

THE USE OF RELEVANCE CRITERIA IN PARTIALLY RELEVANT DOCUMENTS

by
Kelly L. Maglaughlin

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science

Chapel Hill, North Carolina

April, 1999

Approved by:

Advisor

Kelly L. Maglaughlin. The Use of Relevance Criteria in Partially Relevant Documents. A Master's paper for the M.S. in I.S. degree. April, 1999. 55 pages. Advisor: Stephanie W. Haas.

This study investigates the use of relevance criteria in partially relevant documents by comparing it to the use of relevance criteria in relevant and not relevant documents. Participants, 12 social science graduate students, selected relevant and not relevant passages within 20 document representations and judged each document representation as a whole to be: relevant, partially relevant or non-relevant to their information need. Content analysis revealed 31 criteria, discussed positively and negatively, used by the participants when selecting passages and determining the document's relevance. Results indicate 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. Additionally, fewer than 50% of the documents judged relevant or not relevant were totally relevant or totally not relevant. These findings suggest possible solutions to problems with relevance feedback in information retrieval systems.

Headings:

Information needs -- evaluation

Information Retrieval

Relevance Feedback

Relevance Judgments -- evaluation

Acknowledgments

I would like to express my deepest appreciation to Dr. Stephanie Haas for her guidance, support, and unfailing ability to ask the right questions. Special thanks go to Lokman Meho for his willingness to listen and to share both his encyclopedic knowledge and extensive collection of articles. In addition to Lokman, I would also like to thank Seung-Lye Kim and Rong Tang for helping me to clarify my ideas and to develop a viable coding schema. I am also indebted to the Carnegie Foundation for a grant that funded the hundreds of photocopies needed to carry out this research. Most importantly, I am grateful to Barry for going above and beyond the call of duty with his unshakable patience and support and to my friends and family for their encouragement.

Table of Contents

1.0 Introduction	1
2.0 Previous Research	3
2.1 EXPLANATIONS OF RELEVANCE	3
2.2 CRITERIA FOR RELEVANCE	7
2.3 DEGREE OF RELEVANCE	10
2.3 PARTIAL RELEVANCE	12
3.0 Methodology	13
3.1 DATA COLLECTION	13
3.2 CONTENT ANALYSIS	15
4.0 Results	17
4.1 CRITERIA	17
4.1.1 Participant	17
4.1.2 Full text document	18
4.1.3 Author	19
4.1.4 Journal/Publisher	20
4.1.5 Abstract	21
4.1.6 Content	22
4.2 CRITERIA USAGE	24
4.2.1 Criteria for highlighting or underlining a passage	25
4.2.2 Criteria for judging the overall relevance of a document representation	27
4.2.3 Combined criteria for passage selection and judging the overall relevance of a document representation	30
4.2.4 Criteria used in describing “typical” relevant, partially relevant, and not relevant documents	32
5.0 Discussion	34
6.0 Limitations	36
7.0 References	37
8.0 Appendices	45
8.1 APPENDIX A – REFERENCE INTERVIEW FOR ONLINE SEARCH	46
8.2 APPENDIX B – SAMPLE DOCUMENT REPRESENTATION	48
8.3 APPENDIX C – EVALUATION INSTRUCTIONS	49
8.4 APPENDIX D – CATEGORY, CRITERIA, VALUE AND PARTICIPANT USAGE	50

1.0 Introduction

In our 40-year quest to create the ideal information retrieval system, it has only been in recent years that the user has come to be viewed as something more than a tool to evaluate the system. This shift toward the user perspective has, to some degree, influenced the way the term “relevance” is perceived. Much of the current research on relevance now focuses on what real users need from information retrieval systems (Schamber, 1994). Attempting to capture these user needs, several studies have been conducted on the criteria users employ to evaluate retrieved documents (Mizzaro, 1997). Unfortunately, little has been done to compare user criteria to the scale with which the user evaluated the document. We will not be able to fully utilize searcher criteria until we understand how criteria correspond to various relevance judgments such as fully relevant, partially relevant, and not relevant.

Although a few studies have begun the investigation into criteria for fully relevant documents, the question remains, “What does partial relevance mean to users and what criteria do they use when labeling a document as partially relevant?” Many studies that utilize participants with real information needs allow the participants to use their own definitions of relevance (Park 1992, 1993; Barry 1993, 1994; Bruce 1994). According to results from these studies, participants generally indicate that items judged to be relevant met their information needs in some way. At the opposite end of the scale, the documents judged as non-relevant failed to meet the participant’s needs. The criteria of judgments that fall somewhere in between the two extremes of relevance and non-

relevance have yet to be identified. This study was designed to gather knowledge that will begin to fill that gap by investigating user criteria for partial relevance.

2.0 Previous Research

2.1 Explanations of Relevance

Before the question of partial relevance can be discussed in detail, relevance itself must be addressed. The definition for relevance has become the Holy Grail of information retrieval. This is reflected in the different interpretations of relevance in almost every article written about relevance since the Cranfield studies. As Cleverdon maintained in 1966, “The matter of relevance assessment is without a doubt, the most difficult intellectual problem - in fact one of the very few remaining problems - in the evaluation of information retrieval systems” (1966, p. 114). Although thirty years’ worth of research has taken place since this statement was published, it still holds true today.

In seeking a definition of relevance, researchers have dissected the problem in what amounts to two approaches: identifying different types of relevance and defining synonyms for relevance (more thorough discussions on relevance can be found in Saracevic, 1975, 1976; Schamber, Eisenberg, & Nilan, 1990; Schamber, 1994). Several researchers have attempted to define relevance by asserting the importance of different types of relevance.

Most of the various types of relevance proposed in the literature can be divided as Schamber (1994) suggests, into either system oriented or user oriented. Cooper (1971) uses the term logical relevance to describe a relevance decision that has little or nothing to do with the original user’s judgement. “[L]ogical relevance, alias ‘topical-appropriateness,’ which 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). Park (1994) calls this type of relevance topical relevance. Building on Cooper’s definition of logical relevance, Park 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). Logical and topical relevance both describe a relevance decision that has little or nothing to do with the original questioner’s judgement.

Cooper and Park are not alone in identifying system-oriented aspects of relevance. Both Swanson (1986) and Howard (1994) divide relevance into system oriented and user oriented parts, objective relevance and subjective relevance. They 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, 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” (p. 391-2). Howard 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 means that all items containing query terms could conceivably be objectively relevant, regardless of the user’s perception of how those items relate to his or her information need.

In contrast to objective relevance, Swanson and Howard address the relationship between the user and the items retrieved in the concept of subjective relevance. Unlike objective, logical or topical relevance, subjective relevance is totally user centered: “[W]hatever 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 it is the originator of the request that must make a value judgement on the items returned.

In addition to subjective relevance, researchers have also offered situational and psychological relevance, as 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.

Wilson also suggests another user-centered aspect of relevance, psychological relevance. In this definition, Wilson examines not only the moment the relevance judgement is made, but also effects that an item may have on the user’s behavior after the judgement 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, but his definition is more focused than Wilson’s, and in many ways similar to Swanson’s explanation of subjective relevance.

Harter 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).

Along with the various aspects of relevance mentioned above, many researchers have also defined relevance with an assortment of its synonyms. Tessier, et al. (1977) emphasizes user centered relevance with the use of the word *satisfaction*: “What must be emphasized is that satisfaction is ultimately a state experienced inside a user’s head. It is, therefore, a response that may be both intellectual and emotional” (p. 338-4). By stating that a relevance judgement is an internal process, this definition of satisfaction echoes Wilson’s definition of situational relevance.

Foskett (1972) adds another dimension to relevance, novelty, with his explanation of pertinence. His definition of relevance is similar to Swanson’s definition of objective relevance. Foskett states that “‘relevant’ should be taken to mean ‘belonging to the field/subject/universe of discourse delimited by the terms of the request’” (p. 77). Pertinence, on the other hand, “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).

Like Foskett, Cooper (1971, 1973) expands the concept of relevance and suggests the term “utility” as an antithesis to his definition of logical relevance. According to Cooper, utility is user centered and is “a catch-all concept involving not only topic-relatedness but also quality, novelty, importance, credibility, and many other things” (1973, p. 92). In this definition, Cooper not only attempts to capture user-centered relevance, he also offers what, in later studies, users identify as their criteria for relevance.

2.2 Criteria for Relevance

After examining the most prominent phrases and synonyms used by various researchers to describe the concept of relevance, it is obvious that there is by no means a consensus. It is as if the very elusiveness of the definition for relevance drives researchers to continue searching for something that the previous 40 years of searching have failed to reveal. Is it necessary to continue what may be a fruitless quest or can we approach the problem 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 all of the following studies, criteria were gathered directly from the users through think aloud techniques, interviews and questionnaires. Park (1992, 1993) suggests that user criteria can be grouped into three categories. Participants in this study were asked to discuss their information need, and to evaluate citations pertaining to their need. After transcribing the interviews, Park grouped the participant’s 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” (1993, p. 338) and include comparisons between the current research problem and research problems described in the citation.

Where Park’s study categorizes criteria based on their relationship to the searcher, the search, and the problem, Schamber (1991) groups criteria according to aspects concerning the information source itself. This study sought to examine 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, and were also asked to evaluate the information received from those sources. These interviews were transcribed and the participant’s evaluation criteria were examined. From this analysis, ten categories of criteria emerged (ordered by frequency): presentation quality, currency, reliability, verifiability, geographic proximity, specificity, dynamism, accessibility, accuracy, and clarity (1991).

Cool, et al. (1993) combined the approaches taken by Park (1993) and Schamber (1991) by grouping evaluation criteria both by the relationship they had to the searcher and by criteria concerning the information source itself. In this study the researchers performed two separate experiments and reported the results from each. The first experiment 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. The second experiment gathered evaluation criteria from scholars through interviews about the scholar’s information needs and the items they used to meet these needs. The preliminary results, from initial analysis of the data from both experiment, indicated that the criteria 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 judgement - these are modifiers of other facets), and oneself (relationship between person’s situation and other facets) (1993, p. 79).

Barry (1993,1994), like Cool, et al. (1993), found criteria related to both the user and to the information source. Participants in Barry’s 1993 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 lead them not to pursue the full text document. Following this process the participants were interviewed and asked to explain why they circled or crossed out the items. Analysis of the criteria mentioned in these interviews yielded 23 criteria which 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 (1994, p. 154).

Table 1 – Relevance criteria common to three studies

Barry & Schamber (1995)	Cool, et al (1993)	Park (1993)
depth/scope	deep/superficial	scope
accuracy	technicality and scientificness	quality of information
clarity	understandability	readability
currency	age (of document)	up-to-dateness
specificity	specificity	---

Regardless of the research question or categorization of the criteria, all of the studies mentioned found some similar criteria. Barry and Schamber (1995) compare the criteria found in their studies mentioned above and found five that overlapped (Table 1). All five of these criteria can also be found in Cool et al. (1993), and four are identified in

Park (1991). This repetition of criteria may support Barry's assumption "that there is a finite range of relevance criteria across users and situations" (1994, p. 157).

2.3 Degree of relevance

In addition to criteria for relevance, the degree of relevance judgments adds yet another dimension to the problem of relevance. Many studies of system performance and user relevance judgments ask users to evaluate the relevance of an item based on a predetermined scale or on a user-determined scale. These scales vary considerably as can be seen in the following examples: 3 point scale (Saracevic, 1969; Marcus et al. 1978; Janes, 1991b), 5 point scale (Thompson, 1973), 6 point scale (Smithson, 1994), 9 point scale (Caudra & Katter, 1967), 11 point scale (Rees & Schultz, 1967), and magnitude estimation (Eisenberg, 1986,1988; Bruce, 1994).

The problem with this variety of scales, besides the variety itself, is that most researchers do not justify their use of them. Smithson (1994) was one of the few of the studies cited above that explained why scaled judgments were incorporated: "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 scale: 6. very useful, 5. useful, 4. background interest, 3. cannot say, 2. of little use, 1. not useful" (p. 209). Although this may very well assist the user in making a relevance decision, there is nothing to support this assumption.

Many other studies not only failed to justify the use of scaled relevance evaluations, they also failed to utilize it. Several of the studies mentioned above collapsed the judgments by dividing them into two groups at or near the halfway point. In contrast, Saracevic (1969) favored the relevance group; collapsing a three-point scale by placing

the middle, “partially relevant” judgments with the “relevant” judgments. Saracevic’s method was later supported by (Eisenberg and Hu, 1987). In Eisenberg and Hu (1987), participants were asked to indicate on a “100mm” line where they would place the dividing point between relevant and non-relevant. The majority of the participants in the study indicated that the break was closer to the non-relevant end of the line than to the relevant end. 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” (p. 68).

Rees and Schultz (1967) and Janes (1993) both look at user behaviors associated with scaled judgments. Before conducting their study using an eleven 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 the participant’s use of scale points in Rees and Schultz (1967) to those in a similar study by in Caudra 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 judgement. 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 non-relevant, but lies somewhere in between.

2.3 Partial Relevance

The aspects of relevance discussed in the previous two sections, criteria and scale, are rarely compared in the same study and the question, “What do users find in partially relevant items that make them neither relevant nor non-relevant?” has yet to be examined fully. Janes (1993) suggests that the answer to this question may be related to the degree to which the item possesses various criteria:

Perhaps the process people go through is a two- (or multiple-) stage one:

1. Determine, very quickly, if the document is really good or really bad. If so, say so (and the data appears to show that they don't much care how really good or bad it is).
2. If not, then more time and effort must be taken to determine how much of it is good, whether or not it is from a trustworthy source, addresses the right issues, is in the right language, is available and accessible, etc. (p. 113)

Bookstein (1983) adds another aspect to this by suggesting that judgments of partial relevance could be either a reflection of the item's degree of relevance, as Janes suggests, or a reflection of the user's uncertainty in the item's relevance. Spink (1997) suggests 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). While the aspects of partial relevance identified by Janes, Bookstein, and Spink may begin to shed light on the perception of partial relevance, further research is needed before the implications surrounding a judgement of partial relevance can be understood.

3.0 Methodology

3.1 Data Collection

The most important questions about partial relevance focus on the user. They include, “Do users identify different criteria in relevant, partially relevant, and non-relevant items?” In order to begin this investigation the researcher recruited 12 social science graduate students attending the University of North Carolina at Chapel Hill (10 from the Department of Sociology and 2 from the Department of History) with real information needs. Nine of the participants were working on their dissertation, two on their thesis, and the last on a paper for publication.

They were recruited by word of mouth, flyers posted in social science departments and emails posted on social science department listservs. The advertised incentive was a free Dialog search and photocopies of all the articles deemed relevant. Initial contact included verifying that the possible participant was a graduate student and that he or she had an information need directly related to a current research question. In addition to this, an appointment was made for the initial search interview.

The initial search interview consisted of an in-depth reference interview lasting from 20-60 minutes. The interview was audio taped and referred to for clarification as needed. The participant filled out a reference interview questionnaire and then answered questions asked by the researcher to clarify information as needed (see Appendix A for the reference questionnaire). Questions posed to the participant attempted to gather information about: the 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. Immediately following the interview, an appointment was made for the participant to evaluate the retrieved document representations. The participants were asked to set aside two hours for this session. The second interview occurred from two to seven days following the initial interview.

The researcher then conducted the 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 and fields for article or book title, author, journal or publisher name, publication date, and language (see Appendix B for an example document representation). 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 evaluation.

In the second session, participants were asked to evaluate the document representations, highlight passages they considered relevant, mark through the passages they considered non-relevant, and judge the document representations as a whole to be: relevant, partially relevant or non-relevant (see Appendix C for evaluation instructions). After evaluating the 20 document representations, the participants were asked to explain why they chose to highlight and cross out the passages, why they chose to mark the

document representation as relevant, partially relevant, or not relevant, and how they would describe a typical relevant, partially relevant or not relevant document. Due to time constraints, three participants did not answer the final question. At the completion of this interview the participants were given photocopies of the document representations they 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. This second interview was audio taped, transcribed, and analyzed following qualitative and quantitative methods.

3.2 Content Analysis

The content of the interviews was analyzed with the intent of making “replicable and valid inferences” (Krippendorff, 1980, p. 21) about the reasons the participants gave for selecting passages and rating documents. The set of interviews was examined and the participants’ criteria for passage selection were compared to the criteria identified in Barry (1993, 1994), Schambler (1991), Cool et al. (1993), Park (1992, 1993) and Wang and White (1994). None of the criteria sets in this literature fully captured both the information discussed by the participants and the research questions posed by this study. Therefore, as suggested by Stempel (1981), a new set of codes for the participants’ criteria was developed.

The coding system used in this study was developed by following the 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. The researcher’s notes on all the reasons given were examined for similarities. Similar reasons were grouped by criteria.

Each criterion was then given a positive or negative value depending on whether the participant felt that the information was useful or not useful. The criteria identification was an iterative process done by the researcher with advice from several colleagues.

A comparison of interjudge agreement was used to test the reliability of the criteria codes to fully capture information expressed by the participants. The researcher 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 (1996), the coefficient of interjudge agreement of the three judges was found to be .72 and determined to have a 95% confidence limit. The minimum coefficient of interjudge agreement between any two judges was found to be .81 and determined to have a 95% confidence limit. These results were within acceptable limits so the researcher used the criteria codes for content analysis on the remaining interviews.

4.0 Results

4.1 Criteria

The content analysis revealed 31 criteria, discussed positively and negatively, used by the participants when selecting passages to highlight and determining the relevance of a document representation. Based on the focus of the individual criterion, they were grouped into six categories: participant, full text document, author, journal/publisher, abstract, and content (see Appendix D for a list of category, criteria, value, and participant usage). The four common criteria found in previous studies of relevance criteria (Barry and Schamber, 1995; Cool, et al., 1993; Park, 1993), accuracy/validity, currency, depth/scope, and understandability were also identified in this study.

4.1.1 Participant

Criteria in this category relate directly to the participants' feeling, goals or constraints.

Time – references to a whether or not the information would save or waste the participants' time. Example responses: "Um, so anything that I can get that gives sort of a broad overview of a history in a region is helpful - saves me a ton of time." "I have so much literature that I need to go through already on the minority groups that I am including that I have no need or time to include, kind of related stuff about a group that I can't include."

Threatening information – indications that the participant’s research might have been done previously. Example responses: “This looks very similar to what my dissertation is, I had a little panic and said, “oh boy, she’s already done that.” “And that’s exactly what I’m doing...it looks like the findings are the same as mine, hopefully they haven’t done everything I’m doing.” “Yeah, I’m nervous about this article because I was hoping that nothing had been done, so, but I really need to look at it.”

Not relevant to this project – information that was useful to another project the participant was working on but that may or may not have been useful to the current project. Example responses: “This actually is not as relevant to my dissertation but it is relevant to what I am going to be doing next.” “Oh, that’s for another paper I think I would be interested in writing about.”

4.1.2 Full text document

Criteria in this category relate directly to the full text document rather than the document representation that the participant is evaluating.

Audience – information indicating the audience for which the full text document was written. Example responses: “That seemed to me that it was directed pretty seriously at high school students.” “And, I don’t know - I guess they are speaking to their audience, people who may be like really super gung-ho about what sports can do. That’s why.”

Form – information indicating the full text document’s form or type. Example responses: “Um, 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.”

Novelty – indications that the participant has knowledge of or has read the full text document. Example responses: “And so, I think that I have a copy of this article already.” “OK, um, this particular article is one that I read quite extensively and, um, has become sort of a source material for what I’m doing.”

Possible content – information that leads the participant to guess or theorize on the content of the full text article. Example responses: “My guess is that this is much more focused on a particular ownership.” “So my assumption is that this article is about...”

Read the full text – indications of whether the full text document would be sought or not. Example responses: “So it looked like it would be something that would be worth looking at.” “And while that’s really tangentially related to what I’m interested in, it’s probably not enough that I would go seek out this article.”

4.1.3 Author

Criteria in this category relate directly to the author or authors of the document.

Discipline – references to the author’s area of research. Example responses: “And, low and behold, the author is somebody I’m familiar with from her work in doing this type of research.” “And, again, the author, who’s come up again, who’s doing work substantively very similar to my own.”

Institutional Affiliation – references to the author’s sponsor or employer. Example responses: “Well, first of all – [author affiliation: ILO] International Labor Organization is probably the key international organization that’s focusing on labor markets.” “I

noticed that the corporate source was Duke University and that actually made me really interested in it.”

Novelty – indication that the participant was familiar or unfamiliar with author.

Example responses: “Ok, yeah, in the preliminary research that I have done, in particular, this guy has come up...So, I would definitely go and look for this article based on that.”

“I mean it sounds important, I just don’t know the authors.”

Perceived Status – the participant’s perception of the author’s academic standing.

Example responses: “And because of the stature of these two editors I think that is probably going to be a very relevant article.” “I was kind of excited to see that it was also a prestigious author” ““Seymour Martin Lipset,’ a very important political scientist.”

4.1.4 Journal/Publisher

Criteria in this category relate directly to an the journal where the article appears or the publisher of a book.

Main focus – references to the journal’s typical content. Example responses:

“Annual Review of Sociology” is sort of like a summary of research in a broad field.”

“Um, I crossed out [the journal title], again, it’s just more interpersonal literature, um, and they aren’t really looking at organizational frames.”

Novelty – indication that the participant was familiar or unfamiliar with the journal or publisher. Example responses: “And I highlighted the journal because I didn’t know about it.” “Never heard of it. It’s from England, it’s probably not going to be relevant. Which, I mean, can completely deep-six what looks like a good article.”

Perceived quality – the participant’s perception of the journal or publisher’s rank or quality. Example responses: “It is an AJS which makes it one of the top two journals.” “And also, looking that it’s in ‘Sociological Perspectives’ which is sort of a second or third tiered journal in sociology.”

Recency – references to the date of publication. Example responses: “I mean it’s relevant ...when it was published [1996].” “And um, I highlighted the year, “1990,” it seems, it’s slightly old.”

4.1.5 Abstract

Criteria in this category relate directly to quality and usability but not the content of the abstract in the document representation.

Citable – indications that the abstract itself would be cited rather than citing the full text document. Example responses: “But I might still cite it just based off the abstract.” “And in fact, I mean, I could, actually, um, could practically cite this article just based on this.”

Informativeness/understandability – references to the abstract’s ability to represent information in the full text document. Example responses: “I tend to think this abstract is not really related to what the subject of the article is.” “It didn’t really give me enough information to make a real informed decision.”

4.1.6 Content

Criteria in this category relate to the content, not the structure, of the title, abstract, and descriptor fields of the document representations.

Accuracy/Validity – references to the quality of the research. Example responses: “Then I just crossed out ‘social tolerance and social feasibility’ because that seemed to me to be less a kind of factual element and more a value judgment on the bases of the authors.” “Um, they had a ‘sample of 166’ people and then sort of looking at the methodology and saying, OK this looked like something that was responsibly done with a large enough sample.”

Background – references to background or context information. Example responses: “Um, this one I highlighted ... because I am trying to contextualize the women teachers’ experiences that I’m interested in.” “But I gave it only a P for partial because I thought it was background information relevant to the overall process but not something that I am focusing on.”

Citations – indications that the full text document references notable sources. Example responses: “Because they’re using – referring to Bordingham and White...which is a really important article in my field.” “‘Sampson and Laub’s social control theory’ got to take stock of that for sure...Sampson and Laub are probably the most recent people - most recent famous people to get their names in a big article and call it their own.”

Contrast – references to information that contrasts with their own or other research. Example responses: “It would be useful to look at even if I just set it up as a strawman... this is completely not at all what I have found.” “Um, I did highlight though,

“Employment and equality on the frontier,” because that might make an interesting comparative.”

Depth/Scope – references to the breadth or specificity of information covered.

Example responses: “And the title implies that it is going to be broad in scope.” “That seemed to me to be really particular and specific, um, and not necessarily relevant to what I’m doing.”

Domain – references to the field or area of study. Example responses: “I think that’s part of what I don’t need - I don’t need psychological stuff here.” “I would definitely read this article because it’s very focused on my general topic but also because it’s from a different, ah, profession.”

Information to find other information – references to information that could lead to additional information. Example responses: “So, it looks like it might be a good place to look to get data on [my topic].” “[They] were terms that I could see following up on. Um, I don’t think I’ve ever tried [searching for] stigmatization for instance.”

Information that helps stimulate/clarify – references to ideas or methods that help to stimulate participant’s thinking. Example responses: “It would help me to sort of formulate my own ideas and thoughts.” “I’m curious to see how they justify it because I’m going to have to do it too.” “[It will] help me place the features I’m talking about within a framework of understanding.”

Novelty – references to known or unknown information. Example responses: “And they also added a new aspect which was the former colonial statutes.” “Now this more heavily highlighted stuff, that actually is nothing new there.”

Rarity – references to the accessibility of the information itself. This does not include document accessibility. Example responses: “I haven’t seen too many references about that so, that’s why I marked that.” “I’ve seen so little on the career trajectory stuff and sort of career development that has actually talked to anybody.” “Oh, um, ‘Picasso’ is way over done.”

Subject Matter – references to subject, topic, method or result that do not include any of the above criteria. Example responses: “But, so, tripartism, is an issue I’m interested in.” “I’m interested in how kids who are in school work, but it’s very different than an actual work-study program...Um, it’s just a very, um, different concept.” “So that becomes very important because that’s not only – I mean that’s sort of the data I’m using.”

4.2 Criteria Usage

As discussed in the methodology section, participants were asked three different questions: “Why did you decide to highlight or underline a passage?,” “Why did you mark the document representation overall as relevant, partially relevant, or not relevant?,” and “How would you describe typical relevant, partially relevant, and not relevant documents?” The analysis of the responses included comparing criteria mentioned in the answer to each question to the overall judgment of document relevance and to each other.

Table 2 – Participants’ overall relevance judgments on document representations

Participant	Number of Document Representations						
	Not Relevant		Partially Relevant		Relevant		Total evaluated
1	1	5.00%	10	50.00%	9	45.00%	20
2	0	0.00%	4	23.53%	13	76.47%	17*
3	0	0.00%	5	25.00%	15	75.00%	20
4	1	5.00%	12	60.00%	7	35.00%	20
5	5	25.00%	7	35.00%	8	40.00%	20
6	5	25.00%	8	40.00%	7	35.00%	20
7	2	10.00%	2	10.00%	16	80.00%	20
8	5	26.32%	4	21.05%	10	52.63%	19*
9	3	15.00%	10	50.00%	7	35.00%	20
10	11	55.00%	7	35.00%	2	10.00%	20
11	3	15.00%	10	50.00%	7	35.00%	20
12	2	10.00%	8	40.00%	10	50.00%	20
Total	38	16.10%	87	36.86%	111	47.03%	236

* does not equal 20 because the participant had seen the document or document representation before and did not evaluate it.

The breakdown of documents representations by participant and document relevance shows that the majority of the document representations, 47%, were found to be relevant, while only 16% were found to be not relevant (Table 2). Only 10 participants found not relevant document representations. This unequal balance of document representations resulted in very little data about not relevant documents and the criteria used to identify and describe them. In order to present a balanced analysis of the use of criteria in the three relevance judgements, the number of criteria in a relevance category was normalized by the total number of documents in the relevance category.

4.2.1 Criteria for highlighting or underlining a passage

The purpose of asking the participants to highlight or underline portions of the document representations was to determine the aspects of document representations that

are important in deciding its relevance. Analysis of this was accomplished by comparing categories of the criteria to the overall judgement of the document representation.

Table 3 – Normalized use of passage criteria across relevance judgments

Criteria Category	Not Relevant	Partially Relevant	Relevant
Abstract		5.7	3.6
Author	15.8	23.0	40.5
Content	215.8	305.7	408.1
Full Document	10.5	16.1	37.8
Journal/Publisher	2.6	6.9	24.3
Participant		1.1	9.0
Total	244.7	358.6	523.4

The number of criteria mentioned when discussing the passages marked was significantly higher in relevant than not relevant document representations (Table 3). The participants' habits for evaluating documents on their own and possibly only looking for relevant passages may have contributed to this trend. This could also be due to other factors including the participants reading relevant documents more closely or spending more time on relevant documents.

Criteria concerning the content were mentioned more than the sum of all other criteria regardless of relevance judgement (Table 3). This would indicate that while there are many more aspects to relevance than content, the content of a document representation is more important and/or receives more attention than any other aspect.

Criteria in the author, content, full document, journal/publisher categories were identified for all relevance judgments in almost identical proportions. This indicates that the degree of relevance may not depend on the category of passages but in the number of passages in those categories.

Table 4 – Normalized use of passage criteria and value across relevance judgments

Criteria Category and Value	Not Relevant	Partially Relevant	Relevant
Abstract +		1.1	2.7
Abstract -		4.6	0.9
Author +		16.1	29.7
Author -	15.8	6.9	10.8
Content +	65.8	181.6	357.7
Content -	152.6	134.5	62.2
Full Document +	2.6	5.7	32.4
Full Document -	7.9	10.3	5.4
Journal or Publisher +	2.6	2.3	18.0
Journal or Publisher -		4.6	6.3
Participant +		1.1	7.2
Participant -			1.8
Total	247.4	369.0	535.1

Other differences between relevance categories can be seen when the categories are separated by the value, either positive or negative, that the participant indicated when describing the passage (Table 4). Every category of criteria was mentioned more positively than negatively in relevant document representations, the criteria values were mixed in partially relevant document representations, and the criteria values were generally negative in not relevant documents representations. This trend can be seen most clearly in the content category where the relevance of a document representation could be predicted by the number of positive or negative content criteria the document representation contains.

4.2.2 Criteria for judging the overall relevance of a document representation

In an attempt to identify differences between passage selection criteria and document relevance criteria, the participants were asked to discuss their overall relevance judgments of each document representation.

Table 5 –Percentage of criteria categories in passage selection and overall relevance judgments

Criteria Category	Passage	Overall
Abstract	0.9%	5.5%
Author	7.0%	3.8%
Content	81.7%	74.9%
Full Document	5.9%	12.0%
Journal/Publisher	3.4%	2.5%
Participant	1.1%	1.3%

Content analysis determined that overall the criteria and criteria categories mentioned by the participants were the same for both passage selection and document relevance. Additionally the similarity of the category usage was statistically significant (Table 5). This indicates that in general the categories are used in the same proportions for both passage selection and relevance judgments.

Table 6 – Normalized use of criteria in overall relevance evaluations

Criteria Categories	Not Relevant	Partially Relevant	Relevant
Abstract	2.6	16.1	9.9
Author	2.6	1.1	14.4
Content	86.8	162.1	145.0
Full Document	15.8	44.8	34.2
Journal or Publisher	2.6	2.3	8.1
Participant	10.5	1.1	2.7
Total	121.1	227.6	214.4

Not surprisingly, some of the trends identified in the criteria for passage selection were also found in relevance judgement criteria (Table 6). The number of criteria mentioned when discussing the overall relevance was greater in relevant than not relevant document representations and the criteria concerning the content were mentioned more

than the sum of all other criteria regardless of relevance judgment. Unlike the criteria for passage selection, criteria for relevance were identified for all relevance judgments.

Table 7 – Normalized use of criteria and value in overall relevance evaluations

Category and Value	Not Relevant	Partially Relevant	Relevant
Abstract +	2.6	5.7	2.7
Abstract -		10.3	7.2
Author +		1.1	11.7
Author -	2.6		2.7
Content +	13.2	85.1	135.1
Content -	81.6	92.0	14.4
Full Document +	2.6	13.8	27.9
Full Document -	5.3	11.5	0.9
Journal or Publisher +	2.6		8.1
Journal or Publisher -		2.3	
Participant +			
Participant -	10.5	1.1	0.9
Total	121.1	223.0	211.7

Other differences between criteria for passage selection and relevance judgement criteria can be seen when the categories are separated by the value, either positive or negative, that the participant used when describing the passage (Table 7). In passage selection every category of criteria was mentioned more positively than negatively in relevant document representations; in relevance determination, the abstract was mentioned more negatively than positively. The criteria used for partially relevant document representations were also different in that the criteria were mentioned, in general, more negatively than positively. This indicates that even though a partially relevant document representation has more passages with positive than negative criteria, participants felt that the document representation as a whole contains more negative than positive aspects.

4.2.3 Combined criteria for passage selection and judging the overall relevance of a document representation

The findings discussed in the previous sections were generalized across participants and document representations. It is interesting to note, however, that participants, when evaluating a single document, did not always use the same criteria for determining relevance that they used for passage selection. More than twice as many criteria were mentioned in describing passage selection than in describing relevance judgments. Out of the 236 document representations evaluated, 129 document representations had criteria used to describe the relevance judgement that were not used when describing the selection of passages in the document. The criteria mentioned most frequently for relevance judgments, but not passage selection, were depth/scope and read the full text. The discrepancy between these indicates that asking participants to evaluate passages only, and not the document representation as a whole, does not always give the full picture as to why a particular relevance judgement was selected. The following analysis looks at the document representation as a whole by combining the criteria for passage selection and the criteria for judging the overall relevance of a document representation.

The document representations were examined to determine if there were criteria categories that tend to occur together. In other words, whether the presence of one category could predict the presence of another category. All categories, except the positive participant category, occurred with positive content category at least 82% of the time. Other than the content categories, the categories did not occur together more than 36% of the time (positive abstract criteria and positive document criteria). This indicates that the criteria categories, other than content, are distinct, not dependent on each other, and cannot be used to predict the presence of another category.

Table 8 – Distribution of positive and negative criteria in document representations

Document relevance	Total	only negative criteria		only positive criteria		both negative and positive criteria	
Not Relevant	38	21	55.26%	0	0.00%	17	44.74%
Partially Relevant	87	4	4.60%	6	6.90%	77	88.51%
Relevant	111	0	0.00%	50	45.05%	61	54.95%
Total	236	25	10.59%	56	23.73%	155	65.68%

The document representations were also examined to see how the presence of positive and negative criteria differed across the overall relevance judgments (Table 8). As would be expected, a great majority of partially relevant document representations contained both positive and negative criteria. The occurrence of positive and negative criteria in relevant and not relevant document representations was not as straightforward: 55% of the not relevant document representations contained only negative criteria but only 45% of the relevant document representations only contained positive criteria. This means that almost half of the document representations that were judged either relevant or not relevant contained aspects that contradicted the overall judgement. Only 36.9% of all the document representations were judged by the participants to be partially relevant

yet 65.7% of all the document representations contained both positive and negative information. Assumptions cannot be made that a document's relevance judgement directly reflects the value of all the information contained in the document. Neither can assumptions be made that the criteria or even individual occurrences of the criteria have equal weight in the final decision.

4.2.4 Criteria used in describing “typical” relevant, partially relevant, and not relevant documents

After discussing the decisions they had made on each document representation, the participants were asked to describe a typical relevant, partially relevant and not relevant document. The purpose of this was to gather the participants' perceptions of the criteria they used to evaluate document representations and then compare that to the criteria they actually used when determining relevance. The two-hour time constraint on the interview prevented three participants from answering the final question.

Table 9 – Occurrence of criteria used in describing a “typical” document by all participants

Category-Value	Not Relevant	Partially Relevant	Relevant
Author +			
Author -			1
Content +	1	3	20
Content -	16	7	1
Full Document +			2
Full Document -	4	3	
Journal/Publisher +			2
Journal/Publisher -	2		
Participant +	1	1	1
Participant -	1	1	
Total	25	15	27

The results are similar to those found in the relevance judgement criteria. The participants focused on criteria in the content category, and they found relevant documents to have mostly positive criteria, and partially relevant and not relevant document to contain more negative than positive criteria (Table 9). This again differs with the criteria values used in passage selection and reinforces the idea that the participants' perceptions of documents as a whole do not necessarily reflect what the document actually contains.

5.0 Discussion

This study was designed to investigate the use of relevance criteria in relevant, not relevant and, specifically, partially relevant documents. Previous studies have suggested some possible differences of criteria between partially relevant and relevant documents but few studies have been conducted specifically to test these theories. This investigation found evidence to support some but not all of these theories.

Novelty, uncertainty, and a lesser degree of relevant aspects 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 the judgment of partially relevant reflects 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. Participants 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. The participants, in their discussions of passage selection, indicated that partially

relevant documents did not contain as many criteria and the value of the information in partially relevant documents was not as great as that in the relevant documents. This was reinforced by the participant's evaluation of the document representation as a whole where, in fact, they tended to perceive that a partially relevant document had even less value than they had indicated when selecting passages in that document.

Additionally, the participants often used criteria when evaluating document representation as a whole that were not mentioned during passage selection. This indicates that both the evaluation of passages and the document representation as a whole are needed to fully understand why a particular relevance judgment was made.

It is also important to note that fewer than 50% of the documents judged relevant or not relevant were totally relevant or totally not relevant; most contained both negative and positive information. This finding can help to explain some of the current problems with information retrieval systems that use complete documents in relevance feedback. Performing feedback using a document that was judged to be relevant, but in fact contains some not relevant information, will give higher weights to the not relevant information in the document, along with the relevant information, reducing the system's effectiveness. One possible solution would be to allow the users of feedback retrieval systems to make relevance decisions on passages within documents rather than on the document as a whole. Further research on passage feedback retrieval systems is needed before this can be verified.

6.0 Limitations

This study was designed as a beginning to the investigation into the use of relevance criteria across relevance judgments and is limited by the small number of participants. Results from this study can be generalized only to similar populations, information needs, and material types (i.e. text).

7.0 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, Ohio, (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, Illinois, (pp. 103-111). Medford, NJ: Information Today, Inc.

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.

Cleverdon, C.W., Mills, J., & Keen, E.M. (1966) Factors in determining the importance of indexing systems. (Volume 1). Cranfield: College of Aeronautics.

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.

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.

- Eisenberg, M. (1986). Magnitude estimation and the measurement of relevance.
Unpublished doctoral dissertation, Syracuse University, Syracuse, N.Y.
- 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, Massachusetts, (pp. 66-69). Medford, NJ: Learned Information.
- Flick, U. (1998). An introduction to qualitative research. Thousand Oaks, CA: Sage.
- Gifford, C., & Baumanis, G. J. (1969). On understanding user choices: textual correlates of relevance judgments. American Documentation, 20(1), 21-26.
- 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, Ohio, (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., & Benefeld, A. R. (1978). Catalog information and text as indicators of relevance. Journal of the American Society for Information Science, 29(1), 15-30.

Mizzaro, Stefano. (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.

Saracevic, T. (1969). Comparative effects of titles, abstracts and full texts on relevance

judgments, Proceedings of the American Society for Information Science, San
Francisco, California, (pp. 293-299). Westport, Conn.: 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., 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. and Greisdorf, H. (1997). Users' Partial Relevance Judgments During Online Searching. Online and CDROM Review, 21, (5) 271-279.

- 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.
- Swanson, D. R. (1986). Subjective versus objective relevance in bibliographic retrieval systems. The Library Quarterly, 56, 389-398.
- 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.L., & White M.D. (1995). Document use during a research-project: a longitudinal study. Proceedings of the American Society for Information Science, Columbus, Ohio, (pp. 181-188). Medford, NJ: Learned Information, Inc.

Wilson, P. (1973). Situational relevance. Information Storage and Retrieval, 9(8), 457-471.

Zull, C.G. & Baumanis, G.J. (1967). Textual Correlates of Relevance: Preliminary Investigations. Cleveland, OH: Case Western Reserve University, School of Library Science, Center for Documentation and Communication Research.

8.0 Appendices

8.1 Appendix A – Reference Interview 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 bellow
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)

Name _____

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.

13. 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? _____

8.2 Appendix B – Sample Document Representation

12/3,AB/6

DIALOG(R)File 7:Social SciSearch(R)

(c) 1998 Inst for Sci Info. All rts. reserv.

03186239 GENUINE ARTICLE#: ZN630 NUMBER OF REFERENCES: 92

TITLE: Adolescent deviant behavior: Multiple contingency table analyses of overlap between behaviors

AUTHOR(S): Benda BB (REPRINT); Corwyn RF

CORPORATE SOURCE: UNIV ARKANSAS,SCH SOCIAL WORK/LITTLE ROCK//AR/72204

(REPRINT); UNIV ARKANSAS, COLL EDUC, CTR RES TEACHING & LEARNING/LITTLE ROCK//AR/72204

JOURNAL: JOURNAL OF SOCIAL SERVICE RESEARCH, 1998, V24, N1-2, P29-59

PUBLISHER: HAWORTH PRESS INC, 10 ALICE ST, BINGHAMTON, NY 13904-1580 ISSN: 0148-8376

LANGUAGE: English DOCUMENT TYPE: Article

ABSTRACT: This was a study of 1,093 adolescents from six public high schools designed to test a deviance syndrome perspective by examining how much overlap there is between certain deviant behaviors, and by investigating whether the same theoretical elements account for variance in those behaviors. A log-linear analysis of suicide attempts, alcohol use, use of marijuana, and number of sexual partners indicated a significant interaction (or overlap) between the latter three forms of deviant behavior. However, an examination of the configuration of data according to deviant behaviors indicated that the actual numerical overlap was less impressive than the statistical significance or syndrome argument would imply. The present study also provided logistic and OLS regression analyses of many elements of social control theory. Those analyses revealed that few of the theoretical factors were consistent predictors of the various forms of problem behavior. Social work implications of the findings were discussed.

8.3 Appendix C – Evaluation Instructions

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

1. As you are reading a document, highlight any portion of it that is relevant to your research and underline any portion of it that is not relevant to your research.
2. 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.
3. Mark the letter corresponding to your overall judgement in the margin next to document representation:

R = relevant

P = partially relevant

N = not relevant

8.4 Appendix D – Category, Criteria, Value and Participant Usage

Category	Criteria and Value	1	2	3	4	5	6	7	8	9	10	11	12
Abstract	Citable +						2			8			
Abstract	Informativeness +	1	1				1						
Abstract	Informativeness -	7	7	3			1			1		2	1
Author	Author Novelty +	6			2		8	4				1	6
Author	Author Novelty -	1	4	1	1		2	2	3	4	1	2	1
Author	Discipline +		3				2	8		2			
Author	Discipline -							1		1	3		
Author	Institutional Affiliation +	1						2					
Author	Perceived Status +	5	6				3			2			
Author	Perceived Status -		1										
Content	Accuracy-Validity +		5	1			3	1		11			
Content	Accuracy-Validity -	2	1	2	1	4	1			4			
Content	Background +	1	5	2								12	1
Content	Background -											1	
Content	Content Novelty +	1	3		6	1	2	1		2			2
Content	Content Novelty -		2		1		1	4		1	1	3	4
Content	Contrast +		3		4	2	3	1		2		2	2
Content	Corroboration +	1	2	10	4	4	1	1	1	5		4	
Content	Depth-Scope +	15	5	15	6	1		2	3	2	2	7	7
Content	Depth-Scope -	10	4	7	11	1	4	7	5	3		17	7
Content	Domain +		1				1	1	1	2			
Content	Domain -		1	2	2	4	4	2	3	6			2
Content	Guess about content +	4	1	7	3	1	2	2		3	3	4	2
Content	Guess about content -	3	2	3	1		2	2				4	
Content	Has known citation +				1				3	1		1	
Content	Info. to find other information +			6			8	28				2	1
Content	Info. to find other information -						1					1	
Content	Not Relevant to this project +					1							
Content	Not Relevant to this project -			1		1							1
Content	Rarity +	4	4	1			2	3		1		1	
Content	Rarity -			1									1
Content	Subject Matter +	53	38	47	26	14	36	58	69	52	20	40	45
Content	Subject Matter -	7	6	9	7	6	24	7	13	40	35	15	41
Document	Audience -								1			2	3
Document	Document Novelty +		6					1			3		3
Document	Document Novelty -	1				1		4	1				
Document	Form +		1	1			1	6				1	
Document	Form -	1	2					1					
Document	Read the full text +	2	17	1	4			7	3	7	1	5	1
Document	Read the full text -		1				1			8			
Document	Recency +		9				2	1		2			
Document	Recency -		1			3							
Journal/Publisher	Accessibility +		2				1						
Journal/Publisher	Accessibility -		1										
Journal/Publisher	Journal Novelty +			1									
Journal/Publisher	Journal Novelty -		2							2			
Journal/Publisher	Main focus +		2	2			1	4	1	1			
Journal/Publisher	Main focus -		1				1			4			

Category	Criteria and Value	1	2	3	4	5	6	7	8	9	10	11	12
Journal/Publisher	Perceived quality +		12		1		3			5			
Journal/Publisher	Perceived quality -		1					1			1		
Participant	Feels threatened +		1		2	1		1					1
Participant	Feels threatened -		1										
Participant	Time constraints +			2			1						
Participant	Time constraints -	1	1						1	1		2	1