systems (Borlund, 2000a, 2000b; Borlund & Ingwersen, 1997). In our definition situational relevance is a user-centred, empirically based and realistic as well as potentially dynamic type of relevance. Situational relevance expresses the relationship between the user's perception of usefulness of a retrieved information object, and a specific work task situation. The employment of simulated work task situations and situational relevance is an attempt to acknowledge and allow for the subjective type of relevance under controlled circumstances in the evaluation of IIR systems.

Situational Relevance and Information Need Relationships

The introduction of this particular type of dynamic situational relevance satisfies a present demand for a cognitive-oriented type of relevance. The dynamic aspects of the demand is empirically documented with the studies of the nature of users' relevance judgement behavior (e.g., Bruce, 1994; Robins, 1997; Spink et al., 1998), and the studies of the relevance criteria users employ to judge relevance reveal that cognitive and situational factors are involved in the process (e.g., Park, 1993). The demand for situational relevance is further described with the psychological relevance by Harter (1992). Psychological relevance essentially describes the cognitive and dynamic nature of situational relevance, and emphasizes that context is necessary for realistic and dynamic relevance assessments and a corresponding development of the information need (e.g., see Fig. 3). The context, i.e., the user's *perception* of a (work task) situation, is a psychological construct that represents the user's assumptions about the world at a given moment. The context is thus dynamic and changing as contextual effects occur. The real work task situation may not change radically except via interaction between the individual and his social environment (e.g., see Fig. 3). However, this perception of the work task situation may change dynamically, according to what the user perceives of information from the information objects, including his or her own process of cognition. Basically, situational relevance share the same existential conditions required for an information need to be dynamic and able to develop, that is: (1) a given situation that allows for a cognitive perception of the given situation and further triggers an information need, as well as functions as the platform for the assessment of situational relevance; and (2) external provided information (Borlund, 2000a, p. 20).

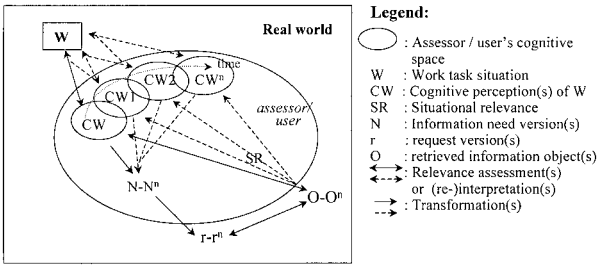

FIG. 3. The interrelationship between judgement of situational relevance and the development of the information need during a dynamic and interactive IR session (Borlund, 2000a, p. 44).

Figure 3 stresses situational relevance and the information need relationship, viewed over time. The remaining types of relevance, from algorithmic relevance to pertinence, are omitted for reasons of simplicity. Figure 3 demonstrates how the cognitive perception **(CW)** of the (simulated) work task situation **(W)** may change over time due to the assessment of situational relevance **(SR)**. Figure 3 also shows how the assessment of situational relevance **(SR)** is carried out based on both cognitive elements and situational factors. The same test person may alter his or her information need **($N–N^n$)** by different interpretations over time of the same situation **(W)**, or several test persons may interpret

the same work task **(W)** in different ways at the same time **($CW–CW^n$)**, resulting in different need versions **($N–N^n$)**.

The employment of situational relevance frames the relevance judgement as a cognitive process. The process includes not only the user's evaluation of whether a given information object **($O–O^n$)** satisfies the information need **($N–N^n$)**, it may lead to a change of the user's cognitive structures. A change in knowledge structures of the user may result in a modification of the user's perception of the information need **($N–N^n$)**. As such, the judgement of situational relevance may be a dynamic looping process resulting in a development of the information need.

Conclusions

Multidimensionality of relevance is commonly illustrated by the various relevance criteria users employ to judge relevance of retrieved information objects. In the previous sections we have shown how multidimensionality of relevance can be viewed with reference to the many different conceptions of relevance employed, for example, classes, types, degrees, criteria, and levels of relevance. In outlining the different conceptions of the multidimensionality of relevance as well as adding the aspect of dynamic relevance, we present a relevance framework, which demonstrates that a consistent and compatible understanding has been reached at an overall level. Further, future employment of the outlined framework ensures consensus on the relevance concept. In addition, by approaching the relevance concept according to multidimensionality and dynamics, the three conclusions stated by Schamber et al. (1990) is verified.

The concept of relevance is the key issue to the functioning and evaluation of IR systems. Schamber et al. (1990) show, with their comprehensive review and study of the relevance research of the past the initial coherent and indicating evidence, that changes in the perception of the concept of relevance was about to happen. Robertson and Hancock-Beaulieu (1992) confirm the change in the perception of relevance by presenting the *relevance revolution*. The result of the relevance revolution is an increasing acceptance that relevance should be judged in relation to the information need rather than the request. We agree with these conclusions and in our attempt to propose experimental conditions that enable to facilitate evaluation of IIR systems as realistically as possible, we acknowledge the multidimensional and (potential) dynamic nature of relevance, and thus suggest to make use of simulated work task situations and nonbinary judgement of situational relevance in the evaluation of IIR systems.

With the presentation of the three figures (Figs. 1–3) we present different views of relevance, even for the same types of relevance. Figure 1 displays the types of relevance employed in traditional IR, over session time. Figure 2 outlines all the possible relevance types at a given instance of an IR session, including situational relevance as viewed by Wilson (1973). The third figure, Figure 3, displays IIR over session time illustrated with the dynamic nature of situational relevance as well as the interrelationship between this type of relevance and the process of information need development. We approach relevance based on the idea of the cognitive view point in which relevance judgements evolve during the process of IR interaction. In a cognitive sense, information, as a mental construct, is equivalent to the assessment of relevance of the retrieved information objects, regardless the subjective type of relevance.

Acknowledgments

The author thanks Associate Professor Morten Hertzum (Roskilde University, Denmark) for the fruitful exchange of views with reference to the concept of relevance in IR. The present work is supported by a grant from the Research Fond of the Danish Ministry of Culture [ref. no.: A 2001 06-021(a)]; this work is carried out as a part of the TAPIR research project headed by Professor Peter Ingwersen at the Royal School of Library and Information Science, Department of Information Studies.

References

Barry, C., & Schamber, L. (1998). Users' criteria for relevance evaluation: A cross-situational comparison. Information Processing & Management, 34, 219–236.

Barry, C.L. (1994). User-defined relevance criteria: An exploratory study. Journal of the American Society for Information Science, 45, 149–159.

Bateman, J. (1998). Changes in relevance criteria: A longitudinal study. In: ASIS Proceedings. 1998 (pp. 23–32).

Belkin, N.J. (1977). Internal knowledge and external information. In M. de Mey, R. Pinxten, M. Poriau, & F. Vandamme (Eds.). International workshop on the cognitive viewpoint (pp. 187–194). Ghent: University of Ghent.

Belkin, N.J. (1978). Progress in documentation: Information concepts for information science. Journal of Documentation, 43, 55–85.

Borlund, P., & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. Journal of Documentation, 53, 225–250.

Borlund, P., & Ingwersen, P. (1998). Measures of relative relevance and ranked half-life: Performance indicators for interactive IR. In B.W. Croft, A. Moffat, C.J. van Rijsbergen, R. Wilkinson, & J. Zobel (Eds.), Proceedings of the 21st ACM SIGIR conference on research and development of information retrieval; Melbourne, Australia (pp. 324–331). ACM Press/York Press.

Borlund, P. (2000a). Evaluation of interactive information retrieval systems. Doctoral dissertation, Åbo Akademi University. Åbo: Åbo Akademi University Press.

Borlund, P. (2000b). Experimental components for the evaluation of interactive information retrieval systems. Journal of Documentation, 56, 71–90.

Brookes, B.C. (1975a). The fundamental problem of information science. In V. Horsnell (Ed.), Informatics 2 (pp. 42–49). London: Aslib.

Brookes, B.C. (1975b). The fundamental equation of information science. In Problems of information science, FID 530 (pp. 115–130). Moscow: VINITI.

Brookes, B.C. (1977). The seveloping cognitive viewpoint in information science. In M. de Mey, R. Pinxten, M. Poriau, & F. Vandamme (Eds.), International workshop on the cognitive viewpoint (pp. 195–203). Ghent: University of Ghent.

Brookes, B.C. (1980). The foundation of information science: Part I: Philosophical aspects. Journal of Information Science: Principles and Practice, 2, 125–133.

Bruce, H.W. (1994). A cognitive view of the situational dynamism of user-centered relevance estimation. Journal of the American Society for Information Science, 45, 142–148.

Campbell, I., & van Rijsbergen, K. (1996). The ostensive model of developing information needs. In P. Ingwersen & N.O. Pors (Eds.), Proceedings of CoLIS 2, second international conference on conceptions of library and information science: Integration in perspective; Copenhagen (pp. 251–268). Copenhagen: Royal School of Librarianship.

Cleverdon, C.W. (1960). Aslib Cranfield research project: Report on the first stage of an investigation into the comparative efficiency of indexing systems. Cranfield: The College of Aeronautics.

Cooper, W.S. (1971). A definition of relevance for information retrieval. Information Storage and Retrieval, 7, 19–37.

Cox, E.P. (1980). The optimal number of response alternatives for a scale: A review. Journal of Marketing Research, 17, 407–422.

Cosijn, E., & Ingwersen, P. (2000). Dimensions of relevance. Information Processing & Management, 36, 533–550.

Cuadra, C.A., & Katter, R.V. (1967a). Experimental studies of relevance judgments: Final report. Volume 1: Project summary (TM-3520/001/00). Santa Monica, CA: System Development Corp.

Cuadra, C.A., & Katter, R.V. (1967b). Opening the black box of "relevance." Journal of Documentation, 23, 291–303.

De Mey, M. (1980). The relevance of the cognitive paradigm for information science. In O. Harbo, & L. Kajberg (Eds.), Theory and application of information research. Proceedings of the 2nd international research forum on information science (pp. 49–61). London: Mansell.

Doyle, L.B. (1963). Is Relevance an adequate criterion in retrieval system evaluation? Proceedings of the American Documentation Institute; 26th annual meeting (pp. 199–200). Chicago, IL.

Eisenberg, M., & Barry, C. (1988). Order effects: A study of the possible influence of presentation order on user judgments of document relevance. Journal of the American Society for Information Science, 39, 293–300.

Eisenberg, M., & Hu, X. (1987). Dichotomous relevance judgments and the evaluation of information systems. In: ASIS Proceedings (pp. 66–69).

Eisenberg, M., & Schamber, L. (1988). Relevance: The Search for a definition. In ASIS Proceedings (pp. 164–168).

Eisenberg, M.B. (1988). Measuring relevance judgments. Information Processing & Management, 24, 373–389.

Froehlich, T.J. (1994). Relevance reconsidered: Towards an agenda for the 21st century: Introduction to special topic issue on relevance research. Journal of the American Society for Information Science, 45, 124–134.

Green, R. (1995). Topical relevance relationships: 1. Why topic matching fails. Journal of the American Society for Information Science, 46, 646–653.

Gull, C.D. (1956). Seven years of work on the organization of materials in the special library. American Documentation, 7, 320–329.

Harter, S.P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43, 602–615.

Hertzum, M., Andersen, H.H.K., Andersen, V., & Hansen, C.B. (2002). Trust in information sources: Seeking information from people, documents, and virtual agents. Interacting with Computers, 14(5), 575–599.

Hjørland, B. (1997). Information seeking and subject representation: An activity–theoretical approach to information science. Westport, CT: Greenwood Press.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. Journal of Documentation, 52, 3–50.

Jacoby, J.W., & Matell, M.S. (1971). Three-point Likert scales are good enough. Journal of Marketing Research, 8, 495–500.

Janes, J.W. (1994). Other people's judgements: A comparison of users' and others' judgements of document relevance, topicality and utility. Journal of the American Society for Information Science, 45, 160–171.

Järvelin, K., & Kekäläinen, J. (2000). IR evaluation methods for retrieving highly relevant documents. In N.J. Belkin, P. Ingwersen, & M.-K. Leong (Eds.), Proceedings of the 23rd ACM SIGIR conference on research and development of information retrieval, Athens, Greece (pp. 41–48). New York: ACM Press.

Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information science. Norwood, NJ: Ablex Publishing.

Lancaster, W.F. (1969). Medlars: Report on the evaluation of its operating efficiency. American Documentation, 20, 119–142.

Lemon, N. (1973). Attitudes and their measurement. London: B.T. Batsford.

Mizzaro, S. (1998). How many relevances in information retrieval? Interacting with Computers, 10, 303–320.

Pao, M.L. (1993). Term and citation retrieval: A field study. Information Processing & Management, 29, 95–112.

Park, T.K. (1993). The nature of relevance in information retrieval: An empirical study. Library Quarterly, 63, 318–351.

Park, T.K. (1994). Toward a theory of user-based relevance: A call for a new paradigm of inquiry. Journal of the American Society for Information Science, 45, 135–141.

Rees, A.M., & Schultz, D.G. (1967). A field experiment approach to the study of relevance assessments in relation to document searching, 2 vols. Cleveland, OH: Center for Documentation and Communication Research, School of Library Science, Case Western Reserve University.

Reid, J. (1999). A New, task-oriented paradigm for information retrieval: Implications for evaluation of information retrieval systems. In T. Aparac, T. Saracevic, P. Ingwersen, & P. Vakkari (Eds.), Proceedings of CoLIS 3, third international conference on the conceptions of library and information science: Digital libraries: Interdisciplinary concepts, challenges and opportunities, Dubrovnik. Zagreb: Zavod za informacijske studije Odsjeka za informacijske znanosti: Filozofski fakultet (pp. 97–108). Lovke: Naklada Benja.

Robertson, S.E., & Hancock-Beaulieu, M.M. (1992). On the evaluation of IR systems. Information Processing & Management, 28, 457–466.

Robins, D. (1997). Shifts of focus in information retrieval interaction. In ASIS Proceedings (pp. 123–134).

Saracevic, T. (1969). Comparative effects of titles, abstracts and full texts on relevance judgments. In ASIS Proceedings (pp. 293–299).

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, 321–343.

Saracevic, T. (1996). Relevance Reconsidered '96. In P. Ingwersen, & N.O. Pors (Eds.), Proceedings of CoLIS 2, second international conference on conceptions of library and information science: Integration in perspective, Copenhagen (pp. 201–218). Copenhagen: Royal School of Librarianship.

Saracevic, T., & Su, L. (1989). Modelling and measuring user–intermediary–computer interaction online searching: Design of study. In ASIS Proceedings (pp. 75–80).

Schamber, L. (1994). Relevance and information behavior. In M.E. Williams (Ed.), Annual review of information science and technology (ARIST) (pp. 3–48). Medford, NJ: Learned Information, Inc.

Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. Information Processing & Management, 26, 755–775.

Sperber, D., & Wilson, D. (1986). Relevance: Communication and cognition. Cambridge, MA: Harvard University Press.

Spink, A., Greisdorf, H., & Bateman, J. (1998). From highly relevant to not relevant: Examining different regions of relevance. Information Processing & Management, 34, 599–621.

Su, L.T. (1992). Evaluation measure for interactive information retrieval. Information Processing & Management, 28, 503–516.

Swanson, D.R. (1977). Information retrieval as a trial-and-error process. Library Quarterly, 47, 128–148.

Swanson, D.R. (1986). Subjective versus objective relevance in bibliographic retrieval systems. Library Quarterly, 56, 389–398.

Tang, R., Shaw, W.M., Jr., & Vevea, J.L. (1999). Towards the identification of the optimal number of relevance categories. Journal of the American Society for Information Science, 50, 254–264.

Tang, R., & Solomon, P. (1998). Towards an understanding of the dynamics of relevance judgments: An analysis of one person's search behavior. Information Processing & Management, 34, 237–256.

Thorne, R.G. (1955). The efficiency of subject catalogues and the cost of information searches. Journal of Documentation, 11, 130–148.

Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. Journal of Documentation, 56, 540–562.

van Rijsbergen, C.J. (1979). Information retrieval, 2nd ed. London: Butterworths.

Voorhees, E.M. (2001). Evaluation by highly relevant documents. In W.B. Croft, D.J. Harper, D.H. Kraft, & J. Zobel (Eds.), Proceedings of the 24th ACM SIGIR conference on research and development of information retrieval, New Orleans, LA (pp. 41–48). New York: ACM Press.

Voorhees, E.M. (1998). Variations in relevance judgments and the measurement of retrieval effectiveness. In B.C. Croft, A. Moffat, C.J. van Rijsbergen, R., Wilkinson, & J. Zobel (Eds.), Proceedings of the 21st Annual International ACM SIGIR conference on research and development in information retrieval, Melbourne, Australia (pp. 315–323). ACM Press.

Voorhees, E.M. (2000). Variations in relevance judgments and the measurement of retrieval effectiveness. Information Processing & Management, 36, 697–716.

Wilson, P. (1973). Situational relevance. Information Storage and Retrieval, 9, 457–469.