

User Perspectives on Relevance Criteria: A Comparison among Relevant, Partially Relevant, and Not-Relevant Judgments

Kelly L. Maglaughlin and Diane H. Sonnenwald

School of Information and Library Science, University of North Carolina at Chapel Hill, CB# 3360 100 Manning Hall, Chapel Hill, NC 27599-3360. E-mail: {maglk,dhs}@ils.unc.edu

This study investigates the use of criteria to assess relevant, partially relevant, and not-relevant documents. Study participants identified passages within 20 document representations that they used to make relevance judgments; judged each document representation as a whole to be relevant, partially relevant, or not relevant to their information need; and explained their decisions in an interview. Analysis revealed 29 criteria, discussed positively and negatively, that were used by the participants when selecting passages that contributed or detracted from a document's relevance. These criteria can be grouped into six categories: abstract (e.g., citability, informativeness), author (e.g., novelty, discipline, affiliation, perceived status), content (e.g., accuracy/validity, background, novelty, contrast, depth/scope, domain, citations, links, relevant to other interests, rarity, subject matter, thought catalyst), full text (e.g., audience, novelty, type, possible content, utility), journal/publisher (e.g., novelty, main focus, perceived quality), and personal (e.g., competition, time requirements). Results further indicate that multiple criteria are used when making relevant, partially relevant, and not-relevant judgments, and that most criteria can have either a positive or negative contribution to the relevance of a document. The criteria most frequently mentioned by study participants were content, followed by criteria characterizing the full text document. These findings may have implications for relevance feedback in information retrieval systems, suggesting that systems accept and utilize multiple positive and negative relevance criteria from users. Systems designers may want to focus on supporting content criteria followed by full text criteria as these may provide the greatest cost benefit.

Introduction

Much of the current research on relevance in information retrieval focuses on what users need from information re-

trieval systems (Schamber, 1994). Attempting to capture these user needs, several studies have investigated the criteria users employ to evaluate retrieved documents (e.g., Barry 1993, 1994; Bruce, 1994; Cool, Belkin, & Kantor, 1993; Howard, 1994; Park, 1992, 1993; Regazzi, 1988; Schamber, 1991; Wang, 1994). According to these studies, participants generally indicate that documents judged as relevant met their information needs in some way, and the documents judged as not relevant failed to meet their needs. The criteria of judgments that fall somewhere in between relevant and nonrelevant have yet to be fully explored (Spink & Greidorf, 1997). Although some studies (e.g., Barry, 1993, 1994; Wang, 1994; Wang & White, 1995) have investigated criteria used to determine relevant documents, or documents participants' intend to pursue, the question remains: what does partial relevance mean to individuals and what criteria do they use when labeling a document as partially relevant?

This study investigates the use of relevance criteria in partially relevant judgments by comparing it to the use of relevance criteria in relevant and not-relevant judgments. Study participants were provided with a set of document representations that were gathered in response to their expressed information need. The study participants were then asked to select passages that contributed or detracted from a document's relevance and to judge whether each document as a whole was relevant, partially relevant, or not relevant to their information need. Participants were then interviewed and asked why they thought the passages were useful or not, and why the documents were relevant, partially relevant, or not relevant. Content analysis of the interviews revealed that the participants used 29 criteria, most of which provided positive and negative contributions to relevance, when selecting passages and determining a document's relevance. These criteria can be grouped into six categories: author, abstract, content, full text, journal or publisher, and personal.

Received January 3, 2001; Revised August 13, 2001; accepted August 13, 2001

© 2002 Wiley Periodicals, Inc.

Published online 28 January 2002 • DOI: 10.1002/asi.10049

TABLE 1. Examples of types of relevance.

System-oriented	User-oriented
Logical (Cooper, 1971)	Psychological (Wilson, 1973)
Topical (Cooper, 1971; Park, 1994)	Situational (Wilson, 1973; Harter, 1992)
Objective (Swanson, 1986; Howard, 1994)	Subjective (Swanson, 1986; Howard, 1994)

The results of this research may increase our understanding of the role criteria play in fully relevant, partially relevant, and not-relevant judgments. With an increased understanding we may be able to better use criteria and partially relevant judgments, in addition to relevant and not-relevant judgments, in interactive information retrieval systems that incorporate relevance feedback.

Previous Research

Explanations of Relevance

Before the question of partial relevance judgments can be discussed in detail, relevance itself must be addressed. The definition for relevance has been frequently discussed and debated as reflected in the different interpretations of relevance in the many articles written about relevance in the last 40 years. Detailed discussions on relevance can be found in Saracevic (1975, 1976), Schamber, Eisenberg, and Nilan (1990), and Schamber (1994). An overview is provided here as background for a discussion of criteria used in relevance judgments.

In discussing relevance, researchers have primarily taken two approaches: identification of different types of relevance, and definition of synonyms for relevance. According to Schamber (1994), relevance can be categorized as either system-oriented or user-oriented. Schamber's categories were used as guides in the following discussion of the different types of relevance discussed in the literature (Table 1). System-oriented relevance refers to a correspondence between the user's query terms and the terms that are indexed and stored in the retrieval system. It includes logical, topical, and objective relevance. Cooper (1971) uses the term logical relevance to describe a relevance decision that has little or nothing to do with the original user's judgment. "[L]ogical relevance, alias 'topical-appropriateness,' . . . has to do with whether or not a piece of information is on a subject which has some topical bearing on the information need in the question" (p. 20). Building on Cooper's definition of logical relevance, Park (1994) maintains that, "topical relevance is context-free and is based on fixed assumptions about the relationship between a topic of a document and a search question, ignoring an individual's particular context and state of needs" (p. 136).

Similarly, Swanson (1986) and Howard (1994) assert that objective relevance has very little to do with the needs of the query originator and more to do with how the system (computer or otherwise) interprets the query. According to Swanson (1986), once the query is written or "objectified"

and passed on to a search intermediary, the user's information need and the written query may no longer be closely tied: "The issue is not what the requester meant to ask but what the request itself actually said" (pp. 391–392). Howard (1994) elaborates on this by stating that objective relevance "is taken to be that relationship which is system-based and usually measured by topicality. That is, the crucial relation is how well the topic of the information request is represented in the topics of the responses" (p. 172). Objective relevance, as defined by both Swanson and Howard, is the relationship between the stated request and the response to that request. This implies that all items containing one or more query terms could conceivably be objectively relevant although IR systems typically consider the number and frequency of query terms in items. However, the user's perception of how those items relate to his or her information need is not considered when calculating objective relevance.

In contrast to logical, topical, and objective relevance, user-oriented types of relevance include subjective, situational and psychological relevance. Swanson and Howard address the relationship between the user and the items retrieved in the concept of subjective relevance. "[W]hatsoever the requester says is relevant is taken to be relevant; the requester is the final arbiter . . . because an information retrieval system exists only to serve its users" (Swanson, 1986, p. 390). In the case of subjective relevance, the originator of the request must make a value judgment on the items returned.

In addition to subjective relevance, researchers have also proposed that situational and psychological relevance are important aspects of relevance. Wilson (1973) suggests that situational relevance encompasses the circumstances surrounding the user's perception of his or her information need. "Situational relevance is relevance to a particular individual's situation—but to the situation as he sees it, not as others see it or as it 'really' is" (p. 460). With this definition Wilson, like Swanson and Howard, proposes that there are aspects of relevance that only the user can identify.

In his definition of psychological relevance, Wilson examines not only the moment the relevance judgment is made, but also effects that an item may have on the user's behavior after the judgment has been made. He states that psychological relevance "has to do with the actual uses and actual effects of information: how people use information and how their views change or fail to change consequent to the receipt of information" (p. 458). Harter (1992) also uses the term psychological relevance, and his definition is similar to Swanson's explanation of subjective relevance. Har-

ter suggests that “[users] would like to find any citation or article ‘bearing to the matter at hand’—despite whether the article is about the topic of the search” (p. 603).

Many researchers have also defined relevance with an assortment of synonyms. Tessier, Crouch, and Atherton (1977) emphasize user-oriented relevance with the use of the word “satisfaction.” Foskett (1972) proposes an additional synonym for relevance, “pertinence.” Pertinence, unlike objective relevance, ‘should be taken to mean, “adding new information to the store already in the mind of the user, which is useful to him in the work that prompted the request”’ (p. 77). Thus, for Foskett, relevance is subjective and includes novelty. Cooper (1971, 1973) proposes the synonym “utility” as an antithesis to his definition of logical relevance. According to Cooper (1973), utility is user-oriented, and is “a catch-all concept involving not only topic-relatedness but also quality, novelty, importance, credibility, and many other things” (p. 92), i.e., things of value to the user. With these synonyms, Foster and Cooper offer what, in later studies, individuals identify as their criteria for relevance.

Criteria for Relevance

After examining the types of relevance discussed in the literature and the prominent synonyms for relevance, it is obvious there is not a consensus regarding the definition of relevance. Can we approach the problem of characterizing relevance from a different perspective? Froehlich (1994) suggests that a single definition may not be the answer: “The absence of a unified definition of relevance does not mean that information scientists cannot determine the diverse criteria that people bring to systems by which to judge its output” (p. 129).

The renewed interest in user relevance criteria since 1985 (Mizzaro, 1997) seems to indicate that Froehlich is not alone in his belief that a great deal of information about relevance lies in user-defined criteria. In many studies, criteria were gathered directly from users through think-aloud protocols, interviews, and questionnaires. For example, Park (1992, 1993) interviewed study participants, asking them to discuss their evaluation of citations pertaining to their need. After analyzing the data, Park grouped participants’ evaluation criteria into three broad categories: internal (experience) context, external (search) context, and problem (content) context. Internal (experience) context encompasses the knowledge of the field currently held by the individual and his or her understanding of the current information need. External (search) context refers to criteria directly related to the current search, such as search quality and perception of availability. Problem (content) context describes criteria related to the “intended uses of the citation” (Park, 1993, p. 338) and include comparisons between the current research problem and research problems described in the citation.

Schamber (1991) examined evaluation criteria mentioned by users of weather information systems. Participants

were asked to describe work situations that required weather related information and the sources from which they sought information. They were also asked to evaluate the information received from those sources. From the analysis of this interview data, Schamber identified 10 categories of criteria. Ordered by frequency, these categories are: presentation quality, currency, reliability, verifiability, geographic proximity, specificity, dynamism, accessibility, accuracy, and clarity.

Cool, Belkin, Kantor, and Frieder (1993) combined the approaches taken by Park (1993) and Schamber (1991). They captured evaluation criteria from college freshmen by asking them to write brief explanations concerning their decision to use or not use items for a research paper. They also collected evaluation criteria from scholars through interviews about the scholar’s information needs and the documents they used to meet these needs. Their results indicate that relevance criteria for these populations fall into six categories: topic (how a document relates to a person’s interests), content/information (characteristics of what is “in” the document itself), format (formal characteristics of the document), presentation (how a document is written/presented), values (dimensions of judgment—these are modifiers of other facets), and oneself (relationship between a person’s situation and other facets).

Barry (1993, 1994) also investigated criteria used to make relevance judgments. Participants in Barry’s study were asked to evaluate document representations by circling information that would cause them to pursue the full text document or by crossing out information that would cause them not to pursue the full text document. The participants were also asked to explain why they circled or crossed out the items. Analysis of the criteria mentioned in these interviews yielded 23 criteria that were grouped into seven categories: information content, user’s previous experience and background, user’s beliefs and preferences, other information and sources within the information environment, sources of the document, document as a physical entity, and user’s satisfaction (Barry, 1994, p. 154).

Barry and Schamber (1995, 1998) compare the criteria identified in their above-mentioned studies, Schamber (1991) and Barry (1994), and found 10 criteria that overlapped (Table 2). In a series of studies, Schamber and Bateman (1996) began work on developing a user criterion scale by selecting 119 criteria from Schamber (1991), Su (1993), and Barry (1994), and asking study participants to interpret the criterion in the context of their own information-seeking process. Results from their first study suggest that the list could be reduced to 83 criterion terms. A second study was performed to explore how participants interpreted and applied the 83 criterion terms. Results from this second study indicate that the criteria can be organized into five groupings, four of which (i.e., currency, availability, clarity, and credibility) overlap with the categories suggested by Barry and Schamber (1998). Schamber and Bateman’s fifth criteria group, aboutness, was an assumed factor in Barry and Schamber (1998), but was not studied explicitly. This

TABLE 2. Overlapping relevance categories identified in studies comparing relevance criteria literature.

Barry and Schamber (1995, 1998)	Schamber and Bateman (1996)
Depth scope/specificity	
Accuracy/validity	
Clarity	Clarity
Currency	Currency
Tangibility	
Quality of source	Credibility
Accessibility	
Availability of information	Availability
Verification	
Affectiveness	
Topical appropriateness ^a	Aboutness

^a Category assumed but not studied directly.

repetition of criteria may support Barry's (1994) assumption "that there is a finite range of relevance criteria across users and situations" (p. 157).

Relevance Judgment Methods

In addition to relevance criteria, the method of obtaining relevance judgments adds yet another dimension to the problem of characterizing relevance. Many studies ask individuals to evaluate the relevance of an item based on a predetermined scale or a user-determined scale. These scales vary considerably as can be seen in the following examples: three-point scale (Janes, 1991b; Marcus, Kugel, & Benefeld, 1978; Saracevic, 1969), five-point scale (Thompson, 1973), six-point scale (Smithson, 1994), nine-point scale (Cuadra & Katter, 1967), 11-point scale (Rees & Schultz, 1967), and magnitude estimation (Bruce, 1994; Eisenberg, 1986, 1988).

The problem with this variety of scales, besides the variety itself, is that some researchers do not justify their use of them. For example, Smithson (1994) reported: "In order to avoid any ambiguity surrounding the word relevance, the user was asked to 'score' documents in terms of 'usefulness' on a six point nominal scale labeled: 6. very useful, 5. useful, 4. background interest, 3. cannot say, 2. of little use, 1. not useful" (p. 209). There are no data that illustrate the validity and reliability of such a scale.

Other studies collapsed participants' responses into two groups divided at or near the halfway point of the scales (e.g., Smithson, 1994). For example, Saracevic (1969) collapsed a three-point scale into a two-point scale by combining the "partially relevant" judgments with the "relevant" judgments. Saracevic's method was later supported by Eisenberg and Hu (1987) and Janes (1991a). In both studies, participants were asked to indicate on a 100-mm line where they would place the dividing point between relevant and not relevant. The majority of the participants in the study indicated that the break was closer to the not-relevant end of the line than to the relevant end. That is, it appears that the

center point of a given scale is not the division between relevant and not relevant; the relevant portion of the scale is actually larger than the not-relevant portion. This finding may indicate several things pertaining to the analyses of scaled relevance judgments. "One interpretation might be that collapsing categories results in underestimating relevance and performance; conversely, it could be argued that use of a two point scale over estimates relevance" (Eisenberg & Hu, 1987, p. 68).

Rees and Schultz (1967) and Janes (1993) look at user behaviors associated with scaled judgments. Before conducting their study using an 11-point scale, Rees and Schultz hypothesized that, "the end points would not be used, and an effective scale of seven or eight points would remain" (p. 117). This proved not to be the case, and the two end points were the most highly used areas. Janes (1993) compares participants' use of scale points in Rees and Schultz (1967) to those in a similar study by Cuadra and Katter (1967). In both studies the end points were used more frequently than the points in between. To explain this trend, Janes (1993) suggests that "People seemed more confident about decisions at the ends of the scales and find these judgments easy, and find decisions about 'middling' documents to be more difficult and uncertain" (p. 113). In a recent study by Tang, Vevea, and Shaw (1999), a variety of scales were compared to determine one that optimized the participant's confidence in the judgment. Although the seven-point scale was found to correlate most highly with user confidence, it was also found that regardless of scale, participants tended to utilize the end points most frequently. This may indicate that while relevance judgments can be affected by the relevance scale, the scale, in and of itself, cannot ease the decision making process when the item is neither relevant nor not relevant. This may be because the notion of middling degrees of relevance is very poorly defined, if at all, in the scales.

Partial Relevance

The aspects of relevance discussed in the previous two sections, criteria and scale, are rarely compared and discussed in the same study. The question, "What do users find in partially relevant items that make them neither relevant nor not relevant?" has yet to be examined. Bookstein (1983) suggests that judgments of partial relevance could be either a reflection of the item's degree of relevance, as Janes (1993) also suggests, or a reflection of the user's uncertainty in the item's relevance. That is, partial relevance is determined using a variety of criteria and values assigned to those criteria.

Spink and Greisdorf (1997) suggest novelty may also be a factor: "the retrieval of partially relevant items played a crucial role in providing these users with new information and directions that may lead them through further stages of their information seeking process" (p. 276). Spink, Greisdorf and Bateman (1998, 1999) also identify 15 criteria used to determine partially relevant documents: "not on the

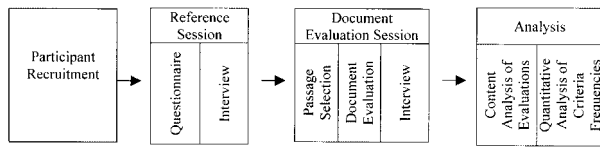


FIG. 1. Data collection and analysis process.

money,” “chronology (timeliness),” “not enough information,” “dealt only partially with the subject,” “contained multiple concepts,” “on target, but too technical,” “lists good resources,” “lists good references,” “identifies a different but related concept (new terms),” “information included too brief,” “future implications (related to current problem),” “on target, but too narrow,” “could be helpful but don’t know yet,” “could be other opportunities,” and “duplicate information.” The criteria, “duplicate or duplicate information” was found to play a role in both partially relevant and not-relevant judgments. No other criteria used in partially relevant judgments were also used in relevant or not-relevant judgments.

Although the aspects of partial relevance identified by Janes, Bookstein, and Spink, Greisdorf and Bateman may begin to shed light on the concept of partial relevance, further research is needed to increase our understanding of the relationship of partial relevance judgments to relevant and not-relevant judgments. The research reported here is a step towards developing an understanding of criteria used to evaluate relevant, partially relevant, and not-relevant judgments.

Methodology

Data Collection

To investigate the criteria people use when making relevance judgments 12 social science graduate students (10 from the Department of Sociology and 2 from the Department of History) attending the University of North Carolina at Chapel Hill participated in this study (Fig. 1). Each participant had a real information need: nine were working on their doctoral dissertation, two on their master’s thesis, and one on a paper for publication. All had done previous research on their topics.

The participants were recruited by word of mouth, flyers posted in social science departments, and e-mail messages posted on social science department listservs. The advertised incentive was a Dialog search and photocopies of all the articles deemed relevant. In individual reference sessions participants filled out a reference interview questionnaire (Appendix A) and participated in an interview. Questions posed to each participant, in both the questionnaire and subsequent, unstructured interview, attempted to gather information about: the participant’s research topic, the participant’s current knowledge of the topic, searches already conducted on the topic, the participant’s expectations of

quality/quantity for this search, and any deadlines associated with the research project.

After this initial session, one of the authors (Maglaughlin) conducted a search based on the information gathered in the initial interview and attempted to locate a minimum of 20 document representations relevant to the participant’s need. Several Dialog databases were searched, including ERIC, Sociological Abstracts, and PsycINFO. Document representations and formats varied slightly from database to database but all included the Dialog header, fields for the article or book title, author, journal or publisher name, publication date, language, and abstract (Appendix C). Between 32 and 105 document representations were found for each participant. Based on nine participants’ desire for current information and the remaining participants’ indication that currency did not matter, the 20 most recent documents representations found for each participant were chosen for use in the document evaluation session.

The document evaluation session was conducted 2 to 7 days following the reference session, and was scheduled to last for no longer than 2 hours. Participants were asked to evaluate the 20 most recent document representations, highlight passages they thought contributed to the document’s relevance, mark through the passages they considered to detract from the document’s relevance, and judge the document representations as a whole to be relevant, partially relevant or not relevant (Appendix B). Of the 244 unique documents gathered for evaluation, participants chose not to evaluate eight of them.

After evaluating the document representations (Appendix C), the participants were interviewed. During the interviews they were asked three different questions: (1) “Why did you decide to highlight or underline a passage?” (2) “Why did you mark the document representation overall as relevant, partially relevant, or not relevant?” (3) “How would you describe typical relevant, partially relevant, and not-relevant documents?” Follow-up questions were asked as needed to clarify participants’ responses to the above questions. This interview was audio taped and transcribed. Due to the 2-hour limit on the session, three participants did not have time to answer the final question. At the completion of this interview the participants were given photocopies of the document representations they had marked and a computer disk containing all of the document representations located for them. Within a week, the participants received photocopies of the articles they felt would be relevant to their research.

Data Analysis

The content of the interviews was analyzed with the intent of making “replicable and valid inference” (Krippendorff, 1980, p. 21) about the reasons participants gave for selecting passages and rating documents. The interviews were examined to identify criteria participants used when making relevant, partially relevant and not-relevant judgments. The participants’ criteria for passage selection and

TABLE 3. Definitions and examples of relevance categories, criteria, and contribution.

Criteria	Definition	Contribution	Example
<i>Abstract</i>			
Citability	abstract can be cited instead of the full text document	+	"But I might still cite it just based off the abstract."
Informativeness	abstract's ability to represent information found in the full text document	+	"This is good that they have this abstract . . . because they describe their findings in some detail."
		-	"It didn't . . . give me enough information to make a real informed decision."
<i>Author</i>			
Author novelty	participant familiarity with the author	+	"In the preliminary research that I have done . . . this guy has come up"
Discipline	author's area of research	-	"I just don't know the authors."
		+	"I'm familiar with from her work in doing this type of research."
Institutional affiliation	author's sponsor or employer	-	"And then again, I saw that it was by [author's name] . . . so, this may not be as important."
		+	"[It] is probably the key international organization"
Perceived status	perception of the author's academic standing	+	"I was kind of excited to see that it was also a prestigious author"
Accuracy-validity	quality of the research	+	"this . . . was responsibly done with a large enough sample."
		-	"[it seems] to be less . . . factual . . . and more a value judgment."
Background	background or context information	+	"This one I highlighted . . . because I am trying to contextualize"
		-	"I thought it was background information . . ."
<i>Content</i>			
Content novelty	participant familiarity with the information	+	"And they also added a new aspect . . ."
Contrast	information that contrasts with his or her own or other research.	-	"that actually is nothing new there."
		+	"that might make an interesting comparative."
Depth-scope	breadth or specificity of information covered.	+	"And the title implies that it is going to be broad in scope."
		-	"That seemed to me to be really particular and specific . . . not necessarily relevant."
Domain	field or area of study	+	"I would definitely read this article because . . . it's from a different profession."
		-	"I don't need psychological stuff here."
Citations	full text document cites notable sources.	+	"which is a really important article in my field."
Links to other information	information that could lead to additional information	+	"So, it looks like it might be a good place to look to get data on"
Relevant to other interests	information that is only useful in another context	-	"This actually is not as relevant to my dissertation but it is relevant to what I am going to be doing next."
Rarity	uniqueness of the information	+	"I haven't seen too many references about that so, that's why I marked that."
		-	"Oh, um, 'Picasso' is way over done."
Subject matter	document topic	+	"So, tripartism, is an issue I'm interested in."
Thought catalyst	information that helps to stimulate participant's thinking	-	"it's just a very different concept."
		+	"It would help me to sort of formulate my own ideas and thoughts."
<i>Full-Text Document</i>			
Audience	information indicating the intended audience	-	"That seemed to me that it was directed . . . at high school students."
Document Novelty	participant has knowledge of or has read the full-text document	+	"I have a copy of this article already."
		-	"I think I know this article . . . and it's too far off."

TABLE 3. (continued)

Criteria	Definition	Contribution	Example
Type	the form or type of artifact	+	"this one looks really relevant because it's a dissertation."
		-	"You have all these books, I can't read them all, but if I get a review essay, I can get the key content."
Possible Content	information that leads the participant to guess about the content of the full-text article	+	"So my assumption is that this article is about . . ."
		-	"My guess is that this is much more focused on a particular ownership."
Utility	whether the full-text document would be sought or not	+	"it looked like it would be something that would be worth looking at."
		-	"while that's . . . tangentially related to what I'm interested in, it's probably not enough that I would go seek out this article."
Recency	references to date of publication	+	"I mean it's relevant . . . when it was published"
		-	"it seems, it's slightly old."
<i>Journal or Publisher</i>			
Journal Novelty	participant familiarity with the journal or publisher	+	"I highlighted the journal because I didn't know about it."
		-	"Never heard of it . . . it's probably not going to be relevant."
Main focus	journal's typical content	+	"[It] is sort of like a summary of research in a broad field"
		-	"it's just more interpersonal literature."
Perceived quality	perception of the journal's or publisher's rank or quality	+	"which makes it one of the top two journals."
		-	"which is sort of a second- or third-tiered journal"
<i>Participant</i>			
Competition	Indication that the article competes with participant's work	+	"I'm nervous about this article because I was really hoping nothing had been done, so I really need to look at it."
Time requirements	whether the information would save or waste participant's time	+	"anything . . . that gives sort of a broad overview . . . saves me a ton of time."
		-	"I have so much literature that I need to go through already . . . that I have no need or time to include [this]."

document relevance were compared to the criteria identified in Schamber (1991), Park (1992, 1993), Cool et al. (1993), Barry (1993, 1994), Wang and White (1994), and Spink, Greisdorf and Bateman (1998, 1999). The combined set of criteria identified in this literature did not appear to fully capture the information discussed by the participants in this study, although there was overlap. Therefore, as suggested by Stempel (1981), a new set of codes for the participants' criteria was developed.

The expanded coding system was developed using theoretical coding methods discussed in Flick (1998). Each interview was segmented by the identified passages in the retrieved document representations and further by each separate reason the participant gave for selecting the passage. All the reasons given were examined for similarities. Similar reasons were grouped together to form a single code or criterion. Following a method similar to Cool, et al. (1993), each criterion was also labeled as "positive" or "negative,"

depending on whether the participant used the criterion as a positive or negative indication of relevance. The criteria identification was an iterative process.

A comparison of intercoder agreement was used to test the reliability of the criteria codes to fully capture information expressed by the participants. One author (Maglaughlin) and two colleagues coded portions of interviews from three participants. There was an 80% agreement between the three judges and a minimum of 88% agreement between any two of the judges. By following the formulas suggested in Cohen (1960), the Kappa coefficient of intercoder agreement of the three judges was found to be 0.72 and determined to have a 95% confidence limit. The minimum coefficient of intercoder agreement between any two judges was found to be 0.81 and determined to have a 95% confidence limit. These results were within acceptable limits so the criteria codes were used to analyze all the remaining interviews.

Limitations

This study was designed as a preliminary investigation into the use of relevance criteria across relevance judgments, and is limited by the small number of participants and document representations. Results from this study can be generalized only to similar populations, information needs, and material types (i.e., text) and subject domain. Replication of this research with additional study participants from a variety of populations, information needs, materials, and subject domains is necessary to increase its reliability and generalizability.

Results

Criteria

Our analysis of the interview content revealed 29 criteria used by the participants when selecting passages that contributed or detracted from the document's relevance and when determining the overall relevance of a document representation. Based on the focus or target of the individual criterion, the criteria were grouped into six categories: abstract, author, content, document, journal or publisher, and participant (Table 3). In addition to criteria, the contribution to the judgment of relevance, where "+" is a positive and "-" is a negative contribution, was also analyzed so that we could investigate the value of the criterion in comparison to the relevance judgment (Table 3).

Most criteria were discussed both negatively and positively with several exceptions. For example, the citability of the abstract was always mentioned with a positive connotation, such as the participant saying that he or she could cite the abstract directly without seeking the full-text document. In addition, the intended audience for the document was always referred to in a negative manner, usually by the participant indicating he or she was not a member of intended the audience.

Criteria similar to four criteria commonly identified in previous studies of relevance criteria (Barry & Schamber, 1995, 1998; Cool et al., 1993; Park, 1992, 1993), i.e., accuracy/validity, currency, depth/scope, and understandability, were identified in this study as accuracy/validity, recency, depth-scope, and informativeness. Two criteria, novelty and specificity or depth-scope, have definitions that overlap with novelty and recency (Wang, 1994), depth (Tang & Solomon, 1998), and specificity (Spink et al., 1999).

However, unlike previous studies of relevance criteria for not, partially, and fully relevant documents (e.g., Spink et al., 1999), these criteria were found to occur in more than one type of relevance judgment. For example, the criteria depth-scope, for example, "It was specific to my query" (Spink et al., p. 611), was listed only as a criteria for relevant items but in this study the criteria "depth/scope" was referred to positively by participants in both relevant and partially relevant judgments.

Comparison of Criteria Used to Evaluate Passages and Documents

Content analysis determined that, in general, the criteria mentioned by the participants and subsequent criterion categories were similar for both passage selection and document relevance judgments. The categories tend to be used in the same proportions for both passage selection and document relevance judgments (see Table 4). However, there are some notable differences in criteria usage between passage and document judgments. Author novelty, discipline, links to other information, and subject matter were mentioned more in discussions of the passages than in discussions of document relevance, with differences in frequency ranging from 3 to 12.5%. Determining whether these differences are a result of this particular sample or a specific role the criteria play in passage and document evaluation will require further research.

Similarly, participants, when evaluating a document, did not always use the same criteria for determining document relevance that they used for passage selection. Out of the 236 document representations evaluated, 129 document representations were judged using criteria that were not used when describing the selection of passages in the document. The criteria mentioned more frequently for document relevance judgments, but not for passage selection, were informativeness, depth/scope, possible content and document utility, with frequency differences ranging from 3 to 7.5%. The discrepancy between these indicates that asking participants to evaluate passages only, and not the document as a whole, does not always give the full picture as to why a particular relevance judgment was selected.

One of the most striking features of the categories is that the criteria in the content category were mentioned more than the combination of all other criteria. Not surprisingly, this seems to indicate that although there are several criteria involved in the making of relevance decisions, content related to the information need is the focus of most relevant judgments.

Co-occurrence of Criterion Categories

To investigate criterion categories, an attempt was made to determine if there were criterion categories that tended to co-occur. That is, does the presence of one category predict the presence of another category. All categories, except the positive personal category, occurred with positive content category at least 82% of the time. Other than the content categories, the other categories did not occur together more than 36% of the time. This indicates that the criterion categories, other than content, are distinct, not dependent on each other, and cannot be used to predict the presence of another category.

Criteria Usage Among Relevant, Partially Relevant, and Not-Relevant Documents

The number of criteria mentioned by participants when discussing passages was significantly higher in documents

TABLE 4. Use of category and criteria in both passage and document evaluation.

Category	%		Criteria	%		Value	%	
	Passage	Document		Passage	Document		Passage	Document
Abstract	0.9	5.5	Citability	0.2	1.7	+	0.2	1.7
			Informativeness	0.7	3.8	+	0.2	0.2
Author	10.0	3.6	Author novelty	4.3	0.8	+	2.4	0.4
			Discipline	4.7	0.8	-	1.9	0.4
						+	4.3	0.6
						-	0.4	0.2
Institutional affiliation	0.2	0.2	+	0.2	0.2			
Content	77.8	69.4	Perceived status	0.8	1.7	+	0.8	1.7
			Accuracy-validity	2.9	1.3	+	1.7	0.6
			Background	1.0	2.5	-	1.2	0.6
						+	0.9	2.5
			Content novelty	2.0	2.9	+	1.3	1.0
			Contrast	1.3	1.5	-	0.8	1.9
						+	1.2	1.5
			Depth-scope	7.0	14.3	+	0.1	0.0
						-	3.2	6.7
						-	3.8	7.5
			Domain	2.0	2.3	+	0.2	0.8
			Previous encounter	0.5	0.2	-	1.8	1.5
			Links to other information	4.0	0.6	+	0.5	0.2
			Relevant to other interests	0.3	0.8	-	4.0	0.6
			Rarity	1.3	0.8	+	0.3	0.8
Subject matter	1.3	0.8	+	1.3	0.4			
			+	38.7	26.9			
Full Text	8.0	17.4	Subject matter	52.8	40.3	+	14.1	13.2
			Thought catalyst	1.6	2.3	+	1.6	2.3
			Audience	0.5	0.2	-	0.5	0.2
			Document novelty	1.3	1.5	+	0.8	1.0
						-	0.5	0.4
			Reading value (doc. type)	1.1	0.6	+	0.7	0.6
			Possible content	2.2	5.5	-	0.4	0.0
						+	1.3	4.0
			Document utility	1.5	9.0	+	1.0	1.5
						-	1.5	6.9
			Recency	1.4	0.6	-	0.0	2.1
						+	1.1	0.6
			-	0.4	0.0			
Journal of Publisher	3.3	2.5	Journal novelty	0.4	0.2	+	0.1	0.0
			Main focus	1.3	0.6	-	0.3	0.2
						+	0.9	0.4
						-	0.5	0.2
Perceived quality	1.5	1.7	+	1.3	1.7			
Personal	1.1	1.7	Competition	0.7	0.4	+	0.3	0.0
			Time requirements	0.4	1.3	+	0.7	0.4
						+	0.3	0.0
-	0.1	1.3						

judged to be relevant than in documents judged not relevant ($p = 0.05$). This was also true for the number of criteria mentioned when discussing document relevance ($p = 0.05$). That is, participants discussed a larger number of criteria whenever they explained why a document representation was relevant. Perhaps the participants read relevant documents more closely, spend more time on relevant documents, or find it easier to talk about positive connections or associations between their information need and a document rather than negative associations

or connections. For example, it may be easier for a person to describe an object when they know what that object is, than when they do not know what the object is.

As discussed earlier, when selecting passages, criteria focusing on content were mentioned more than the combination of all other criteria (Table 5). Criteria focusing on content are also the most cited category for document selection, regardless of the type of relevance judgment (Fig. 2). This would indicate that while there are many more

TABLE 5. Use of criteria within relevance judgments.

Category	Document relevance criterion focus					
	Not relevant		Partially relevant		Relevant	
	Passage %	Document %	Passage %	Document %	Passage %	Document %
Abstract		2.17	1.56	7.07	0.67	4.62
Author	6.38	2.17	6.23	0.51	7.58	6.72
Content	87.23	71.74	82.87	71.21	76.26	67.65
Full text	5.32	13.04	7.17	19.70	9.26	15.97
Journal or publisher	1.06	2.17	1.87	1.01	4.55	3.78
Participant		8.70	0.31	0.51	1.68	1.26
Total	99.99	99.99	100.01	100.01	100	100

aspects to relevance than content, the content of a document representation is more important and/or receives more attention than any other aspect across all relevance judgments. However, the percentage of content criteria in passage selection decreased and other criteria increased (i.e., participant, author, full text, and journal or publisher) as the judgments changed from not relevant to partially relevant to relevant (Fig. 2). This would seem to indicate that content has a slightly more important role in the evaluation of not relevant and partially relevant documents than in relevant documents.

Content criteria, followed by full-text criteria, was also the most frequently used criteria during document relevance judgments regardless of the type of judgment (Fig. 2). The next most frequently used criteria varied depending on the relevance judgment. In relevant documents the third most frequently mentioned criteria was author; in partially relevant documents it was abstract criteria and in not-relevant documents it was participant criteria. The higher occurrence of participant criteria in not-relevant documents may indicate that the participant’s context plays a greater role in not-relevant documents than in the judgment of partially relevant and relevant documents.

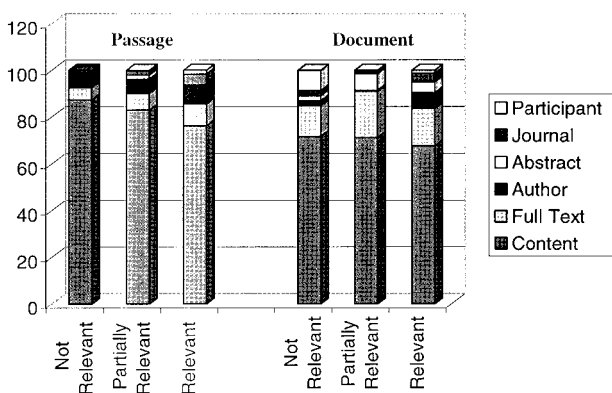


FIG. 2. Use of criteria within relevance judgments.

Value Usage Compared to Document Relevance Judgments

The document representations were also examined to see how the presence of positive and negative values of criteria differed across the document relevance judgments.

Approximately half of the document representations that were judged either relevant or not relevant contained at least one criterion with a value that contradicted the document judgment (Table 6). Only 37% of all the document representations were judged by the participants to be partially relevant, yet 65.68% of all document judgments were based on both positive and negative values of criteria. This indicates that assumptions should not be made that a document’s overall relevance judgment directly reflects the value of all the information contained in the document.

Category and Contribution Usage Compared to Document Relevance Judgments

By looking at both the category and the values of criteria in that category that the participant indicated when describing the passages or document judgments, other differences between criterion categories can be seen. When evaluating passages, criteria in every category were mentioned from a positive perspective more often than a negative perspective for passages in documents later judged to be relevant. This was not the case for passages from documents judged to be not relevant and partially relevant (Table 7). The criteria values were generally negative in not-relevant documents for all categories except journal. In partially relevant document judgments, the relationship between positive and negative aspects of the criteria varied across categories (Fig. 3) with positive values of author, content, and participant criteria being mentioned more frequently and negative values of abstract, full text, and journal criteria mentioned more frequently. This trend can be seen most clearly in the content category where the relevance of a document could be predicted by the percentage of positive or negative content criteria values ascribed to passages in the document representation (Fig. 3).

TABLE 6. Distribution of positive and negative values of criteria in document judgments.

Document relevance judgments	Total occurrence of criteria used in judgments		Only negative criteria values		Only positive criteria values		Both negative and positive values		Totals
Not relevant	38	16%	21	55.26%	0	0.00%	17	44.74%	100%
Partially relevant	87	37%	4	4.60%	6	6.90%	77	88.51%	100%
Relevant	111	47%	0	0.00%	50	45.05%	61	54.95%	100%
Totals	236	100%	25	10.59%	56	23.73%	155	65.68%	100%

As with the passage evaluation, the relationship between positive and negative values of the criteria varied across categories and types of documents relevance judgment (Fig. 4). In document relevance determination, abstract criteria were only mentioned positively while participant criteria were only mentioned negatively. This differed from passage evaluation where abstract criteria were mentioned more negatively in partially relevant judgments and more positively in relevant judgments, and participant criteria were mentioned more positively in both partially relevant and relevant judgments. This may indicate that the most noteworthy aspects of these criteria in document evaluation are the positive aspects of the abstract but the negative aspects of participant criteria.

Discussion

Towards an In-depth Understanding of Criteria: Synthesis of Criteria

Among the many challenges to the study of relevance criteria is the diverse methodology used by researchers. For example, identification of criteria may be an artifact of the study domain. The study participants in Schamber (1991) were asked to discuss a situation when they needed weather-

related information for their jobs, making geographic proximity a very important criterion. The criteria identified in studies may also be influenced by the design of the study. For example, in this study, participants were promised full-text versions of articles they deemed relevant before they evaluated the document representation. Therefore, availability, unlike in other studies (e.g., Barry, 1994; Park, 1992; Schamber & Bateman, 1998), was not a criterion in this study.

In addition to different study designs, researchers also tend to use different terms for similar criteria. For example, references to the timeliness of information are called currency (Schamber, 1991; Schamber & Bateman, 1998) and also recency (Wang, 1994). The criteria defined as the quality of a document's publisher or a journal's source is also described using a variety of terms including: reliability (Schamber, 1991), reputation/visibility (Barry, 1994), authority (Wang, 1994), credibility (Schamber & Bateman, 1998), and perceived quality, as noted in this article. Although all of these terms are valid ways of describing conceptually identical criteria, their variety makes comparing and contrasting criteria across studies difficult. Given these limitations, Table 8 is an attempt to synthesize criteria that was found in more than one of eleven studies of criteria. The synthesis is based on a comparative content analysis of

TABLE 7. Use of passage criteria and criterion values in document relevance judgments.

Criterion	Evaluation focus	Document relevance					
		Not relevant		Partially relevant		Relevant	
		Passage (%)	Document (%)	Passage (%)	Document (%)	Passage (%)	Document (%)
Abstract	+		2.17	0.31	7.22	0.51	4.68
	-			1.24		0.17	
Author	+			4.35	0.52	5.56	5.53
	-	3.30	2.17	1.86		2.02	1.28
Content	+	27.47	10.87	49.07	38.14	66.84	63.83
	-	63.74	67.39	36.34	41.24	11.62	6.81
Full Text	+	1.10	2.17	1.55	6.19	6.06	13.19
	-	3.30	4.35	2.80	5.15	1.01	0.43
Journal or publisher	+	1.10	2.17	0.62		3.37	3.83
	-			1.55	1.03	1.18	
Participant	+			0.31		1.35	
	-		8.70		0.52	0.34	0.43
Total		100.01	99.99	100	100.01	100.03	100.01

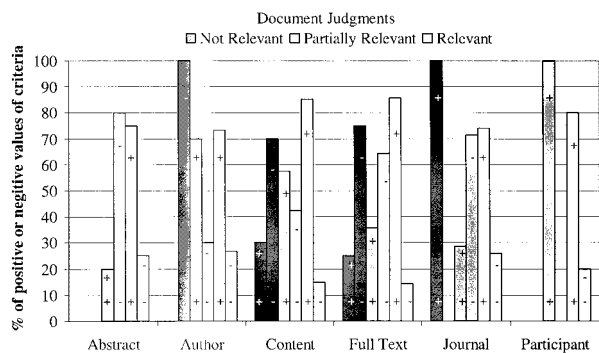


FIG. 3. Positive and negative contributions of criteria used in evaluation of passages.

the literature and the criterion definitions developed in this study.

The more frequently a criterion is identified, the more likely the criterion is applicable across document domains and situations. For example, four criteria—breadth, subject matter (topic), currency, and author—are identified in at least 10 (out of 11) studies. Three criteria—accuracy, user experience, and affectiveness—were identified in at least eight studies.

The average number of studies in which a particular criterion is identified is 6.80. This is encouraging. The field may be beginning to reach consensus regarding criteria used in making relevance judgments. Future research is necessary to investigate whether differences among studies may be attributed to the fact that different types of documents, domains, and participants were studied, or whether the differences in criteria are an indication that there are theoretical constructs yet to be discovered and understood with respect to relevance judgment criteria.

Application of Criteria in IR

Criteria Contribution

The majority of criteria found to be in common across studies were mentioned as contributing both positively and negatively to relevance judgments in our study. As illustrated in Table 6, both positive and negative aspects of criteria are used for 55% of the criteria for documents judged relevant, 89% of the criteria used for documents judged partially relevant, and 45% of the criteria used in documents judged not relevant. Furthermore, fewer than 50% of the documents judged relevant or not relevant were evaluated as having criteria that was totally positive or negative; most were judged to contain both negative and positive criteria. These findings may explain some current problems with information retrieval systems that use complete documents in relevance feedback. Calculating feedback using a document that was judged to be relevant, but in fact, is not 100% relevant, will give higher weights than may be warranted to aspects of the document that are not relevant from the user's perspective, thus reducing the system's effectiveness. Calculating relevance feedback using a

document that was judged to be relevant, but in fact, contains some passages or attributes that detract from the document's relevance, may give higher weights to the not-relevant aspects of the document, than may be warranted, reducing an IR system's effectiveness. Clearly, in addition to allowing multiple criteria to be specified, these data indicate that allowing users to specify both positive and negative aspects of these criteria may help increase the performance of relevance feedback in information retrieval systems. An additional possible solution would be to allow users of feedback retrieval systems to specify criteria for passages within documents rather than on the document as a whole.

Content Criteria

The category mentioned most frequently in identifying relevant, partially relevant, and not-relevant documents in this study is content; it includes: accuracy/validity, background, novelty, contrast, depth/scope, domain, citations, links, relevant to other interests, rarity, subject matter, and thought catalyst (Table 9). The frequency of which content criteria is used may indicate that IR systems that incorporate relevance feedback, content criteria may be appropriate to include as the highest cost/benefit category.

A second category to consider is the full text document criteria. It was the second most frequent category of criteria mentioned by participants in this study when describing partially relevant and not relevant documents (Table 5 and Fig. 2).

Novelty, uncertainty, and a smaller number of relevant criteria have all been suggested to explain partially relevant documents. Spink and Greisdorf (1997) found that most partially relevant documents contain more novel information than relevant documents. In this study, the difference between the number of novel criteria identified in partially relevant and not relevant documents was not statistically significant.

Bookstein (1983) suggests that partially relevant judgments reflect either the user's uncertainty in the item's relevance or the item's degree of relevance. Both of these theories are supported by the findings in this study. Partic-

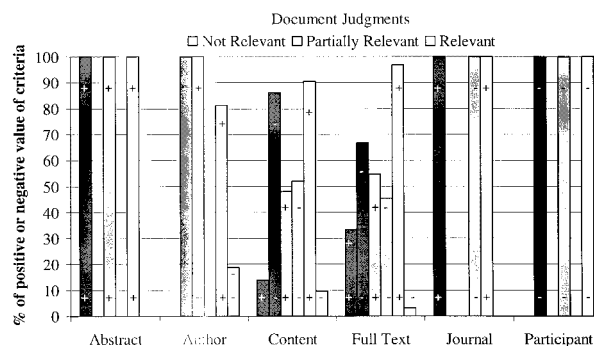


FIG. 4. Positive and negative values of criteria used in evaluation of documents.

TABLE 8. Synthesis of common concepts for relevance criteria in literature.

Category	Criteria concepts	Schamber (1991)	Park (1992)	Cool et al. (1993)	Barry (1993, 1994)	Wang (1994)	Schamber and Bateman (1996)	Tang and Soloman (1998)	Bateman (1998a, 1998b)	Spink et al. (1999)	Tang et al. (1999)	This study	Total times identified	Average per category
Author	Credibility/status	N/A	✓	✓	✓		✓		✓		✓	✓	7	7.0
Content/ topic/ aboutness	Subject matter/topic ^a	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11	
	Breadth/completeness/depth/level/ scope/specificity	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	10	
	Accuracy/credibility/quality/ validity/verifiability	✓	✓	✓	✓	✓	✓		✓		✓	✓	9	
	Clarity/presentation quality/readability/ understandability	✓	✓	✓	✓		✓		✓		✓		7	
	Novel/new information		✓	✓	✓	✓		✓		✓		✓	6	
	Connections/lists/links to other information		✓	✓	✓					✓	✓	✓	6	
	Background information		✓	✓	✓			✓			✓	✓	6	
	Methodological information		✓	✓	✓		✓				✓	✓	6	
	Stimulus/thought catalyst			✓	✓				✓			✓	4	
	Geography focus/proximity	✓						✓			✓		3	
Full text	Currency/recency/timeliness	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	10	
	Document/article type	✓	✓	✓				✓			✓	✓	7	
	Availability/accessibility/ Obtainability	✓	✓		✓	✓	✓		✓			N/A	6	
	Novelty				✓		✓	✓			✓	✓	5	
	Utility			✓				✓		✓	✓	✓	4	
Journal/ publisher/ source	Authority/quality/reliability/ reputation/value/visibility	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10	7.5
	Novelty		✓	✓	✓		✓					✓	5	
Oneself/ participant/ user	Affectiveness/appeal/competition	✓	✓	✓	✓	✓			✓	✓	✓	✓	9	
	Belief/experience/understanding		✓	✓	✓	✓		✓	✓	✓	✓	✓	8	
	Time constraints/requirements		✓		✓	✓						✓	4	

^a Category often assumed but not studied directly.

TABLE 9. Frequency of category usage across relevance judgments.

Category frequency	Not relevant	Partially relevant	Relevant
Highest	Content	Content	Content
↓	Full text	Full text	Full text
Lowest	Participant	Abstract	Author

ipants indicated that they were unsatisfied with the informativeness of the abstract more frequently in partially relevant than relevant document representations, and that they made more guesses about the content of the full-text document with partially relevant document representations. The results from this study also support theories from Bookstein (1983) and Janes (1993) that partially relevant documents are selected based on the same criteria as relevant documents, they just do not meet as many criteria or do not satisfy the criteria to the same degree.

In looking at how some of these other criteria may be used in relevance feedback, it is fortunate that the category perhaps most difficult to incorporate in an IR system, i.e., the participant category, is only frequently used when a document is judged to be not relevant (Table 5 and Fig. 2). The participant category that incorporates personal attributes of the information seeking situation would most likely also be difficult to incorporate as relevance feedback in an IR system (e.g., it is hard to imagine a system that could accurately or consistently determine if a document competes with the user's work).

Further research on feedback retrieval systems is needed to evaluate these alternatives. This study is one attempt to further our understanding of criteria and their role in relevant, partially relevant, and not-relevant judgments.

Acknowledgments

We would like to express our appreciation to Stephanie Haas for her early guidance and support in this research. The Carnegie Foundation provided partial funding for data collection. This material is based on work partially funded by the STC Program of the National Science Foundation under agreement No. CHE-987664 and NIH National Center for Research Resources, NCRR 5-P41-RR02170.

Appendix A—Reference Interview Questionnaire for Online Search

1. E-mail Address: _____
2. Do you prefer other means of contacting you? If yes, please indicate how and where? _____

3. School/Dept.: _____
4. Educational Level: _____
5. Is this the first time you have been interviewed for this purpose? _____
6. Did you ever try by yourself to search for information on similar systems? _____
7. What will the end product of your research be? **a.** paper **b.** thesis **c.** dissertation **d.** other, please specify below
8. What is your topic about? (Describe it in as much detail as you can)
9. Have you searched on this topic before? If so, what did you find? (Please describe briefly)
10. If you know any, please list the key concepts you judge to be important for your topic.
11. If you know any, please name a few journals you feel are important in the field.
12. If you know any, please name a few authors who have written on the topic. If you know any, please name a few databases you wish me to search for information on your topic.
14. What kind of materials are you looking for? (Circle as appropriate) **a.** articles **b.** books **c.** conference proceedings **d.** dissertations **e.** all
15. In what language(s) would you like the information?

16. How far back and/or current do you need the information to be? _____

Appendix B—Instructions Given to Participants at the Time of the Interview

Please read and evaluate these document representations in the following manner.

- As you are reading a document representation,
- a. Highlight any portion of it that is relevant to your research.
 - b. Underline any portion of it that is not relevant to your research.

After you have finished reading it, judge the document representation as a whole to be either “relevant,” “partially relevant,” or “not relevant” to your research and mark the letter corresponding to your overall judgment in the margin next to document representation:

- R = Relevant
- P = Partially relevant
- N = Not relevant

Appendix C—Participant's Markings on Document Representations

14/5/17
DIALOG(R)File 37: Sociological Abstr.
(c) 1998 Sociological Abstracts Inc. All rts. reserv.

476751 9613979

Mapping Interactions within and between Cohesive Subgroups

Frank, Kenneth A.
Dept Counseling/Educational Psychology/Special Education Michigan State U.
East Lansing 48824-1034
Social Networks 1996, 18, 2, Apr, 93-119.
CODEN: SONEQ4
PUB. YEAR: 1996
COUNTRY OF PUBLICATION: Netherlands
LANGUAGE: English
DOCUMENT TYPE: Abstract of Journal Article (ajaj)
Presents a model of subgroup interactions based on the work of L. Festinger, S. Schachter, & K. Back (FSB) (1950). FSB constructed a map of interactions in an apartment court by tracing the communication patterns of a planted tumor through the various physical subgroups in the court. These a priori subgroups are replaced by a posteriori subgroups, & their use of physical geography to measure distances is replaced by a metric multidimensional scale (MDS). Data from professional discussions among teachers in a high school are used to test this extension of the FSB model. Teachers (N = 23) were asked to list the 5 other teachers with whom they most often discussed professional matters, & to weight the frequency of these discussions. A metric MDS was then employed to generate a map of interactions within & between cohesive subgroups. It is suggested that this map allows a characterization of the processes at the teacher & school levels. Further, it is argued that this model accounts for these influences much better than an alternative model based on blocks of structurally similar actors. 4 Tables, 3 Figures, 99 References. Adapted from the source document. (Copyright 1996, Sociological Abstracts, Inc., all rights reserved.)
DESCRIPTORS: Social Networks (D792000); Network Analysis (D555600); Sociometric Analysis (D813400); Primary Groups (D658500); Self-Interaction (D749900); Social Influence (D789300); Group Dynamics (D340200); Comparatives (D151800); Teachers (D854100)

9/5/13

DIALOG(R)File 56: ABC Bibliographies Modern
(c) 1998 ABC-CLIO. All rts. reserv.

00242098 26-05726

Andre Verdet's secret, to make the gaze sing
Restany, Pierre
Oeil (Switzerland), no. 451, (May 1994) p. 31-8. (8 col.)
DOCUMENT TYPE: Article

Examines the work of the French poet, painter and sculptor Andre Verdet (b.1913). The author describes Verdet's upbringing, his early poetry and his involvement with the French Resistance during the Second World War. He notes the influence on his work of his mentor Pablo Picasso (1881-1973), who produced the cover for his 1955 collection of poems *Provence Noire*, with photographs by Gilles Ehrmann, and who encouraged the publication of his book of art works *Sertillages de Provence* in 1956. The author comments on Verdet's increasing interest in cosmology, and his ability to incorporate Biblical allusions and the theories of Einstein into his work.

DESCRIPTORS: Verdet, Andre; war; cosmology; Physics and Art; Art; France; Iconography, Bible; Science and Art; Painting, France; Sculpture, France; book covers; Ehrmann, Gilles; Photography, France; Picasso, Pablo; Bible

5/1.AB/61
DIALOG(R)File 7: Social SciSearch(R)
(c) 1998 Inst for Sci Info. All rts. reserv.

02913531 GENUINE ARTICLE#: RY393 NUMBER OF REFERENCES: 94
TITLE: THE 3 FACES OF INTERPERSONAL DEPENDENCY - STRUCTURAL ANALYSES OF SELF-REPORT DEPENDENCY MEASURES
AUTHOR(S): PINCUS AL; GURTMAN MB
CORPORATE SOURCE: PENN STATE UNIV, DEPT PSYCHOL, 542 MOORE BLDG/UNIVERSITY PK//PA/16802; UNIV WISCONSIN PARKSIDE, DEPT PSYCHOL/RENOSHA//W1/53141
JOURNAL: JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY, 1995, V69, N4 (OCT) , P744-758
ISSN: 0022-3514
LANGUAGE: ENGLISH DOCUMENT TYPE: ARTICLE

ABSTRACT: Using the interpersonal circumplex and the 5-factor model of personality as a structural framework, the authors identified 3 aspects of interpersonal dependency in structural analyses of the universe of content defined by the combined item pool of widely used self-report dependency measures: love dependency, exploitable dependency, and submissive dependency. Markers of these aspects of interpersonal dependency were reconstituted, and their relations with neuroticism (N), conscientiousness (C), and openness to experience (O) were examined. All forms of interpersonal dependency are related positively with N. Exploitable and submissive dependency are increasingly negatively correlated with C and O, whereas love dependency is positively correlated with C and O. Implications regarding the assessment of dependency and multifactor models of dependency are discussed, and hypotheses for future research linking dependency and depression are proposed.

References

- Barry, C.L. (1993). A preliminary examination of clues to relevance criteria within document representations. *Proceedings of the American Society for Information Science*, Columbus, OH (pp. 81-86). Medford, NJ: Learned Information, Inc.
- Barry, C.L. (1994). User-defined relevance criteria: An exploratory study. *Journal of the American Society for Information Science*, 45(3), 149-159.
- Barry, C.L., & Schamber, L. (1995). User-defined relevance criteria: A comparison of two studies. *Proceedings of the American Society for Information Science*, Chicago, IL (pp. 103-111). Medford, NJ: Information Today, Inc.
- Barry, C.L., & Schamber, L. (1998). User criteria for relevance evaluation: A cross-situational comparison. *Information Processing & Management*, 34(2/3), 219-236.
- Bateman, J. (1998a). Changes in relevance criteria: A longitudinal study. *Proceedings of the American Society for Information Science* (pp. 23-32). Medford, NJ: Information Today, Inc.
- Bateman, J. (1998b). Modeling changes in end-user relevance criteria: An information seeking study. Unpublished doctoral dissertation, University of North Texas, Denton, TX.
- Bookstein, A. (1983). Outline of a general probabilistic retrieval model. *Journal of Documentation*, 39(2), 63-72.
- 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(5), 142-148.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, 20(1), 37-46.
- Cool, C., Belkin, N.J., Kantor, P.B., & Frieder, O. (1993). Characteristics of texts affecting relevance judgments. In M.E. Williams (Ed.), *Proceedings of the 14th National Online Meeting* (pp. 77-84). Medford, NJ: Learned Information, Inc.
- Cooper, W.S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval*, 7(1), 19-37.
- Cooper, W.S. (1973). On selecting a measure of retrieval effectiveness, part 1: The "subjective" philosophy of evaluation. *Journal of the American Society for Information Science*, 24(2), 87-100.
- Cuadra, C.A., & Katter, R.V. (1967). Experimental studies of relevance judgments final report. volume 1: Project summary. Cleveland, OH: Case Western Reserve University, School of Library Science, Center for Documentation and Communication Research.

- Eisenberg, M. (1986). Magnitude estimation and the measurement of relevance. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Eisenberg, M.B. (1988). Measuring relevance judgments. *Information Processing & Management*, 24(4), 373–389.
- Eisenberg, M., & Hu, X. (1987). Dichotomous relevance judgments and the evaluation of information systems. *Proceedings of the American Society for Information Science*, Boston, MA (pp. 66–69). Medford, NJ: Learned Information.
- Flick, U. (1998). *An introduction to qualitative research*. Thousand Oaks, CA: Sage.
- Foskett, D.J. (1972). A note on the concept of “relevance.” *Information Storage and Retrieval*, 8(2), 77–78.
- 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(3), 124–133.
- Harter, S.P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, 43(9), 602–615.
- Howard, D.L. (1994). Pertinence as reflected in personal constructs. *Journal of the American Society for Information Science*, 45(3), 172–185.
- Janes, J.W. (1991a). The binary nature of continuous relevance judgments: A case study of users’ perceptions. *Journal of the American Society for Information Science*, 42(10), 754–756.
- Janes, J.W. (1991b). Relevance judgments and the incremental presentation of document representations. *Information Processing & Management*, 27(6), 629–646.
- Janes, J.W. (1993). On the distribution of relevance judgments. *Proceedings of the American Society for Information Science*, Columbus, OH (pp. 104–114). Medford, NJ: Learned Information, Inc.
- Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.
- Marcus, R.S., Kugel, P., & Benenfeld, A.R. (1978). Catalog information and text as indicators of relevance. *Journal of the American Society for Information Science*, 29(1), 15–30.
- Mizzaro, S. (1997). Relevance: The whole history. *Journal of the American Society for Information Science*, 48(9), 810–832.
- Park, T.K. (1992). The nature of relevance in information retrieval: An empirical study. Unpublished doctoral dissertation, School of Library and Information Science, Indiana University, Bloomington, IN.
- Park, T.K. (1993). The nature of relevance in information retrieval: An empirical study. *The 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(3), 135–141.
- Rees, A.M., & Schultz, D.G. (1967). A field experimental approach to the study of relevance assessments in relation to document searching: Final report: Volume 1. Cleveland, OH: Case Western Reserve University, School of Library Science, Center for Documentation and Communication Research.
- Regazzi, J.J. (1988). Performance measures for information retrieval systems: An experimental approach. *Journal of the American Society for Information Science*, 39(4), 235–251.
- Saracevic, T. (1969). Comparative effects of titles, abstracts and full texts on relevance judgments. *Proceedings of the American Society for Information Science*, San Francisco, CA (pp. 293–299). Westport, CT: Greenwood Publishing Corporation.
- 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. (1976). Relevance: A review of the literature and a framework for thinking on the notion in information science. *Advances in Librarianship*, 6, 79–138.
- Schamber, L. (1991). Users’ criteria for evaluation in a multimedia environment. *Proceedings of the American Society for Information Science*, Washington, DC (pp. 126–133). Medford, NJ: Learned Information, Inc.
- Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology*, 29, 3–48.
- Schamber, L., & Bateman, J. (1996). User criteria in relevance evaluation; Toward development of a measurement scale. *Proceedings of the American Society for Information Science*, Baltimore, MD (pp. 218–225). 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(6), 755–776.
- Smithson, S. (1994). Information retrieval evaluation in practice: A case study approach. *Information Processing & Management*, 30(2), 205–221.
- Spink, A., & Greisdorf, H. (1997). Users’ partial relevance judgments during online searching. *Online and CDROM Review*, 21(5) 271–279.
- Spink, A., Greisdorf, H., & Bateman, J. (1998). Examining different regions of relevance: From highly relevant to not relevant. *Proceedings of the American Society for Information Science*, Columbus, OH (pp. 3–12). Medford, NJ: Learned Information, Inc.
- Spink, A., Greisdorf, H., & Bateman, J. (1999). From highly relevant to not relevant: Examining different regions of relevance. *Information Processing & Management*, 34(4), 599–621.
- Stempel, G.H. (1981). Content analysis. In G.H. Stempel & B.H. Westley (Eds.), *Research methods in mass communication* (pp. 119–131). Englewood Cliffs, NJ: Prentice-Hall.
- Su, L.T. (1993). Is relevance an adequate criterion for retrieval system evaluation: An empirical inquiry into user’s evaluation. *Proceedings of the American Society for Information Science* (pp. 93–103). Medford, NJ: Learned Information, Inc.
- Swanson, D.R. (1986). Subjective versus objective relevance in bibliographic retrieval systems. *The Library Quarterly*, 56, 389–398.
- Tang, R., & Solomon, P. (1998). Toward an understanding of the dynamics of relevance judgment: An analysis of one person’s search behavior. *Information Processing & Management*, 34(2/3), 237–256.
- Tang, R., Vevea, J.L., & Shaw, W.M. (1999) Towards the Identification of the optimal number of relevance categories. *Journal of the American Society for Information Science*, 50(3), 254–264.
- Tessier, J.A., Crouch, W.W., & Atherton, P. (1977). New measures of user satisfaction with computer-based literature searches. *Special Libraries*, 68(11), 383–389.
- Thompson, C.W.N. (1973). The functions of abstracts in the initial screening of technical documents by users. *Journal of the American Society for Information Science*, 24(4), 270–276.
- Wang, P. (1994). A cognitive model of document selection of real users of information retrieval systems. Unpublished doctoral dissertation, University of Maryland, College Park, MD.
- Wang, P.L., & White M.D. (1995). Document use during a research-project: A longitudinal study. *Proceedings of the American Society for Information Science*, Columbus, OH (pp. 181–188). Medford, NJ: Learned Information, Inc.
- Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval*, 9(8), 457–471.