

User-Defined Relevance Criteria: An Exploratory Study

Carol L. Barry

*School of Library and Information Science, 275 Coates Hall, Louisiana State University,
Baton Rouge, LA 70803*

The objective of this study was to describe the criteria mentioned by users evaluating the information within documents as it related to the users' information need situations. Data were collected by asking users in an academic environment to evaluate representations and the full text of documents that had been retrieved specifically for each user's information need situation. Users were asked to mark the portions of the document representations or of the full text of documents that indicated to the users whether they would or would not pursue the information within documents. An open-ended interview technique was then employed to discuss each marked portion with users. The interviews were audiotaped, the tapes transcribed, and the transcriptions were content analyzed in order to identify and describe evaluation criteria.

The results indicate that the criteria employed by users included tangible characteristics of documents (e.g., the information content of the document, the provision of references to other sources of information), subjective qualities (e.g., agreement with the information provided by the document) and situational factors (e.g., the time constraints under which the user was working). The implications of this research for our understanding of the concept of relevance, and for the design and evaluation of information retrieval systems, are discussed.

Introduction

The focus throughout the following discussion is on information retrieval systems that provide users with references to printed, textual materials (i.e., bibliographic retrieval systems). The discussion does not encompass situations in which users are searching for known items, or situations in which users are routinely provided with references as part of a selective dissemination of information system. Rather, the focus is on the user who can present a request for unknown or unfamiliar information, and purposively approaches the system (or the system intermediary) in an attempt to discover such information.

This exploratory study was based on the assumption that, although the retrieval mechanism employed by information systems is based primarily on subject matching to retrieve topically appropriate documents, motivated users examining information as it relates to current information need

situations evaluate that information based on criteria beyond the topical appropriateness of documents. This research was an attempt to identify and describe those criteria.

Throughout the following discussion, there are many references to users' "information need situations." In the context of this research, an information need situation is very simply defined as some situation in which a user finds himself or herself, in which there is a need for information. This is not meant to imply that there is such a phenomenon as an "information need" which is static and unchanging. Rather, this research is based on the assumption that such an information need situation encompasses all factors that the user brings to the situation: previous knowledge, awareness of information that is available, affective or emotional factors, the expected use of the information, any time constraints under which the user is working, and so on. It is also assumed that such information need situations, as based on these types of factors, are dynamic and constantly changing.

Another assumption on which this research is based is that judgments of relevance are ultimately judgments of documents, even when those judgments are based on representations of documents provided by information retrieval systems. A decision by a user to pursue or not pursue a document, based upon the representation of the document provided by the system, is not seen as a judgment about the representation itself. Rather, it is seen as a judgment about the document itself, based upon the description provided. Decisions about relevance are not decisions about the relevance of document representations; they are decisions about, or at least predictions about, documents, which are often based solely upon the representations that are available.

Before describing the research study, the literature that led to the assumption that motivated users evaluate information based on criteria beyond the topical appropriateness of documents is discussed.

The Topical Approach to the Definition of Relevance

The primary retrieval mechanism employed by bibliographic retrieval systems is based on the matching of subject terms used to describe requests and subject terms used

to describe documents. The evaluation of these systems has traditionally been based on the ability of the system to satisfy the need that brought the user to the system. The most common measures of system performance are recall and precision. Both measures deal with the system's ability to retrieve "relevant" documents; i.e., documents that are useful to the requester in relation to the situation that prompted the request.

Notice the shift in terminology between the descriptions of what systems do and how systems are evaluated. The system matches *subject terms* that comprise *document representations* and *queries*. Evaluation is then based on the system's ability to retrieve references to *documents* that are *relevant* within the user's current *information need situation*. In order to accept such evaluations as appropriate measures of system performance, one would have to accept a number of underlying assumptions. First, it must be assumed that the subject terms used in the query can adequately describe a user's information need situation. Second, it must be assumed that the subject terms assigned to a document can adequately describe the content of the document. Finally, it must be assumed that subject matching results in the retrieval of relevant documents.

These assumptions form the basis of the topical approach to the definition of relevance, which Saracevic calls the "systems view" of information retrieval (1970). This view posits that relevance is solely a property of the internal mechanism of the system and that relevance is the result of a match between the subject terms of a query and the subject terms assigned to documents. This view implies that all documents correctly retrieved by the system will be, by definition, relevant to the user. This topical approach to the definition of relevance focuses on the internal mechanism of the retrieval system, virtually ignoring the role of the user. In the actual application of systems, however, it is the user who ultimately decides if the retrieved documents are useful and in some way satisfy the need that brought the user to the system. Early observations of user judgments of relevance indicated that criteria other than topical appropriateness were influencing users' decisions. As Saracevic states, by the late 1950s there was "...official recognition that relevance may not be just a simple system phenomenon related to the effectiveness of matching within a retrieval system" (1970, p. 116).

This recognition of the possible nontopical aspects of relevance led to two major research studies in the 1960s. Cuadra and Katter (1967a) identified 38 variables that might affect judgments of relevance. Rees and Schultz (1967) examined the qualities of judges engaged in evaluating information during various stages of biomedical research. The research study examined more than 40 variables. Although both studies examined judgments by experts rather than actual users, the results indicated the central role played by the individual in the relevance judgment process and seriously questioned the validity of a topical definition of relevance. Rees and Saracevic (1966) incorporated the findings of this research into the following characteristics of the relevance judgment process:

A sharp distinction can be made between relevance to a question and relevance to the need underlying a question;

Only the user himself can judge the relevance of the document to this need; i.e., the relevance judgment is subjective;

For [a] user a relevance judgment will change; i.e., it is not constant;

Documents 'inform' various people differently; i.e., the cognitive set has to be taken into account in relevance judgements;

Relevance is a comparative rather than a qualitative concept; i.e., it is not inherent to a document; and

Various types of judgments exist because of the different purposes for which information is required. (p. 229)

The Move toward a User-Defined Concept of Relevance

This changing view of the relevance judgment process initiated an ongoing attempt to describe, primarily on an *a priori* basis, the differences between topical appropriateness and user judgments of relevance. These discussions have emphasized the subjective aspects of relevance judgments and suggested factors other than topical appropriateness that might influence users (Beghtol, 1986; Bookstein, 1979; Boyce, 1982; Cooper, 1971, 1973, 1978; Cuadra & Katter, 1967b; Foskett, 1972; Kemp, 1974; MacMullin & Taylor, 1984; Maron, 1977; Rees, 1966; Swanson, 1977, 1986, 1988; Wilson, 1973, 1978). Several authors have developed and presented models of the information seeking process which characterize judgments of relevance as cognitive processes that are dependent upon the knowledge and perceptions of the user, and as dynamic processes in which the user's information need is a changing and fluid situation (Belkin, 1980; Belkin, Oddy, & Brooks, 1982; Dervin, 1983; Harter, 1992; Taylor, 1962, 1968, 1985, 1988).

Although various authors have chosen many different terms to refer to these concepts, and have developed slightly different models to describe the information seeking process, it is possible to summarize the consensus that has been reached among these authors:

The aboutness or topical appropriateness of documents is not a sufficient condition for judgments of relevance by the person who initially requested the information;

The evaluation of relevance is closely tied to the requester's experience, cognitive state and perceptions, and relevance can only be judged by the requester;

The requester's information need situation is typically a dynamic and fluid state which will be updated and revised as new information is received; and

Evaluations of relevance will involve interactions among various factors including but perhaps not restricted to the requester's situation and goals; the requester's knowledge level and beliefs; the information being evaluated;

the way in which information is represented; the availability of other information within the environment; and the time, effort and cost involved in obtaining information.

In recent years, empirical studies have been conducted in which relevance criteria were elicited directly from users. Park (1993) identified factors mentioned by faculty and college students examining document representations that had been retrieved for each user's specific situation. Park interprets the results as showing three major categories of variables affecting relevance assessments. The first category, internal context, reflects the user's interpretation of a citation based on his or her own prior experiences or perceptions. Examples of variables in this category included levels of expertise in a subject literature, awareness of published literature, previous research experience, and educational background. The second category, external context, includes factors that stem from the individual's search and current research. Factors included perception of the search quality, purpose of the search, perception about the availability of information, priority of information needs, stage of research, and the end product of the research. The third category, problem (content) context, examines the motivations underlying the intended uses of a citation. Such uses included obtaining definitions, background information, methodologies, frameworks for the problem, and analogies. Park also explored the ways in which users assessed citations to documents; that is, how did users interpret the information provided by titles, bibliographic citations and abstracts? Some of the factors mentioned by users included the subject matter indicated by the title, readability, author's status, quality of publications, and types of documents.

Schamber (1991) conducted a study that examined criteria mentioned by occupational users of weather information asked to discuss how they evaluated weather information presented by sources. The results of this study identified ten categories of criteria mentioned by respondents: accuracy of information; currency of information; specificity of information, including detail and concreteness; the geographic proximity of the information to the user's location; the reliability or reputation of the source of the information; the accessibility to information (e.g., the availability of information, the cost of obtaining information, the ease with which information could be obtained); the verifiability of information through other sources; clarity of presentation; dynamism, or the user's ability to interact with the system; and presentation qualities such as entertainment value and affective responses to information.

A study by Nilan et al. (Halpern & Nilan, 1988; Nilan, Peck, & Snyder, 1988) explored users' criteria for source evaluation, focusing on users in serious life- and health-threatening situations in which most sources were human or interpersonal. The criteria mentioned by users evaluating sources of information included coverage; authority or expertise; friendliness or approachability; trust or respect; self (experience, knowledge, logic); relationship of the

source to the user; power or control; social pressure; uncertainty; serendipity; appearance; confidentiality; agreement or confirmation; financial considerations; time considerations; only feasible source or method; best available source or method; ease of access; ease of use; and the source's access to technology or equipment.

The results of these studies have a number of implications. First, the assumption that factors other than topical appropriateness influence users' evaluations of information is supported. Second, there is a great deal of overlap between the criteria elicited directly from users in these studies and criteria that have been previously suggested in the literature. Finally, it appears that users are apparently able to recognize and discuss nontopical aspects of information and sources that are influencing their evaluations.

Implications for Information System Design and Evaluation

The literature and research reviewed above supports the assumption that users approach information retrieval systems in the hopes of finding information that has some meaning for them. Current bibliographic retrieval systems typically retrieve references to documents based primarily on matching the topicality of the information request and documents. Research and thinking into the relevance judgment process indicates that topicality does not automatically result in relevance for users and that users are seeking documents that have some qualities that go beyond topical appropriateness.

This does not automatically imply that retrieval based upon topicality is an inappropriate mechanism. Research into the relevance judgment process has indicated that topicality plays a significant role in the determination of relevance (Saracevic, 1970). But, again, qualities other than topicality also seem to affect the determination of relevance for users. That being the case, it is reasonable to think that retrieval mechanisms based primarily on topical matching may be failing to address the needs that users bring to the systems. Some authors have expressed this view (Swift, Winn, & Bramer, 1978a, b; Weinberg, 1987; Wilson 1978).

This growing body of literature that addresses the differences between topical appropriateness and user evaluations of relevance has culminated in numerous calls for empirical research into human information processing behavior, and into the subjective, cognitive, and dynamic aspects of the relevance judgment process. Dervin and Nilan (1986) essentially summarize this attitude in their call for a paradigm shift within the field. They describe the traditional paradigm as one in which "information is seen as objective and users are seen as input-output processors of information. It is one that searches for trans-situational propositions about the nature of the use of information systems" (p. 16). They suggest that a more fruitful alternative paradigm would be one in which "users [are seen] as beings who are constantly constructing, as beings who are free (within system constraints) to create from systems and situations whatever they choose. It focuses on how people construct

sense, searching for universal dimensions of sense-making. It focuses on understanding information uses in particular situations and is concerned with what leads up to and what follows intersections with systems. It focuses on the user" (p. 16).

Schamber, Eisenberg, and Nilan (1990) also compare the traditional and alternative perspectives of human information behavior in information science and conclude that, with the adoption of "an alternative perspective, with its emphasis on the meanings contained in internal cognitions, information science may be able at last to find clues to the meaning of relevance. We see the interaction between information users, users' situations, and information itself to be a holistic, dynamic, communicative, and interpretive phenomenon... we consider the existing model to be too linear, mechanistic, and static to serve as a valid conceptual framework for exploring the human relevance judgment process" (p. 770).

The disparity between information retrieval design (based primarily on topic matching) and evaluation measures based on a user-defined concept of relevance may also explain, at least in part, the relatively poor performance of existing information retrieval systems. Belkin (1980) states that information retrieval system performance rates average about 60% recall and 40% precision, which is far from the ideal goal of 100% recall and 100% precision. These figures are not surprising if one accepts the assumption that information retrieval systems, evaluated by any measures that are implicitly or explicitly based on the concept of user-defined relevance, are being tested for their ability to perform a function for which they were not designed. One possible solution to the evaluation problem is to accept that information retrieval systems can achieve, at best, high levels of topicality and should be evaluated on that basis alone. Another alternative, one which would advance the field of information science rather than simply admitting present limitations, would be to explore the possibility of incorporating users' relevance criteria into the retrieval mechanism itself.

What foundation is there for believing that information retrieval systems can be designed to incorporate qualities other than the aboutness of documents? Perhaps the strongest indication that information retrieval systems do not have to rely solely on topical matching is the conclusion by researchers that the document itself is the single most influential variable in the relevance judgment process (Saracevic, 1970). Although two users might judge the meaning of the same document differently (whether based upon the full text of the document or representations of the document), it is assumed that both judgments somehow derive from characteristics of that document. The argument supporting the feasibility of research into the criteria employed by users during the relevance judgment process, as a possible means of incorporating such criteria into the information retrieval mechanism, is as follows: the user within an information need situation is looking for information that possesses meaning for the user. The determination of meaning is based, not solely on the aboutness

of the information, but also on other qualities desired by the user. The user is somehow able to determine whether those qualities exist for a given document, evidenced by the user's ability to judge a document as meaningful or not, or to predict from document representations whether the document will be meaningful or not. The document itself is a central variable in the judgment process; i.e., the determination of meaning is an interaction between the qualities the user is seeking and the characteristics of the document. By identifying the characteristics that indicate to a user whether those qualities are present or not, we may be able to take information retrieval beyond the topical approach.

It may not be possible to translate all of the complex interactions that influence judgments of relevance into an information retrieval mechanism. It may be possible, however, to improve the information retrieval mechanism to some degree by attempting to incorporate clues that users can employ to detect qualities other than topical appropriateness. As stated by Wilson, "Any feature of a document at all might make it retrieval-worthy for some requester, so nothing short of the impossible complete description of a document would capture all of its possible relevance. We have to settle for less than perfection. But we do not have to settle for simple content retrieval systems, as most of our systems are" (1978, p. 23).

Problem Statement and Methodology

This research was an attempt to explore the relevance judgment process as it applies to one type of information that has traditionally been of great interest to the field: textual information provided by documents. The research question posed was as follows: What criteria allow users to determine whether connections or lack of connections exist between the information within documents and the users' information need situations? Within this study, relevance was conceptualized as any connection that existed between the users' information need situations and the information provided by documents. This concept was operationalized as users' decisions to pursue or not pursue documents. The intent of the methodology was to create, as nearly as possible, a realistic environment in which motivated users could evaluate information as it applied to real and current information need situations.

Respondents were solicited by advertising a free online search to individuals willing to participate in a research project. The respondents were 18 students and faculty at Louisiana State University, from the following university departments: Geography and Anthropology, Psychology, English, History, and Literature. The stated purposes for the respondents' requests for information included preparation for undergraduate level class assignments, graduate level class assignments, masters theses, doctoral dissertations, and professional presentations and publications.

A written search request form was used to conduct a preliminary online search for each respondent. The results of this preliminary search were used as the basis for a

presearch interview with each respondent. Based on the presearch interview, a final online search was conducted. The 15 documents for which document representations would be presented to the respondent (hereafter referred to as the *stimulus documents*) were randomly selected from the full set of citations retrieved by the final search. For each stimulus document, a set of document representations (i.e., indexing terms, abstracts, bibliographic citation, and notes as to the presence of references, tables, illustrations, etc.) was prepared for presentation to the respondent. Each respondent was also presented with the full text of three documents randomly selected from the set of 15 stimulus documents. The document representations and full text of documents comprised the *stimulus materials* that were examined by respondents.

The presentation order of the stimulus materials for each stimulus document and the presentation order of the stimulus documents were randomized. Each respondent saw all of the stimulus materials for the first document, then all of the materials for the second document, and so on. Respondents were instructed to examine the stimulus materials and to circle any portion of the stimulus materials that prompted a reaction to pursue some aspect of the document (i.e., the document itself, references provided by the document, etc.). The respondents were also instructed to circle and cross out any portion of the stimulus materials that indicated something the respondent would not pursue. Once a respondent had examined all of the stimulus documents, the researcher and respondent discussed each marked portion of the stimulus materials. For each marked item, the respondent was asked "What is it about [the item circled or crossed out] that caused you to circle it [or cross it out]?" The researcher then repeated the respondent's answer and asked if the respondent had anything else to say about this item. This process of neutral questioning was continued until the respondent was offering no new reasons for having marked an item. The interview continued until all marked items had been discussed or until time constraints forced an end to the interview (respondents had been told that the session would not take longer than two hours).

Each interview was audiotaped, and the audiotapes were transcribed and used to create a data set for each respondent. Each data set consisted of all of the responses given by one respondent. A response was defined as anything said about *one marked item* in the stimulus materials. The 18 data sets contained a total of 989 responses to 242 documents.

A content analytic technique was then used to inductively identify and describe the relevance criteria mentioned by these respondents [for full explanations of the content analytic method, see Krippendorff (1980) and Stempel (1981)]. Coding categories were developed inductively from the responses found in the raw interview data. Each category was then tested for reliability using random samples of the data coded by the researcher and one or more independent coders. Percentage of agreement in intercoder reliability was calculated as a simple ratio: the number of agreements between two coders divided by the number of possible agreements (Stempel 1981). The

minimum standard of acceptability for most content analytic studies has been established as 90% and for exploratory studies as 80% (Krippendorff, 1980). When percentages of agreement did not reach acceptable levels, the coding categories were revised and tested for reliability on different random samples of the data. This step was repeated until an acceptable level (84%) of reliability was reached. The coding categories were then used to code the entire data set. [An in-depth explanation of the development of coding categories and rules is provided in Barry (1993).]

The principle advantage of this approach to data collection and description was the opportunity it provided to observe users evaluating information within an actual information seeking and use situation. Because the respondents were motivated users involved in current information need situations and were judging potentially relevant documents that had been retrieved specifically for each respondent, it is believed that this approach was able to capture the situational factors and user perceptions involved in the relevance judgment process. The open-ended interview technique, combined with the use of a neutral questioning technique to probe responses, allowed respondents to discuss virtually any aspects of their situations or of the documents that were influencing their decisions, with no preconceived restrictions or expectations to limit their responses. The content analytic technique provided an unobtrusive means of data description, based totally upon an inductive examination of the responses.

Results

The content analysis identified 23 categories of relevance criteria mentioned by respondents. These categories, as shown in Table 1, can be grouped into broad classes of criteria that pertain primarily to the information content of documents; the user's previous experience and knowledge; the user's beliefs and preferences; other information and sources of information within the environment; sources of documents; the document as a physical entity; and the user's situation.

Criteria Pertaining to the Information Content of Documents

This first group of criterion categories pertains primarily to the information content of the document. This is not to suggest that these criteria are somehow inherent to documents. For example, different users might examine the same document and draw different conclusions as to the recency of the information. However, these criteria do seem to be primarily identifying some characteristics of the information itself.

Depth/scope is defined as the extent to which information provided by the document was in-depth and focused. Examples of responses coded for depth/scope include: "Panic order is too specific, I'm looking for a broader discussion." "It is giving a very thorough and complete view of this." "It's far too summary and generalized to

TABLE 1. Groups of criterion categories.

Grouping	Criterion categories
Criteria pertaining to the information content of documents	Depth/scope Objective accuracy/validity Tangibility Effectiveness Clarity Recency
Criteria pertaining to the user's previous experience and background	Background/experience Ability to understand Content novelty Source novelty Stimulus document novelty
Criteria pertaining to the user's beliefs and preferences	Subjective accuracy/validity Affectiveness
Criteria pertaining to other information and sources within the information environment	Consensus External verification Availability within the environment Personal availability
Criteria pertaining to the sources of documents	Source quality Source reputation/visibility
Criteria pertaining to the document as a physical entity	Obtainability Cost
Criteria pertaining to the user's situation	Time constraints Relationship with author

draw anything significant out of it." "This is telling me it's not anything in depth, just a nice little discussion."

Objective accuracy/validity is defined as the extent to which information provided by the document was accurate, correct or valid. Examples of responses coded for objective accuracy/validity include: "What this author is saying is correct, that's true." "This statement is wrong, I know it to be wrong." "It isn't valid to generalize to that extent." "Given the size of their sample, they're using the right production function here."

Tangibility is defined as the extent to which the document provided information related to real, tangible issues or the extent to which definite, proven information was provided. Examples of responses coded for tangibility include: "This is a point of the actual economic situation in Germany, the reality of the situation." "They say that recent research has shown this, it's been proven." "This is all just speculative theory." "This is just someone's opinion about what was happening, not anything concrete."

Effectiveness is defined as the extent to which a technique or procedure presented within the document was effective or successful. Examples of responses coded for effectiveness include: "They show that they used this technique successfully, it works." "Looking at the results, I'm skeptical that his technique is effective."

Clarity is defined as the extent to which information was presented in a clear or readable manner. Examples of responses coded for clarity include: "I found this to be the most readable of the articles." "Most documents from that

publisher are well written and edited." "Now we come into a morass here where he starts repeating himself, and not very well."

Recency is defined as the extent to which information provided by the document was recent and up-to-date. Examples of responses coded for recency include: "This is from 1990, that's very recent." "This is an up-to-date discussion of this subject." "This is all outdated."

Criteria Pertaining to the User's Previous Experience and Background

The second group of criterion categories pertains more to the previous experience and background of the user. In other words, what does the user already know as he or she is examining a particular document, either about the subject area in general, the information actually provided by the document, or the sources of the document?

Background/experience is defined as the degree of knowledge with which the user approached the document, as indicated by mentions of background or training in the areas discussed by the document. Examples of responses coded for background/experience include: "I don't know anything about stable isotope chemistry." "This is from linguistics and that's not my background."

Ability to understand is defined as the user's judgment that he/she would be able to understand or follow the information presented. Examples of responses coded for ability to understand include: "Geologists are so different

from what I'm doing, I can't understand what they're saying." "I won't be able to follow this." "This is something I've never understood, but here's another application of it that looks like something I could understand."

Content novelty is defined as the extent to which information provided by the document was novel to the user. Examples of responses coded for content novelty include: "I had never heard this proposition before." "This is a notion I'd not come across before." "I've heard this line too many times." "It's just that I'd heard it all before."

Source novelty is defined as the extent to which sources of the document (i.e., authors, journals, publishers, etc.) were novel to the user. Examples of responses coded for source novelty include: "I had never heard of these authors, so I would want to follow up on them." "I would be interested, not just in this article, but in looking at this journal that I've never heard of."

Stimulus document novelty is defined as the extent to which the stimulus document itself was novel to the user. Examples of responses coded for stimulus document novelty include: "This is an article he's written that I didn't know about." "I've seen other articles by this author, but not this one." "I already have this article, so I wouldn't get it."

Criteria Pertaining to the User's Beliefs and Preferences

The third group of criterion categories pertains, not so much to the user's previous background, as to the user's beliefs, attitudes, and personal preferences.

Subjective accuracy/validity is defined as the extent to which the user agrees with information within the document or the extent to which information within the document supports the user's point of view. Examples of responses coded for subjective accuracy/validity include: "I am emphatically agreeing with what the author is saying here." "This is going in a direction I would argue, I could use it to bulwark my argument." "His graph is showing an effect I disagree with." "The author and I have taken opposite points of view on this matter."

Affectiveness refers to emotional responses to any aspect of the document. Responses coded for affectiveness include: "Peter Straub is just my favorite horror author." "I simply enjoy reading articles about the footraces." "This is just an incredibly boring topic." "It's just on my part a dislike for that magazine."

Criteria Pertaining to Other Information and Sources within the Environment

The fourth group of criterion categories pertains to the relationship of the information content or sources of the document being examined to other information or sources within the environment.

Consensus within the field is defined as the extent to which there is or is not consensus within the intellectual field relating to the information within the document. Examples of responses coded for consensus include: "There is

little debate over this issue of public policy for handicapped groups, so this would not be as useful to me." "This is a subject that's very much up for debate." "There are three contentious groups involved in this issue, there's a lot of disagreement about this."

External verification is defined as the extent to which information presented by the stimulus document is supported by other sources of information. Examples of responses coded for external verification include: "They say that peer evaluation promotes sharper writing and some of my reading indicates the same thing." "I have a friend who's doing research into this and she found the same thing." "He says that this technique produces better skills and many other studies have shown that it doesn't."

Availability within the environment is defined as the extent to which information provided by the stimulus document is available through other sources. Examples of responses coded for availability within the environment include: "A lot has already been written about King's movies." "There is a lot of information out there about this one little group." "There is so little information available about Ford Madox Ford, I'll look at anything about him." "I have not been able to find any other information about this technique."

Personal availability is defined as the extent to which the user already has information like that presented within the stimulus document. This criterion category does not refer to the availability of information within the environment as a whole, but rather to the information already collected by the user. Examples of responses coded for personal availability include: "I already have quite a few articles on church history, so I don't need any more." "I don't have anything that discusses grass anatomy, so I could use that."

Criteria Pertaining to the Sources of Documents

The fifth group of criterion categories pertains to evaluations of the sources of the document, rather than the actual information content of the documents.

Source quality is defined as the extent to which general standards of quality can be assumed based on the source of a document. Examples of responses coded for source quality include: "I know this institute and their work is always good." "This is a first rate journal." "This journal will publish anything, most of it is utter rubbish."

Source reputation/visibility is defined as the extent to which the source of a document is well known or reputable. Examples of responses coded for source reputation/visibility include: "This author has a very good reputation in the field." "The editor is well known in the field of historical research into the New Deal." "This is a very prestigious organization."

Criteria Pertaining to the Document as a Physical Entity

The sixth group of criterion categories pertains to the document as a physical entity, as an object to be obtained.

Obtainability is defined as the extent to which it would be possible or easy to obtain a copy of the document. Responses coded for obtainability include: "I think I could get that, there are copies floating around the U.S." "If it's coming from the USDA that would be easy to get, it's over in the library." "This is a master's thesis, and those are just really hard to track down." "This is a science bulletin out of a university in China, I'll never get a copy of that."

Cost is defined as the extent to which there would be cost involved in obtaining a copy of the document. Responses coded for cost include: "I have to decide whether to fork over twenty dollars or thirty dollars, whatever they're asking for dissertations these days." "I wouldn't have to pay anything to get this, I can just get a copy."

Criteria Pertaining to the User's Situation

The final group of criterion categories pertains more to the situational factors influencing the user.

Time constraints is defined as the extent to which time constraints are a factor in the user's situation. Examples of responses coded for time constraints include: "Maybe I'd look at this if I were just an avid Stephen King fan, but I'm restricted by time constraints here." "If I had time I might get it, but I may run out of time."

Relationship with author is defined as the extent to which the user has a personal or professional relationship with the author of a document. This category has been included as a

situational factor because responses that included mentions of such relationships consistently described the relationship (the author is my major professor, the author is a friend of mine) and then indicated that, based on this relationship, the user wanted to be aware of what this author was doing. Such responses were very different from situations in which users predicted the content of documents or the quality of work based on the author of the document. Rather, such responses seemed to describe an effort to monitor work by friends and colleagues. Examples of responses coded for relationship with author include: "The author is a friend and an ally, so I want to see what he's doing." "The author is my major professor, so I should really look at this."

Again, the results of this study are based on 989 responses made by 18 respondents who examined a total of 242 stimulus documents. Within those 989 responses, there were 444 mentions of the criterion categories described above. Table 2 indicates the frequency with which each criterion category was mentioned, the number of documents on which a category was mentioned, and the number of respondents who mentioned each category. Table 3 illustrates the mentions of criterion categories grouped by the classes of categories described above.

It should be noted that the frequency data were affected by both the stimulus materials that were presented to respondents and by the interview procedure. It is possible that some respondents did not mention certain criterion categories simply because there was nothing within the set

TABLE 2. Criterion category frequencies.

Criterion category	Frequency	Percent	Documents	Respondents
Depth/scope	64	14.4	52	16
Objective accuracy/validity	13	2.9	12	8
Tangibility	29	6.5	22	10
Effectiveness	16	3.6	13	6
Clarity	9	2.0	8	5
Recency	25	5.6	21	6
Background/experience	19	4.3	18	10
Ability to understand	9	2.0	6	5
Content novelty	53	11.9	35	10
Source novelty	10	2.3	10	8
Stimulus document novelty	5	1.1	5	4
Subjective accuracy/validity	45	10.1	31	13
Affectiveness	25	5.6	21	9
Consensus within the field	20	4.5	17	11
External verification	19	4.3	15	9
Availability/environment	21	4.7	19	11
Personal availability	5	1.1	3	3
Source quality	14	3.2	14	7
Source reputation/visibility	18	4.1	16	8
Obtainability	10	2.3	10	3
Cost	2	0.4	2	2
Time constraints	6	1.4	6	6
Relationship with author	7	1.6	6	4
Column total	444	99.9 ^a	—	—

The frequency column indicates the total number of times a criterion category was coded. The percent column indicates the percentage of the 444 total criterion category mentions accounted for by each criterion category. The document column indicates the number of documents on which a given criterion category was coded at least once. The respondents column indicates the number of respondents who mentioned a given criterion category at least once.

^aDoes not equal 100 due to rounding error.

TABLE 3. Mentions of groups of criterion categories.

Criterion category groups	Total mentions	Percent	Respondents	Percent
Criteria pertaining to information content of the document	156	35.1	18	100.0
Criteria pertaining to user's background/experience	96	21.6	15	83.3
Criteria pertaining to user's belief and preferences	70	15.8	14	77.8
Criteria pertaining to other information and sources within the environment	65	14.6	18	100.0
Criteria pertaining to sources of the document	32	7.2	11	61.6
Criteria pertaining to the document as a physical entity	12	2.7	6	33.3
Criteria pertaining to the user's situation	13	2.9	8	44.4
Column total	444	99.9 ^a	—	—

^aDoes not equal 100 due to rounding error.

of stimulus materials to prompt such a response. The very act of anchoring responses to a given set of documents may have restricted the criterion categories mentioned by respondents. Also, the intent of the open-ended interview technique was to generate as many mentions of criterion categories as possible. There was no attempt to control for the length of responses by individuals who were simply more verbose or repetitive than other respondents. The emphasis in this research is on the identification and description of relevance criteria. The results should be viewed as descriptive and exploratory, and the frequency data should not be taken as an indication of the relative importance of criterion categories to users. However, the frequency data may be useful in suggesting possible trends among responses and avenues for further research.

Conclusions

One assumption upon which this research was based is that motivated users evaluating information within the context of a current information need situation will base their evaluations on factors beyond the topical appropriateness of documents. This assumption is supported in this research simply by the identification of the criteria mentioned by these respondents. As can be seen in Table 3, *every* respondent mentioned factors beyond the topical appropriateness of documents during their evaluation of the stimulus materials. This conclusion lends support to the arguments that situational factors other than the inherent topical content of documents influence the relevance judgment process; that the situation encompasses any factors that the user brings to the situation, such as experience, background, knowledge level, beliefs, and personal preferences; and that evaluations of individual documents take place within the larger context of the information environment.

Another assumption upon which this research was based is that there is a finite range of relevance criteria that is shared across users and situations. That is, each individual

does not possess a unique set of criteria by which information is judged. The intent of this study was to identify the full range of criteria mentioned by these users. The only means of determining that a full range had been obtained was to examine the redundancy of responses; redundancy was reached when no new mentions of criterion categories were occurring. In every possible ordering of the respondents in this study, redundancy for all criterion categories was reached after the ninth respondent had been interviewed. That is, regardless of the order in which these respondents may have been interviewed, no new criterion categories would have been mentioned once the ninth respondent had been interviewed. This is generally consistent with the findings of previous studies, in which redundancy of criteria mentions was achieved through interviews with fewer than 10 respondents (see Fletcher, 1988; Nilan & Fletcher, 1987; Schamber, 1991).

The results of this study can only be said to apply to this group of users who were evaluating printed, textual material for the purposes of preparing some written, scholarly work. It is not possible to generalize these results to other groups of users, examinations of different types of information, or other information need situations. This study was seen as an incremental step in describing the relevance judgment process. It would not be possible to examine every conceivable type of user, in every situation, examining all types of information, within a single study. Rather, it is hoped that a number of studies examining this process by various users within various situations may be synthesized to present an overall view of the relevance judgment process. A first step in this direction will be the synthesis of the findings of the three major empirical studies that have solicited relevance criteria and factors affecting judgments of relevance directly from users (this study, Park, and Schamber). Such a synthesis is currently in progress. As can be seen from the brief descriptions of the Park and Schamber studies given above, there is apparently a great deal of overlap in the criteria identified by the three studies.

Implications for Future Research

This research identified and described relevance criteria mentioned by users within an academic environment who were examining printed textual materials. Again, this research was exploratory and descriptive, and was intended to provide an incremental step toward better understanding of the relevance judgment process. Perhaps the greatest research need is for similar research into the information behavior of other user populations, examining different types of information for different purposes. The comparison of the results of this study with the work of Park and Schamber promises to yield some preliminary conclusions about the similarities of behavior across users and situations, but three studies are not enough to draw valid generalizations about all information users. Of particular concern is the fact that all of the studies were dealing with educated users who were experienced within their fields. Given the general agreement, which was validated by this research, that the user's knowledge and experience plays a central role in the evaluation of information, it seems essential to examine this evaluation process among users with varying degrees of knowledge and experience.

It is hoped that research such as this may lead to, not only a better understanding of user behavior, but also to the improvement of information retrieval systems. A first step in this goal would seem to be an identification of the clues within document representations that allow users to determine the presence or absence of desired criteria within documents. Such research might further our understanding of what clues should be presented to users of computerized information retrieval systems, and perhaps suggest areas in which the retrieval mechanism itself could be extended beyond subject matching.

The criteria elicited from users in this study were the result of users responding to information presented by document representations and the full text of documents. One question that should be asked is whether users could have predicted their responses before examining these materials. In other words, are users able to specify the criteria they desire on an *a priori* basis, and to what extent? Certainly, before there can be any discussion of changing retrieval mechanisms to include user-defined criteria, we should determine whether users can actually identify those criteria as part of the search request.

Another area of research might examine the relative importance of criteria to users. This research focused on identifying and describing criteria, not ranking the importance or weight of criteria to users. If research could identify the criteria that are the most important to or given the greatest weight by users, such results would suggest the criteria that should be given the first priority in terms of systems design.

It would be interesting to explore the connections between specific portions of document representations and mentions of criteria by respondents. Is the user's ability to predict the presence or absence of specific criteria depen-

dent upon the specific document representations provided? Obviously, this is the case for some of the criteria identified by this study. For example, the user's ability to judge the reputation of the author depends upon being presented with information about the author. However, for most of the criteria identified in this study, the connections are not this obvious. The data gathered by this research is currently being examined to explore possible connections among relevance criteria and specific portions of document representations.

Yet another possible area of research is the extent to which search intermediaries might be able to elicit descriptions of desired criteria from users and then evaluate documents on the basis of those descriptions. Two studies have examined the ability of judges other than users to predict the users' judgments of relevance (Barhydt, 1967; Janes & McKinney, 1992). The judges in the studies included systems experts, subject experts, and professional search intermediaries. The findings of these studies are presented in terms of the number of matches between judges' evaluations of documents and users' evaluations of documents. Both studies found that judges cannot predict users' evaluations with certainty. Given the extent to which the evaluation process depends on subjective factors such as background knowledge, experience, beliefs, and attitudes, these findings are not surprising. However, it might be interesting to identify the criteria that can only be determined by the user and the criteria, if any, that might be successfully evaluated by an intermediary on the user's behalf.

Acknowledgments

I am grateful to Susan Bonzi, Jeffrey Katzer, Michael Eisenberg, and Michael Nilan for their constructive comments and guidance throughout the research process, and to Dialog Information Services, Inc. for generously providing database access.

References

- Barhydt, G. C. (1967). The effectiveness of non-user relevance assessments. *Journal of Documentation*, 23, 146-169.
- Barry, C. L. (1993). *The identification of user relevance criteria and document characteristics: Beyond the topical approach to information retrieval*. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Beghtol, C. (1986). Bibliographic classification theory and text linguistics: Aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, 42, 84-113.
- Belkin, N. J. (1980). The problem of 'matching' in information retrieval. In O. Harbo & L. Kajberg (Eds.), *Theory and application of information research* (pp. 187-197). London: Mansell.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval. *Journal of Documentation*, 38, 145-164.
- Bookstein, A. (1979). Relevance. *Journal of the American Society for Information Science*, 30, 269-273.
- Boyce, B. (1982). Beyond topicality: A two stage view of relevance and the retrieval process. *Information Processing & Management*, 18, 105-109.
- Cooper, W. S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval*, 7, 19-37.

- Cooper, W. S. (1973). On selecting a measure of retrieval effectiveness. *Journal of the American Society for Information Science*, 24, 87–100.
- Cooper, W. S. (1978). Indexing documents by gendanken experimentation. *Journal of the American Society for Information Science*, 29, 107–119.
- Cuadra, C. A., & Katter, R. V. (1967a). *Experimental studies of relevance judgments: Final report. Vol. I: Project summary* (NSF Report No. 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.
- Dervin, B. (1983). *An overview of sense-making research: Concepts, methods and results to date*, Paper presented to the International Communication Association, Dallas, TX.
- Dervin, B., & Nilan, M. (1986). Information needs and uses. In M. Williams (Ed.), *Annual review of information science and technology* (Vol. 21, pp. 3–33). White Plains, NY: Knowledge Industry.
- Fletcher, P. T. (1988). *An exploration of situational dimensions in the information behaviors of general managers in state government*. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Foskett, D. J. (1972). A note on the concept of "relevance." *Information Storage and Retrieval*, 8, 77–78.
- Halpern, D., & Nilan, M. (1988). A step toward shifting the research emphasis in information science from the system to the user: An empirical investigation of source evaluation behavior in information seeking and use. *Proceedings of the 51st Annual Meeting of the American Society for Information Science*, 25, 169–176.
- Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, 43, 602–615.
- Janes, J. W., & McKinney, R. (1992). Relevance judgments of actual users and secondary judges: A comparative study. *Library Quarterly*, 62, 150–168.
- Kemp, D. A. (1974). Relevance, pertinence and information system development. *Information Storage and Retrieval*, 10, 37–47.
- Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.
- MacMullin, S. E., & Taylor, R. S. (1984). Problem dimensions and information traits. *Information Society*, 3, 91–111.
- Maron, M. E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, 28, 38–43.
- Nilan, M. S., & Fletcher, P. T. (1987). Information behaviors in the preparation of research proposals: A user study. *Proceedings of the 50th Annual Meeting of the American Society for Information Science*, 24, 186–192.
- Nilan, M. S., Peek, R. P., & Snyder, H. W. (1988). A methodology for tapping user evaluation behaviors: An exploration of users' strategy, source and information evaluating. *Proceedings of the 51st Annual Meeting of the American Society for Information Science*, 25, 152–159.
- Park, T. K. (1993). The nature of relevance in information retrieval: An empirical study. *Library Quarterly* 63, 318–351.
- Rees, A. M. (1966). The relevance of relevance to the testing and evaluation of document retrieval systems. *Aslib Proceedings*, 18, 316–324.
- Rees, A. M., & Saracevic, T. (1966). The measurability of relevance. *Proceedings of the American Documentation Institute*, 3, 225–234.
- Rees, A. M., & Schultz, D. G. (1967). *A field experimental approach to the study of relevance assessments in relation to document searching. Vol. I. Final report* (NSF Contract No. C-423). Cleveland, OH: Case Western Reserve University.
- Saracevic, T. (1970). The concept of "relevance" in information science: A historical review. In T. Saracevic (Ed.), *Introduction to information science* (pp. 111–151). New York: R. R. Bowker.
- Schamber, L. (1991). *Users' criteria for evaluation in multimedia information seeking and use situations*. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing & Management* 26, 755–776.
- Stempel, G. H., III. (1981). Content analysis. In G. H. Stempel III & B. Westley (Eds.), *Research methods in mass communications*. Englewood Cliffs, NJ: Prentice-Hall, 119–143.
- 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.
- Swanson, D. R. (1988). Historical note: Information retrieval and the future of an illusion. *Journal of the American Society for Information Science*, 39, 92–98.
- Swift, D. F., Winn, V., & Bramcr, D. (1978a). 'Aboutness' as a strategy for retrieval in the social sciences. *Aslib Proceedings*, 30, 182–187.
- Swift, D. F., Winn, V., & Bramer, D. (1978b). A sociological approach to the design of information systems. *Journal of the American Society for Information Science*, 30, 215–223.
- Taylor, R. S. (1962). The process of asking questions. *American Documentation*, 13, 391–396.
- Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College and Research Libraries*, 29, 178–194.
- Taylor, R. S. (1985). Information values in decision contexts. *Information Management Review*, 1, 47–55.
- Taylor, R. S. (1988). *Value-added processes in information systems*. Norwood, NJ: Ablex.
- Weinberg, B. H. (1987). Why indexing fails the researcher. *Proceedings of the American Society for Information Science*, 24, 241–244.
- Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval*, 9, 457–471.
- Wilson, P. (1978). Some fundamental concepts of information retrieval. *Drexel Library Quarterly*, 14, 10–24.