Relevance Criterion Choices in Relation to Search Progress

By

Art Taylor

Marie Radford

_____

Xiangmin Zhang

_____

William Amadio

_____
Outside Member

Tefko Saracevic

_____
Chair

Craig Scott

_____
Director

A Dissertation Proposal
submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in
Communication, Information, and Library Studies
written under the direction of
Professor Tefko Saracevic

Art Taylor
25 Grandin Dr.
Flemington, NJ 08822

Introduction

Relevance is a foundational concept for the study of information retrieval (IR) systems. Early research in IR used a dichotomous concept of relevance (the document was relevant or not), assumed a static relevance judgment decision and greatly diminished or eliminated the role of the user. More recent research, however, has taken a cognitive, user-centered view of the relevance judgment process as both dynamic (changing over time) and multidimensional (varying among users). As part of this relevance judgment process, research has shown a user considers various criteria beyond topicality in making their relevance judgment. That these criteria are related to relevance judgments is clear, but few studies have examined how these relevance criteria may change as the user progresses through the information search process (ISP). A dynamic relevance judgment process suggests a dynamic cognitive state, with user relevance judgments changing over time as cognitive state changes. As these relevance judgments change, it is likely that the criteria used to make those judgments will also change. The study proposed here will attempt to detect such changes by examining user relevance judgments and criteria choices over time, as the user progresses through the ISP.

Statement and Significance of Problem

A user with an information need must close an information gap (Dervin, 1983), a process which may involve multiple search sessions and retrieval of documents or document representations. As documents or document representations are retrieved and examined, users' interaction with these texts changes their cognitive state. As users retrieve documents, they make relevance judgments about the documents based on various criteria. As the users' cognitive state changes, the criteria which are important to their relevance judgments may also change. In this case, topicality would still be a required criteria, but other criteria may have increased importance as the users' subject area knowledge (cognitive state) changes. Current search engines recognize few criteria beyond topic, and provide no

facility to adjust to changes in the cognitive state of users. Previous information science research has provided little guidance in what relevance criteria are important to users, and when in the search process those criteria are important. Identifying associations between relevance criteria choices, relevance judgments, and search stage would provide insights into changes in the users' cognitive state. Findings of associations would confirm and extend previous findings (Vakkari, 2000; Taylor, Cool, Belkin, & Amadio, 2006; Wang & White, 1999) and would inform the design of search systems. Improved search system design could extend basic topical search queries with additional criteria, and could adapt IR processing to the users' cognitive state changes as they progress through the ISP.

Variables and Research Questions

The proposed study will examine the relationship between relevance judgments, the criteria used to make those judgments. and progress through the information search process (ISP). Subjects will be presented with a research problem which will require them to gather information. As the subjects gather information, information will be captured on the relevance judgments of the subjects, the criteria subjects use to make those judgments, and the subject's progress through the ISP. The variables to be examined in this study are as follows:

- the *stage in the search process* which is operationalized as the subjects' selection of search stage from a predetermined list of search stage descriptions as detailed below;

- the *stage in task completion* which is operationalized as that point in time that relates to a specific deliverable that the subject must produce at that point in time;

- the subjects' *relevance criteria* choice which is operationalized as the subjects' choice of criteria which were critical in making their relevance judgment. Criteria will be chosen from a list of predetermined criteria presented to each subject as they evaluate a document. Subjects will also assign a weight indicating the importance of their criterion choice at that point in the search process;

- the subjects' *relevance judgment* which is operationalized as their judgment (relevant, not relevant, or partially relevant/unsure about relevance) on whether the document will be useful in solving their information problem.

Using these variables, the following research questions will be examined in this study.

1. Do users have a preference for specific criteria choices in relation to certain relevance judgments?

2. Do users have a preference for a specific criteria choice in relation to *stage in task completion* as indicated by the criterion selection and the corresponding weight assigned by the user?

3. Do users have a preference for specific criteria in making their relevance judgment in relation to a user-identified *stage in the search process* as indicated by the criterion selection and the corresponding weight assigned by the user ?

4. Do users have a preference for groups of relevance criteria in relation to a *user-identified stage in the information search process* as indicated by the criteria selections and the corresponding weights assigned by the user? Are there groupings of relevance criteria choices which relate to search stage?

Research Model

This research approaches the information seeking process of human information behavior from a problem-solving, user-centered, cognitive perspective. The user has a specific information need, and must gather information to fill that need. Time, context and situation must be considered, with work task viewed as part of context and situation. Several models inform this approach.

Background

Dervin and Nilan (1986) viewed the process of information seeking as one sensitive to context and situation. Users are active participants in this process, not passive receptors of information.

Emphasis is placed on the user both before and after information system use and the potential for change in the user's cognitive model state during information use is acknowledged. Internal (cognitive) and external factors affect the user over time. Dervin (1983) recognizes the impact of time within a holistic view where "all information is subjective" and fixed in a "time-space" frame, thus task and situation in relation to time become part of what constitutes information for a subject (ibid, p. 5). Information is constructed within a time-space framework and is not fixed or constant, but is instead malleable and changing over time.

A similar perspective is offered by Newell and Simon (1972) who regard the individual as an *information processing system.* Problem solving (information seeking) takes place in a problem space, is goal directed and continues through a series of knowledge states until a desired knowledge state is reached. Notably their model identifies a number of memory systems which are used by the subject in their pursuit of a desirable knowledge state. Langley and Rogers (2005) extend the problem space hypothesis to consider problem solving as a cyclic activity where the subject is evaluating and reacting to objects in their environment. Context and situation are not addressed directly by this model. Time however is represented indirectly through the cyclic nature of the model. Over time there are numerous cycles of problem solving during which the knowledge state (cognitive model) changes. Knowledge is not fixed in time, but varies constantly. Within this model, a subject attempting to fill an information need would interact with an IR system, and interact and react to the objects retrieved by the IR system. These objects would be the texts and document representations retrieved by the system. These interactions may bring the subject closer to filling their information need, or they may require the subject to backtrack and revisit a previous search path. In each case, over each cycle, the subject's knowledge state potentially changes.

Belkin (1982) examines problem solving within the context of information needs and IR systems. His work treats the undesirable knowledge state of Newell and Simon's problem solving

4

model as a state in which the subject has an anomaly in their state of knowledge, and that is their 'problematic situation.' The information search process is a series of anomalous states of knowledge (ASK) as the subject pursues that desirable knowledge state in which their problematic situation is resolved or has reached an acceptable conclusion. The subject's pursuit of a desirable knowledge state involves a series of search episodes, each of which involve interactions between the user and the IR system (Belkin et al, 1995). These interactions involve scanning text, reading abstracts or other forms of document representations, or potentially reading the entire document (Bates, 1989). As users proceed through the search process, their knowledge state changes thus changing their anomalous state of knowledge and their corresponding information needs. The subject's cognitive state exists in reference to time expressed as search episodes. Over time (a series of search episodes) the subject's cognitive state changes as the subject's ASK changes.

Model

As an individual seeks to solve an information need, they progress through a information search process, and in the process (over time) they search for documents, retrieve documents, and read the documents or representations of the documents found. Reading documents and absorbing the material in the documents leads to changes in the individual's subject area knowledge, and thus changes the current state of their cognitive model as it relates to their subject area knowledge. At this point they are in a new *cognitive state*. As individuals make relevance judgments, they reference their new cognitive state. The criteria individuals use to make their relevance judgment represents this cognitive state as it relates to their information problem at that fixed point in time. Individuals repeat this process of conducting a search episode as often as is necessary to either complete their search task and satisfy their information need, or arrive at a point where they are satisfied they have gathered all available information. Each search episode involves reading documents or document representations, absorbing

information, and making relevance judgments,

Chart 1 presents this model using a knowledge state diagram to identify the cognitive model states specific to a single information search episode. This diagram shows the retrieval and evaluation of documents within the information search process. Subjects begin the search task in state S1. This state encompasses the user's internal cognitive model prior to the start of the information search process. After formulating the search in operation O1, the user arrives at state S2. It is possible that in this state, the user's cognitive model has not changed as a result of formulating the search, so state S1 and S2 could be the same. It is also possible that the subject may have had a revelation as part of formulating the search query and thus S2 represents an altered cognitive state different from S1. In O2, the search is executed, results are returned and the subject evaluates the results of the search either by reading or skimming the documents or the document representations. This operation will most likely change the subject's cognitive state. It is reasonable to expect that the subject will learn what information is available and the nature of that information (in the form of document characteristics such as depth, breadth, scope). As a result of this review, the user's knowledge of the subject area will most likely change. They are now in state S3.
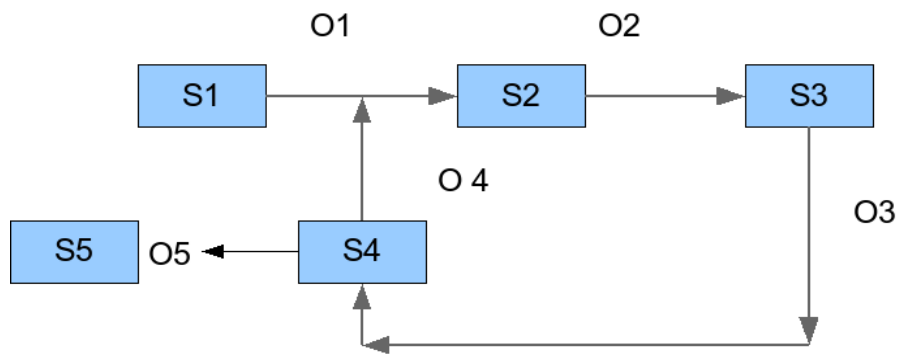
S1 – initial, pre-search
S2 – search executed and documents available
S3 – examined and read documents / document representation
S4 – documents reviewed and relevance judged
S5 – satisfactory conclusion of search process

O1 – formulate initial search and execute
O2 – evaluate document/ document representation; absorb information
O3 – judge relevance using various criteria
O4 – reformulate search
O5 – complete search

Chart 1:State Diagram of the Information Search Process

Once the documents have been read and evaluated in state S3, a relevance judgment is made

using various criteria. Following this relevance judgment, operation O3, the individual is in state S4. At

this point the individual evaluates the information they have absorbed and the documents currently

selected and judged relevant, and makes a decision whether or not to continue searching for

information. If they choose to continue, they reformulate the search in operation O4 and continue the

process of selecting documents, absorbing information and judging relevance. If they are satisfied with

the information they have absorbed and the documents gathered, they arrive at state S5 where an

adequate number of documents has been gathered and the information search process session is

complete.

Chart 2 shows the subset of operations from the model in Chart 1 that comprise the document

evaluation and relevance judgment process. As this diagram shows, subjects use the document

representation in tandem with specific criteria to make their relevance judgment. These criteria are

crucial, providing the lens through which they evaluate the document. This model represents as single

iteration within a larger information gathering process, which will most likely require multiple

iterations to complete. These multiple iterations themselves exist within a work task, with multiple

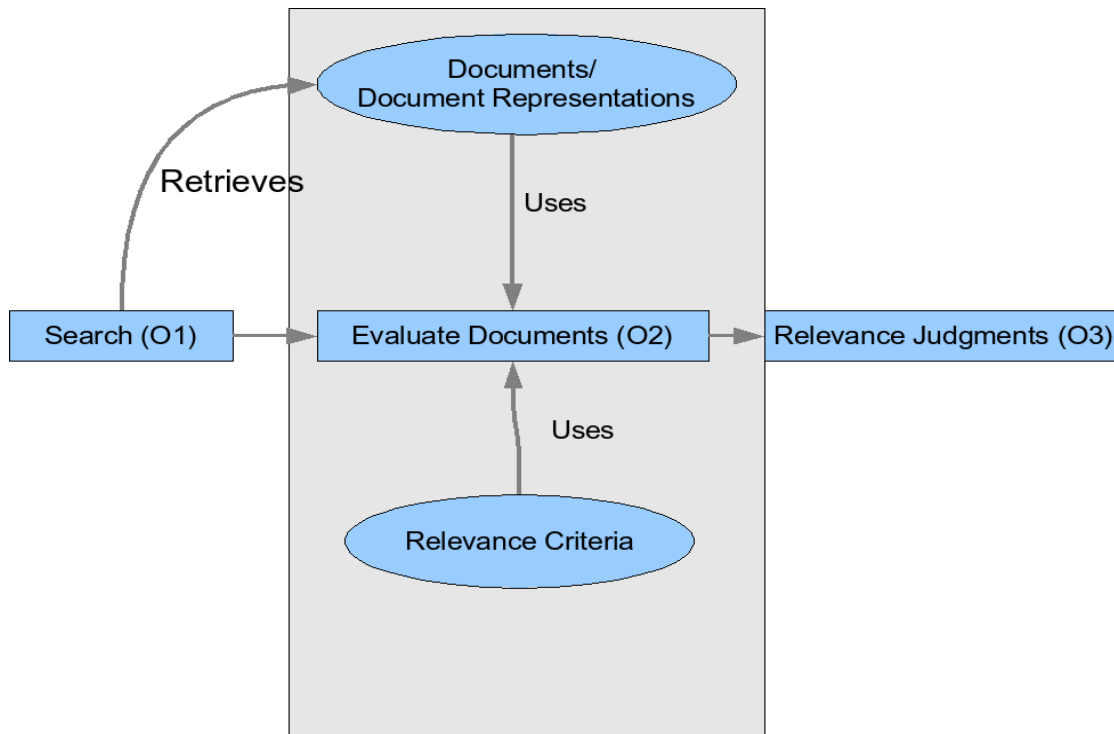iterations of search episodes as shown in Chart 2 taking place as part of this work task.

Chart 2: Search Episode within Information Search Process

Implications for this Research

As this model indicates, the user's cognitive model state is changing continuously as they search

for information. These changes in their cognitive model are manifested in the relevance judgment and

the criteria used to make that relevance judgment (state S3 operation O3 in Chart 1). Since the user's

cognitive model state is changing as they repeat the search/evaluate/judge process, it is reasonable to

expect that criteria that are a manifestation of that cognitive state will change. While criteria such as

topic should remain constant over the duration of the information search, other criteria should be more

likely to deviate from the original set of criteria. The goal of this research is to examine these variables.

How much these criteria would vary, and how often would depend on the *topic* and the *work task*, and the individual's *original cognitive model* as it relates to the subject area (their subject area knowledge). These factors represent intervening variables which were held constant in these studies by assigning similar work tasks and using subject pools of with similar knowledge, background, and search skills.

Implications for this Research

As this model indicates, the user's cognitive model state is changing continuously as they search for information. These changes in their cognitive model are manifested in the relevance judgment and the criteria used to make that relevance judgment (state S3 operation O3 in Chart 1). Since the user's cognitive model state is changing as they repeat the search/evaluate/judge process, it is reasonable to expect that criteria that are a manifestation of that cognitive state will change. While criteria such as topic should remain constant over the duration of the information search, other criteria should be more likely to deviate from the original set of criteria. The goal of this research is to examine these variables. How much these criteria would vary, and how often would depend on the *topic* and the *work task*, and the individual's *original cognitive model* as it relates to the subject area (their subject area knowledge). These factors represent intervening variables that must be addressed by this research.

This model represents the repetition of the search process but does not specifically identify a time frame. It is reasonable to expect that the cognitive changes measured in a short search task would not be as diverse or large as those measured in a search task with a longer time frame, since the longer time frame would allow more time (more cycles) for the individual to absorb the information being reviewed. Therefor the research proposed should be longitudinal, progressing over a period of several weeks.

A diverse set of criteria, and a set of criteria which resonates with the subject is also required. A complex subject area, one which is unfamiliar to the individual and requires the individual to gather

information through a search process, would involve a set of criteria which is reasonably representative of the subject area knowledge domain. This research will need to identify those criteria.

Literature Review

Debate about the nature of relevance has been constant throughout the history of information science. As early as Vickery (1959) relevance was identified as a foundational concept of the field. The indistinct philosophical foundations of the concept of relevance coupled with the difficulty of measuring the cognitive and situational aspects of it have led to a number of definitions of the term in information science with no single, canonical interpretation of the concept. Though the use of a dichotomous and static view of relevance was the basis for measurement of IR system performance, there has always been concern that some of the underlying assumptions about relevance as a measure were flawed. Cuadra and Katter (1967) noted the consensus at the time was that separate individuals making relevance judgments often disagree and those relevance judgments by the same user may change over time. These two observations echo the assessment by Schamber, Eisenberg, and Nilan (1990) nearly 25 years later that relevance is both dynamic (changing over time) and multifaceted (varying among users).

If relevance is highly variable, then its use as a constant measure is dubious. However, despite these early indications of dynamic relevance, much early information science research was based on a static, system-oriented view of relevance. This view treated the user as a constant, fixed entity. Saracevic (1975) distinguishes between a "system view" of relevance where the focus is on the system, and a "destination view" of relevance where *destination* represents interactions with the user. This brings the focus of relevance to the user's interpretations and perceptions of documents being evaluated. It is a subjective view of the user as the user relates the document to their knowledge and situation. Since a user's knowledge and situation are in constant flux, this perspective leads to a high

degree of variability, a dynamic relevance which changes in relation to time.

This interest in the user over the system eventually led to a deeper examination of the cognitive aspects of information science (Wilson, 1984; Belkin, 1990). Studies which followed this approach sought to examine human perception and cognitive processes in relation to information retrieval. Contrary to a systems-centered view of relevance, with this approach the user is not a constant but is a highly variable component of the search process. Understanding and meaning from the perspective of the user is paramount. In examining the cognitive nature of relevance, Bookstein (1979) notes that a concept of a "relevance to a request" as suggested by a system-centered view of relevance is flawed. Relevance is based on understanding and meaning and can only be judged by the user, not by a system. Relevance is user-centered and should therefore be defined in terms of the level of satisfaction of the user. Swanson (1986) continued the study of user-centered relevance making a distinction between "objective relevance" and "subjective relevance." He explains that with subjective relevance, the user is the final judge of what is relevant or not. Objective relevance is defined in more abstract terms with relevance judgments ultimately belonging to the world of objective knowledge. Based on these definitions, a contradiction can occur in which a document could be judged to be objectively relevant and subjectively not relevant (see also Wilson, 1973; Cooper, 1971). A significant body of research is based on objective interpretations of relevance which form the basis for evaluative metrics such as recall and precision. But these objective measures fail to examine and measure a significant portion of the user's cognitive process in making relevance judgments, and thus fail to adequately assess system performance.

Relevance which is user-centered is considered dynamic, and multidimensional. This subjective relevance is highly variable, yet is based on specific and presumably measurable cognitive processes on the part of the user. There has been some recognition that traditional topical relevance definitions may be inadequate to describe the complex and subjective relevance judgment process. Schamber et al

11

(1990) focused on relevance from the user's perspective and took issue with the topical bias of relevance definitions. They noted that relevance definitions were based on topicality using a best match between the topic identified in a query and the topic of the document (p. 758). They considered this approach to be problematic as it is based on the assumption that the subject terms used for topical searching represent meanings both in the query and the document. Identifying meanings in text is an elusive exercise, and research indicates there is some question that the user is even aware of the meaning of their search at various stages in the search process (Cooper, 1971; Belkin, 2000). Schamber et al (1990) note that topic-based searches fail to capture the full breadth of the user's information need and in order to form a better understanding of the dynamic and situational aspects of relevance judgments, research should examine user's relevance criteria choices and interactions in the selection of those criteria with information seeking behavior (p. 773). The research proposed here seeks to examine this relationship.

A number of definitions, frameworks and models exist for the concept of relevance in information science with no consensus on which is correct. The following section examines a number of these frameworks in relation to the proposed research project.

Definitions and Frameworks

Relevance is dynamic, changing as time progresses. Mizzaro (1998) specifically refers to this property of relevance as the "time dimension." He indicates that what may be relevant at one point in time may not be relevant at another point. In his formal model, Mizzaro sees a user in a "problematic situation" (from Belkin et al, 1982) progressing through three operations: perception, expression and formalization which results in a query. These operations are a function of time.

It is not clear where the user fits into Mizzaro's framework where relevance is defined as a relation of the document or surrogate to the query with no mention of the user's perception of that relationship (ibid, p. 310). Such an analysis seems to be missing the cognitive role of the user although

the presentation of the task model recognizes the benefits of moving beyond topical IR systems. The "stereotypes of tasks" presented identifies characteristics of documents that build on cognitive, user-centered research. These document characteristics include, but are not limited to, type of document, document character (theoretical, review), page length of document and date of document (publication date, meeting date) (see also Barry, 1994, 1998; Barry and Schamber, 1998; Park, 1993).

Saracevic (1996, 2006) provided examinations of the progress of relevance research in information science. He noted that relevance remains a key measure for the retrieval of information objects with users as the ultimate judge of that relevance. A critical review of the systems, communication, situational, psychological, and interaction frameworks led to a conclusion that relevance as a concept in information science is not a simple, self-contained, singular concept, but is a multifaceted *system of relevances*. As such, researchers must recognize all levels of the system and their influence on the relevance decision. The author also notes the existence of *manifestations of relevance* as attributes or dimensions of relevance. These manifestations move beyond the commonly identified topical relevance and examine the complex set of dimensions or criteria which are part of the relevance assessment process. The area of "clues research" is identified as the criteria user's identify when making their relevance judgment. These relevance clues are also identified as *criteria for relevance* judgments, and *relevance criteria* in other research.

Cosjin and Ingwersen (2000) built on the work of Saracevic (1996) and others to develop a revised table of attributes and manifestations of relevance (Cosjin and Ingwersen, p. 547). The manifestations of relevance identified are topical, cognitive/pertinence, situational/utility and socio-cognitive. These are categorized as *affective relevance* (from Saracevic, 1996). These affective manifestations of relevance represent expressions of cognitive changes and can be associated directly with the relevance criteria and categories identified in the user-centered cognitive studies by Barry (1994) and others (Barry & Schamber, 1998; Park, 1993).

Cosjin and Ingwersen (2000) emphasize that "interaction" as an attribute of relevance is dependent on time, suggesting that as a user progresses through a search process, affective relevance manifestations may change. Cosjin and Ingwersen note that the progression of time "influences the user's [relevance] decisions" and it is the cognitive changes which occur over time through interaction that lead to this influence (ibid, p. 544). Theoretical work by Mizzaro (1998) also makes this observation, and studies by Vakkari (2000), Spink, Greisdorf and Bateman (1998) and Wang and White (1999) provide some suggestions that this influence exists.

Analysis of this body of research leads to the conclusion that the interaction of time, and affective and cognitive aspects of the user affect relevance judgments. In this proposed study, relevance is defined as the user's perceptions of the document's importance to their information need at a point in time. It is therefore both multidimensional (varying among users) and dynamic (varying over time). This concept of relevance provides a "real world" view of relevance suitable for information science research (Schamber et al, 1990; Wilson, 1973; Harter, 1992; Saracevic, 1996; Borlund, 2003; Borlund & Ingwersen, 1998). It involves a system of relevances (Saracevic, 1996) and focuses on the cognitive and situational level of a stratified model in which users examine documents and absorb information to fill their information need, a process which changes their cognitive state. For purposes of this study, relevance judgments are defined as the process of a user evaluating a document or document representation as being relevant, partially relevant, or not relevant to their information need. Relevance criteria are those factors that contribute to the user's relevance assessment for a positive (document is relevant), negative (document is not relevant), or uncertain (partially relevant / don't know if it's relevant or not) assessment.

Research into Criteria Used to Make Relevance Judgments

A more complete understanding of multidimensional and dynamic relevance calls for further

study of the cognitive context of the relevance judgment. Schamber et al (1990, p. 773) propose examination of the criteria used by users to perform relevance judgments in relation to information behavior and an evaluation of the consistency of these criteria choices. These criteria are part of the user's expression of relevance and when combined with a relevance judgment measured as integral or categorical value they provide a richer expression of the judgment process.

Barry (1994) conducted a study which identified 23 categories of relevance criteria which applied not only to the information content of the document, but to subjective aspects of document interpretation such as the user's beliefs and previous knowledge, contextual factors such as other sources of information in the environment, the user's situation, and the quality of the source of the document (reputation, visibility, authority). Barry's methodology required subjects to identify "items" on the document that prompted them to "pursue" or "not pursue" a document. Documents were selected at random from a set of documents retrieved so a full range of relevant, partially relevant and not relevant documents were examined.

Schamber (1991) conducted similar relevance criteria research with 30 users in three different occupational fields. Barry and Schamber (1998) later combined the data collected from the Barry (1994) study with Schamber (1991). Both studies examined a full range relevance judgments. Despite the diversity of subjects' backgrounds, there was consistency in the criteria selected by the groups in the two different studies. The authors note that user's selection of relevance criteria is "somehow linked" to the user's background, knowledge or experience. Specifically how this background or experience influences relevance choice is not reported.

Barry's (1994, 1998) studies were effective in identifying a set of document attributes and contextual and situational characteristics which searchers use to assess a document as relevant or not. The criteria were categorized into the groupings identified in Table 1 which represent a cross-section of the attributes and manifestations of relevance as identified by Cosjin and Ingwersen (2000). The

relevance criteria and categories reported in the study have been identified in other studies (Park, 1993; Maglaughlin & Sonnenwald, 2002; Tang & Solomon, 1998). The criteria identified, however, did conflate environmental/situational characteristics such as obtainability/cost with document characteristics such as depth/scope and recency. While this identification and categorization is consistent with Barry's exploratory research goals, it does mix the cognitive and situational aspects of relevance judgments. Further relevance criteria research analysis should provide a distinction between these aspects.

Table **1 - Relevance Criteria Groupings Reported by Barry (1994, 1998)**

Grouping	*Criteria category*
content of documents	depth/scope,objective accuracy, tangibility, effectiveness, clarity,recency
user's experience and background	background/experience, ability to understand, content novelty, source novelty, stimulus document novelty
user's beliefs and preferences	subjective accuracy/validity, affectiveness
sources of documents	Source quality, source reputation/visibility
document as a physical entity	obtainability/cost
user's situation	time constraints, relationship with author

Park (1993) performed a content analytic study to identify criteria important to users making relevance judgments. The study involved 10 subjects including a cross-section of college faculty, doctoral and masters students across several different disciplines. The results were used to generate three major categories of relevance assessments and identified several relevance criteria reported by subjects which were consistent with those found by Barry (1994).

Schamber and Bateman (1996) used results of three previous studies by Schamber (1991), Su (1993) and Barry (1994) in an attempt to reduce and synthesize the number of relevance criteria used and produce a measurement instrument involving user's relevance criteria. The authors note that some

subjects in their study had a problem with negative applications of criteria and appeared to have underreported that in their results. Results provided some indication that users understood the concept of relevance criteria and could understand and use categorizations of those criteria.

Maglaughlin and Sonnenwald (2002) worked with 12 graduate students who examined 20 documents each and identified relevant passages in each of those documents. The criteria of "currency" was eliminated from the reported results, but it is reported that nine participants indicated that they wanted current documents, so for 75% of the sample, a total of 180 documents, currency was an implicit relevance criteria. Documents were rated as relevant, partially relevant or not relevant. The results identified 29 relevance criteria consistent with previous research (Barry, 1994; Park, 1993). Researchers found more relevance criteria in relevant documents than in non-relevant documents, possibly indicating relevant documents are read more closely.

Crystal and Greenberg (2006) asked 12 subjects to examine documents found on the Web and identify relevance criteria in the document surrogate and the document. Using content analysis and statistical analysis they identified a number of relevance criteria. Results identified a few criterion were commonly identified by subjects, and a larger set of criteria which were identified less frequently. The criteria of "topicality" and "research group" were criteria frequently identified by their subjects, consistent with the suggestion by Wang and Soergel (1998) that epistemic value (research group) must be satisfied before other search criteria are considered in the search process.

Xu (2007) examined relevance criteria used in 'hedonistic searches" which the author identifies as searches for pleasure. Results reported are consistent with studies which indicate work task has as significant influence on information seeking behavior (Li, 2008). Xu (2007) surveyed 113 subjects who were allowed to browse for information for fun, identified as *affective stimulation*. The relevance judgments examined were considered a form of affective relevance (from Saracevic, 1996). Xu examined what is termed *informative relevance* as the amount of information a document provides in

general, not necessarily as part of a problem solving effort. Since a hedonistic search is "for fun,"

subjects are not solving a problem, but merely trying to gather information. Xu hypothesized that

affective relevance, treated as the emotional impact of a document, is closely related to informative

relevance. Xu reported strong statistical results which suggest that "topicality," "novelty," and

"reliability" contribute to informative relevance, but "scope" and "understandability" do not. Xu also

reported that "topicality" and "understandability" impact affective relevance, but "novelty" does not.

These results add further evidence that context and situation affect relevance judgments and the criteria

used to make those judgments.

Summary of Relevance Studies Examining Criteria Choices

As these studies illustrate, there are criteria beyond topicality which users employ to evaluate

whether or not a document is relevant. The recognition of these criteria extends back to Cuadra and

Katter (1967) who identified them as intervening variables in the relevance judgment process. More

recent research by Barry (1994, 1998; Maglaughlin & Sonnenwald, 2002; Xu & Chen, 2006) have

identified a set of criteria which is consistent across multiple independent studies. It is important to

note that Barry's studies involved all documents evaluated by users, regardless of range and direction of

the relevance assessment (relevant, partially relevant, not relevant). Other studies have duplicated this

methodology and have argued for the importance of evaluating partially relevant documents and

negative relevance judgments (Spink et al, 1998; Hjorland, 2000). Any study of relevance criteria

choices should therefore capture a full range of relevance judgments, from relevant, to partially

relevant, to not relevant.

A number of information science studies have examined information seeking behavior, but only

have handful have examined relevance judgments in relation to the ISP. Researchers have recognized

the need for this research, and some have stressed the situational behavior of relevance should also be

examined in this context (Saracevic, 2006, p. 93). Understanding these interactions can provide insight into the user's cognitive processes and identify document criteria deemed valuable in making relevance judgments from the user's perspective. The studies presented in the following section have pursued this goal.

Studies Examining Relevance Assessments in Relation to the Information Search Process

A clear understanding of the ISP is crucial to the examination of the dynamic nature of relevance judgments. Most of the studies reviewed here have based their ISP model on the framework proposed by Kuhlthau (1991). Kuhlthau examined the information seeking behavior of high school students and college seniors over a series of five studies and developed an information search process model. Research was conducted in a naturalistic environment using a variety of methods. Based on this research, the author formulated a model of the ISP that included a series of six stages identified as initiation, selection, exploration, formulation, collection and presentation. The model is often interpreted as being strictly sequential, though Kuhlthau interpreted these stages as potentially being iterative and recursive (1993, p. 69).

A small number of relevance studies were conducted in the 1960s. Within the traditional IR model, relevance was considered to be a relationship between a system output and an information requirement, and was recognized as a match between a search query and a document. It was ultimately considered a property of the system (Cuadra & Katter, 1967; Rees & Schultz, 1967; Saracevic, 1996). Despite early signs that user evaluations of relevance were varied, researchers largely ignored such variations in pursuit of a valid metric for IR system performance. In response to growing concern over variations in user relevance judgments, a few studies from this time period examined the nature of relevance from a user's perspective and attempted to identify the factors which contributed to dynamic relevance.

Rees and Schultz (1967) noted observations by Vickery (1959) and others that a concept of relevance as a property of a system was flawed. The authors also noted that relevance was not limited to dichotomous judgments (a document is either relevant or not relevant), and that a user-centered approach to relevance could form the basis for a useful metric (Rees & Schultz, p. 8). The researchers considered a relevance judgment to be a decision by the user which provides a measure of the relation between the document and the information problem ("the initial request," p. 16). Relevance was also considered to be graduated: the document might be relevant to the information problem, or the document might be some degree of less relevant.

The researchers examined relevance judgments from 184 judges across three search stages and identified several manifestations of relevance which they described as an "aspect of relevance." Aspects of relevance were *overall relevance*, *formulative relevance*, *methodological relevance* and *overall usefulness*. Subjects were from a variety of backgrounds with varying levels of education, professional experience and professional orientation. The research topic was medically related, and all subjects had some experience in the medical field. The research methodology involved having judges execute a simulated research project. A search process was developed and three search stages were identified: 1) formulation of the research problem, 2) experimental work, and 3) data analysis.

Based on the evaluation of 400 relevance ratings from each of the judges, a variety of results were reported. Researchers found fewer documents were rated as relevant in later search stages. Researchers also found variations in the manifestations (aspects) of relevance across search stages, though statistical results on these variations were not reported directly. Also noted was statistically significant relationships between the background of the subject and their relevance judgments at specific stages. These results suggest that some portion of the variations in relevance (dynamic and multidimensional) may be the result of interactions between search stages and various manifestations of relevance.

Cuadra and Katter (1967) considered the common view of relevance in the mid-1960's to be that of a "black box" -- a research component whose inner workings are ignored. As long as a relevance judgment were made by a judge, a criterion measure was provided; details of how the judgment was made were considered irrelevant. The authors noted that it was common, however, for two expert judges to disagree (multidimensional relevance), and that an expert judge may often change his or her assessment over time (dynamic relevance). The researchers theorized that the "discriminatory response" (relevance assessment) was a function of the document and information requirements. The relevance assessment was also considered a function of "user states" which were identified as what current research considers cognitive state, task and situation.

Cuadra and Katter had 140 subjects examine nine abstracts and make relevance judgments using a graduated relevance scale. They then applied a "treatment" where the judges were directed to make a "simulated" judgment based on 14 "assigned point of views" (p. 269). Subjects made a second set of relevance judgments based on these assigned points of view and researchers measured the difference between the first set of judgments and the second. Based on these responses, a number of groupings were identified which were considered intervening variables in the relevance judgment process. Though these groupings and the authors speculation hint at manifestations of relevance and relevance "clues," the research protocol used did not ask the judges what criteria were used to make their judgments. The authors speculation at specific influences of the relevance judgment process tend to overstate their findings.

The cognitive turn in information science led to renewed focus on the role of the user and their cognitive processes. Studies such as Belkin (1982) and Bates (1989) examined the information search process and acknowledged changes in cognitive stage on the part of the user, but they did not focus on relevance judgments or, more specifically, criterion involved in the relevance judgment process. In the 1990's a number of researchers began to examine relevance judgments and criterion for those

judgments in relation to progress through the search process. Cool, Belkin, Frieder, and Kantor (1993) examined the relevance judgment process across several ISP stages and asked subjects to explain why they were making those judgments. They used a convenience sample of approximately 300 undergraduates taking an introductory computer science course at a U.S. university. Students were required to write an essay on a topic of general computer science interest using at least five sources. Students answered a questionnaire about each document they reviewed. The students were asked to specify when (date and time) they evaluated the document, whether they anticipated using the document for their paper, and to explain why they made that decision. The students were also asked to indicate where they were in the process of completing the paper, at the time of judging each document. The authors' analysis examined facets of document usefulness as expressed by subjects during the relevance judgment process. As a result of this analysis, they identified six such facets of the relevance judgment process: topic, content/information, format, presentation, values, and oneself.

Wang and Soergel (1998) examined criterion for relevance judgments as identified by subjects who were experts in the field in which they were conducting searches (p. 130). Based on their analysis, the 'epistemic value' of a document was the prerequisite for all other values for the document. An emphasis on epistemology may be partially explained by the sample bias towards knowledgeable subjects. Criterion of relevance included "quality" and "orientation/level" but the criteria of 'ability to understand' reported by Barry (1994, 1998) is missing. This omission is potentially due to the expertise level of their subjects who, as experts, were able to comprehend all documents reviewed. The researchers identified several decision rules subjects used to make relevance judgments using one or more criterion of relevance. The researchers did not report changes in the importance of relevance criteria over the search process.

Bateman (1998) examined choices of relevance criterion in relation to progress through the ISP. Relevance criteria identified by Barry (1994), and Schamber and Bateman (1996) were reduced for

clarity and grouped into nine categories to provide subjects with a context with which to interpret the criteria. Bateman's (1998) study involved 35 graduate students who were asked to complete surveys on the information sources they considered most valuable (thus highly relevant). Bateman (1998) notes that subjects did not report moving through the ISP in a "uniform manner" and instead reported an uneven distribution of stages, with some respondents reporting being in multiple stages at once (ibid, p. 27). She does not report variations in criterion importance across the ISP, contradicting other studies which report that criteria such as "novelty" appear to be more important to users in later stages (Vakkari, 2001; Vakkari & Hakala, 2001; Tang & Solomon, 2001; Wang & White, 1999; Hirsh, 1999). The limited sample size and the descriptive statistics used may not have been sensitive enough to detect these changes. Also, Bateman was working with only highly relevant documents and this approach may have skewed the result set towards a more homogeneous set of documents. This method led to the exclusion of partially relevant documents which may involve more malleable relevance judgments and associated choices of criteria.

Tang and Solmon (1998) conducted a series of studies which examined relevance judgments of a single graduate student preparing a term paper. The authors limited their examination to two ISP stages: relevance judgments based on the reading of bibliographic entries, and relevance judgments made after reading the document referenced by the bibliographic entry. The subject was allowed to re-evaluate the documents selected based on bibliographic descriptions, was allowed to mark documents as "partially relevant" and was allowed to go back and re-evaluate those documents. A second observation session was conducted one month after the first session, allowing the subject to read the documents, mentally process the contents and then perform an evaluation based on "usefulness." The authors report that the "subject's approach appeared to be more certain" (ibid, p.253) in evaluating the results of their search (as the subject's mental model changed) later in the search process. Some relevance criterion such as "topical relatedness" and "recency" are reported to have decreased

importance (in terms of the frequency with which they are invoked) later in the search process. The authors also report some "fuzziness" of relevance observations during the process (the subject cannot determine whether or not the document is relevant), suggesting the need for partial relevance judgments.

Hirsh (1999) performed a study with ten fifth grade children (ages 10-11) who were assigned a four week long research project to examine criterion choices of relevance judgments using an ISP which referenced Kuhlthau's (1991) model. Descriptive statistics were reported based on two interviews, one conducted at the beginning of the search process during the subject's first search session, and the other during the third week of the research project. The findings add to the evidence that the use of topicality decreases later in the search process as users begin evaluating documents on a wider range of relevance criteria.

Wang and White (1999) focused on the reading of documents in a long-term study of a convenience sample of 15 experienced researchers with the pool of eight professors, six doctoral students and one masters student. Three ISP stages were used: *selecting*, *reading* and *citing*. Researchers identified topicality, novelty and recency as the most commonly selected relevance criteria and added 'cognitive requisite' (the ability to comprehend a document), a criteria which appears to be very similar to the criterion for relevance reported as the 'ability to understand' identified by Barry (1994), or 'understandability' as identified by Cool et al. (1993). The subjects tended to select more documents as relevant than those they actually used, and applied more diverse relevance criteria in later stages. The greatest variety of selection criteria were reported in the "citing" stage. The authors noted that multiple criteria are commonly used when a positive relevance judgment is made. The study adds evidence for the dynamic nature of the search process, but does not report specific preferences for criteria used to determine relevance in relation to ISP stages.

Tang and Solomon (2001) conducted relevance criteria studies using both laboratory and

naturalistic approaches. The laboratory experiment involved 90 undergraduate students who were given an assignment to conduct research and prepare an outline. The study limited evaluation to only two stages of the ISP identified as stage 1 - reading a bibliographic description of the document, and stage 2 - reading the document. Changes were noted in the selection of the subject's rating of the importance of criteria used to determine relevance. Results suggest a change in the selection of some relevance criteria, when moving from stage one to stage two for the criteria of clarity, importance, newness, recency, topical focus, topical relatedness, but the authors do not report the statistical significance of the results.

Vakkari (2000; Vakkari & Hakala, 2000) performed a broader study of the model shown in Figure 1, examining six ISP stages and a number of criteria used to determine relevance. Eleven students were used in the longitudinal study which involved preparing a proposal for a master's thesis. The study examined changes in relevance criteria choices in relation to task performance. Results from these studies suggest that users identified more documents as relevant early in the search process and identified fewer documents as relevant later in the search process. The researcher noted that the categories of "novelty" and "interest" were selected more during the later stages of the ISP, and "topicality" was the most commonly selected criteria. The author speculates that users, having selected a set of relevant documents earlier in the search process, are more interested in finding novel information (documents different than their current selected set) later in the search process. Though this study provided some useful insights and was empirical research, a small sample size was used, and its analysis and conclusions were based on limited reporting of descriptive statistics.

Taylor, Cool, Belkin, and Amadio (2006) also performed a broader study of the model shown in Figure 1. Researchers identified criteria used for relevance judgments by performing content analysis of comments made by subjects during document selection. A random sample of 40 subjects from the results of a previous study (Cool et al., 1993) with 300 undergraduate students was used. Researchers

used four search stages consistent with Kuhlthau (1993). Findings suggested criteria selection changes, by subjects, as they progressed through a search for documents relevant to an information need. The authors report a statistically significant relationship in terms of frequency of selection for the preference of certain criteria in early search stages ("recency" and "specificity") and for other criteria in later search stages ("source novelty" and "interest").

Task and Situation Influences on Information Seeking Behavior

The fulfillment of an information need is commonly driven by a *work task,* considered a sequence of activities directed at fulfilling the information need (Hansen, 1999). Ingwersen and Jarvelin (2005) suggests there are classifications of information that must be considered. Information seeking behavior may be different for different classifications such as *the problem at hand*, the *knowledge domain*, and *problem solving* tasks. Task complexity and other factors also contribute to these differences (Vakkari, 1999).

Li (2008) examined the relationships between work tasks, search tasks and information seeking behavior. Li performed two studies. The first study was used to determine a list of facets and subfacets of work tasks and search tasks. This involved semi-structured interviews and content analysis of transcripts with 24 subjects. The output of this study provided a list of facets and subfacets of work task and a relationship to search task. Statistical analysis of these results provided an indication of which facets and subfacets had the most impact on search tasks and information seeking behavior. This information was used to select work task types for the second study which examined work tasks and search tasks in greater detail.

A number of different work tasks and their relationship to search tasks and information behavior were examined in study two. Using the results of study one, work tasks were selected which had varying complexities and goals. Results showed that most subjects used Web resources before using library resources, but that library usage more common for high complexity work tasks. Level of

education was also found to have an effect on some search tasks. A number of work task facets were found to have some effect on the search tasks and information behavior: time, goal, process, urgency, subjective task complexity, knowledge of task topic, and salience of tasks.

Gross (2002, 2000) conducted research on the use of information services by subjects who were performing research either as proxies, or with a deep level of unfamiliarity for the topic being searched. Gross noted that search models such as Kuhlthau's (1991) ISP assumed a visceral information need, but information needs are often artificial. She noted that in situations where the information query is imposed, relevance assessments become difficult for the subject. Since IR research often involves imposed information needs, the research suggests that the level of topic familiarity and the ability to formulate search queries effectively becomes a consideration in experiment design. It is clear, based on this research, that to design experiments to examine the influence of progression through the ISP on relevance judgments, researchers must be careful to control work task influences as much as possible.

Assessment of Prior Research

The user studies presented here examined the dynamic nature of relevance judgments by observing subjects both directly and indirectly during the ISP. The selection of categories for ISP stages or search behaviors varied, and was, in some cases, extremely narrow, potentially missing subtle changes in the behavior of users. With the exception of Vakkari (1999) and Taylor et al. (2007), this research did not focus on specific interactions between choices among criteria for relevance judgment and the progression through the ISP. With the exception of Tang & Solomon (2001) and Taylor et al. (2007), sample sizes were too small to allow for strong statistical analysis beyond descriptive statistics. Though some of these studies suggest that subjects use more than one criterion to make their relevance judgments, none of them specifically examined or analyzed the groupings of criteria used by subjects.

Of the studies identified here, only Vakkari (2000) and Taylor et al. (2007) examined the relationship between ISP or stage in task completion, and criteria choices during the relevance

judgment process, in combination with the use of three levels of relevance including partial relevance. This is important since research has shown that partial relevance is common early in the search process, and it is these partially relevant documents which are re-evaluated later in the search process using various criteria (Spink, Greisdorf, & Bateman, 1998; Tang and Solomon, 1998).

Justification for Research Approach

To advance understanding of these processes, this study proposes to examine these interactions during the relevance judgment process directly, and to use stronger statistical methods to provide additional clarity and depth. The proposed research will examine these criterion choices and their evolution during the ISP using the complete range of relevance judgments (including partial relevance). Additionally, use of a larger sample size and more sensitive statistical methods will add to the statistical strength of this study. Subjects will be asked to identify criteria choices directly, rather than using the indirect method of content analysis of subject comments or interviews used in previous studies (Bateman, 1998; Wang & White, 1999; Vakkari, 2000; Taylor et al., 2007). Finally, the proposed study will also perform statistical analysis to determine whether there are consistent groups of criteria used to determine relevance during the ISP.

The studies identified here have all examined one, or in a few cases, several aspects of the nature of dynamic relevance judgments. This study proposes to examine a number of these aspects of relevance dynamics in a single study, allowing various interactions to be examined with a single subject pool. This study will also examine groups of relevance criteria selections in relation to the ISP stage, a relationship which has not been examined. This study also proposes to examine relevance criteria interactions in depth, using a more detailed level of criteria choices (examining more criteria) and proposes to use a sufficiently large sample size to provide additional statistical strength to the results.

Preliminary Research

In preparation for this proposed research, a series of studies were conducted. The results of these preliminary studies have informed the development of this proposal as discussed below.

Research Practicum

The goal of the research practicum was to perform exploratory research to determine if choices for relevance criteria varied over the course of a search process. In order to accomplish this goal, the following procedures were used. A convenience sample of 39 subjects was drawn from the student population of business school students at an American university. The subject pool was a mix of approximately 20 % paid subjects and 80 % unpaid subjects. All subjects were assigned the same specific search task and performed the task while being monitored in a computer lab at the university. Online, Web-based surveys were used to collect data on the subject's background, their relevance criteria choices, relevance assessments and ISP stage. All data was self-reported by the subject. Subjects worked alone on their search problem and took between 45 minutes and two hours to complete the test session. The information search was conducted with online searches using the ABI/Inform online database to find documents in that database to help solve their search problem.

Subjects were asked to retrieve at least 10 documents to solve an assigned research questions. Documents retrieved could be any combination of "relevant" and "not relevant" documents. Relevance was captured on an interval scale and allowed for a full range of relevance including partial relevance.

On completion of the search test, a user interview was conducted using open-ended questions about the documents selected and the subject's reasons for considering the document relevant or not relevant. These subject interviews were recorded and later transcribed.

The relevance criteria categories developed by Barry (1994) and Park (1993) were used in this study to identify which specific categories and subcategories were used by the subjects during the search process. The total number of relevance criteria categories was reduced for clarity and simplicity

to avoid survey exhaustion. The ISP stage references used in this study were a combination of

Kuhlthau's (1993) ISP stages and Ellis's (1997) search patterns. Subjects were shown a list of search

stage descriptions; a help page was also provided with a more detailed description for each search

stage.

Table 2 - Relevance Criteria Selections by Users as Percent of Total

Criteria	Percent Selected*
Clarity of presentation	10.04%
Ability to understand	10.02%
Depth/scope	9.64%
Precision	9.64%
Specificity	9.48%
Amount of information	9.25%
Interest in topic	9.21%
Instructional	8.93%
Recency	8.32%
Authority of author	8.04%
Bias of author	7.42%
Total	100.00%

* percentage of total selected for all relevance judgments

Table 3- ISP Stage Reported as Percent of Total

ISP Stage	Percent Selected
Extracting information	17.10%
Browsing for information	16.71%
Trying to focus on topic	15.39%
Learning about the topic	13.31%
Becoming informed	10.76%
Verifying information	10.53%
Defining and extending focus	9.81%
Completion and testing	6.39%
Total	100.00%

* percentage of total selected for all relevance judgments

The results shown in Table 2 indicate that criteria for relevance were selected in range

approximately between 8% and 10 %. The consistency of these percentages provides some indication

that most criteria were selected consistently by subjects, with what appears to be little discrimination

between criteria based on frequency of selection.

The selection of ISP stage shown in Table 3 is based on subject-identified search stages. In

terms of frequency of selection, most documents were reviewed in the "extracting" and "browsing" stage. An orderly progression through an ISP would presume that subjects would select each search stage in the roughly the same proportion. The selection of search stages by subjects in this study does not show a consistent progression using the ISP model presented to the subjects. The uneven distribution may have been a result of the short time frame allowed for the experiment (approximately one hour) which did not allow subjects to make an orderly progression through the ISP. These results may, however, indicate that the ISP models used to identify search stage are not representative of the cognitive process of subjects. It is possible that the stages with the largest percentage of selections in Table 3 (extracting, browsing, focusing, learning) represent the search process of subjects, and that the other stages (verifying, becoming informed, defining and extending focus, completion and testing) represent uncommon search behaviors.

Analysis of results concerning changes in the selection of specific relevance criteria in relation to the ISP stage did not find a statistically significant association. Subjects did appear to have more confidence in their assessments and did use a variety of criteria in later stages, a finding consistent with consistent with Vakkari (2000), possibly indicating subjects were more selective and discerning in later stages of the search process.

In reviewing the methods used in the research practicum, several potential issues were revealed. It is possible that a short duration study of the ISP may not allow examination of progress through a search process and resultant changes in criteria used for relevance judgments. Instead, a longitudinal study may allow subjects more time to select documents, absorb the content, and make more thoughtful assessments of relevance. Additionally, the number of relevance criteria chosen by subjects for each document was larger than expected, suggesting that either subjects use a large number of criteria for each relevance judgment, or that the number of criteria provided for selection was too large and subjects were confused about which to select. It may be more appropriate to provide subjects with a

smaller set of relevance criteria from which to choose. The design of methods for Pilot Study 1 reflected these observations.

Pilot Study 1

The goals of Pilot Study 1 were to explore relationships between relevance criteria choices and progress through the information search process. Additionally, this study tested the methods and data collection instruments which were changed based on analysis of the research practicum results. The research was conducted in June of 2007 using a convenience sample of 16 subjects who were undergraduate business students taking an online business course at an American university. The methods were the same as those proposed for the main study with two exceptions. In this pilot study, the selection of the search stage was made by the subjects once, at the start of the search process, when the subjects entered a search query (see Appendix H). Also, this study had a single deliverable, a requirement for subjects to complete a research project, and did not require the preparation of interim deliverables. Subjects had three weeks to complete their research project and were required to use the Web-based search engine described below to record information about their relevance judgments and criteria choices during the ISP. Participation was voluntary; subjects who completed the research were given extra credit.

The 16 subjects recorded 558 relevance criteria choices. Analysis of the data entered indicate an uneven reporting of search stages by subjects as shown in Table 4. The lack of preference for the search stages identified as "differentiating" and "verifying" could be attributed to the subjects' conducting an initial series of searches to collect documents, and then not reusing the search engine (and data collection instrument) as they progressed through the later stages of the ISP. It also appears from this data that the concept of "extracting" represents a consistent choice for search stage, and "exploration" (with the description of "scanning for information") was chosen much less frequently than "extracting."

Table 4 - Pilot Study 1 ISP Stage Selections

Search Stage Code*	Count	Pct of Total
extracting	228	40.86%
initiation	177	31.72%
exploration	89	15.95%
verifying	41	7.35%
differentiating	23	4.12%
Total Result	**558**	100.00%

* Subjects chose search stage based on the definitions shown in Appendix H

These results provide some indication that subjects may not report, or may not progress through a Web-based search process in a manner which is entirely consistent with prior research by Ellis (1997) or Kuhlthau (1993). To increase the likelihood of subjects providing a more consistent reporting of their search progress, interim deliverables were added to the method for second pilot study and the main study. These interim deliverables encourage subjects to work more consistently throughout the duration of the research effort and thus provide relevance judgment data for search stages which were underreported in the pilot study. They also provide an opportunity to analyze criteria selections and relevance judgments in relation to a stage in task completion. Additionally, the data collection instrument was changed to allow a subject to select a search stage with the evaluation of each document/web site. This differs from the procedure used in the first pilot study where the search stage was selected once when entering the search query, and was then assumed to be the same for each document evaluated.

Pilot Study 2

To test modifications to the methods for the main study, and to further examine relationships between relevance criteria choices and progress through the information search process, pilot study two was conducted. A convenience sample of 82 subjects from a population of undergraduate students at an American university participated in the study in Fall of 2007. The methods for this study were the same as those proposed for the main study (see the Proposed Research Design section), with the exception that subjects did not indicate a level of importance for relevance criteria selected. Subjects

examined and reported on a total of 818 distinct Web documents. Data collected included the URL of the Web page where the document was retrieved, a search stage (selected by the subject from a list of search stages), and various criteria that the subject identified (selected from a list of criteria) as pertinent to their relevance judgment.

The research project assigned to the subjects requires the creation of a set of presentation slides on a research topic. This final project deliverable can be considered the *work task*, and the interim deliverables represent stage in task completion. Table 5 provides a count of Web documents evaluated by subjects in relation to the project deliverable which was due when the relevance judgment was made (the stage in task completion). The data in Table 5 suggests that the majority of the documents were evaluated in preparation for the "detailed outline" deliverable, suggesting that this production of this deliverable required the subjects to gather information and learn about the topic before proceeding. A large number of documents were also identified in preparation for the "rough draft" and "final presentation" deliverables, suggesting the subjects may be adding sources to their reports later in the research process, filling existing information gaps and possibly 'backfilling' and adding sources to potentially raise their project grade.

Table 5 - Documents Assessed by Deliverable Due

Deliverable Due	Documents Assessed
Abstract	81
Detailed Outline	293
Rough Draft	187
Final Presentation	257
Total	**818**

Results for 82 subjects.

Table 6 reports the total number of relevance judgments made in relation to the information search process stage. As reported in the pilot studies, subjects appeared to demonstrate a preference for selecting some search stages over others. The *initiation*, *extracting*, and *exploration* stages are most commonly selected, and *differentiating* and *verifying* appear to be underreported. Subjects do not

34

report a consistent, smooth progression through all search stages provided.

Table 6 - Documents Selected by Stage in the Search Process*

Stage	Documents Assessed
initiation	527
differentiating	229
exploration	543
extracting	1550
verifying	256
Total	**2578**

*n=82

An examination of the criteria subjects reported in selecting pages may provide some indication

of their reasons for selecting or rejecting Web page documents as they progress through the search

process. Since there were a large number of criteria provided and reported by the subjects, Table 7 is

simplified to list only those criteria which changed most as the subject moved through the search

process. This data in this table identifies relevance judgments reported by search stage, as a percentage

of total relevance judgments for that search stage. Based on the counts for the selection of criteria in

these stages, subjects appeared more concerned with the "amount of information" and "depth" in earlier

search stages ("initiation" and "differentiating") than in later stages. Based on this analysis, the criteria

of "novelty" of sources appears to be slightly more important in later stages, as does the "structure" of

the document and the time constraints of the subject. This could be because subjects are looking for

new sources of information and documents of different structure, providing a different approach and

new information for their subject area.

Table 7: Criteria Choices as a Percentage Selected for a Search Stage

Relevance Criteria	Search Stage			
	initiation	differentiating	extracting	verifying
affectiveness	0.57%	4.37%	2.77%	3.52%
amount of information	8.92%	12.23%	10.77%	7.42%
authority	1.90%	3.49%	2.65%	3.91%
bias	1.90%	1.75%	1.94%	3.13%
depth	8.92%	12.23%	9.03%	7.42%
novelty	1.14%	3.93%	2.06%	2.73%
recency	8.35%	7.86%	6.97%	6.25%
source quality	4.93%	3.49%	4.06%	3.13%
structure	6.26%	5.68%	7.55%	8.20%
time constraints	0.57%	1.31%	1.29%	2.34%

* reported as a percentage of the total criteria choices (not all are shown) within a search stage

An examination of the frequency counts of relevance criteria selected grouped by project deliverable required provides similar, though not identical, insights into the progress of the subjects through the ISP, as shown in Table 8. (Note the subjects were allowed to select more than one criteria per document evaluated, so it is possible to have a criteria count greater than the 818 documents evaluated by the subjects.) Subjects working on the final presentation, who were approaching the end of their research effort, were more likely to select "amount of information" and "understandability" as important criteria in their document choice. The criteria of "accuracy of the document" is also important. It is possible that the subject's are looking for greater understanding of the topic as they prepare to conclude their work, and are thus looking for additional depth and accuracy.

Table 8: Comparison of Criteria Code Selections for Rough Draft and Final Presentation

Criteria Code	Rough Draft	Final Presentation
accuracy	68	82
affectiveness	14	38
amount of information	56	98
authority	15	30
bias	15	15
breadth	36	47
depth	62	72
novelty	9	22
recency	41	64
source quality	24	28
structure	39	64
time constraints	13	17
understandability	72	123
Total Result	**588**	**873**

These results do demonstrate some minor variations in the counts of the selection of relevance criteria across search stage. But analysis based on frequency of selection for these criteria does not reveal a statistically significant variation in the importance of these criteria to subjects in relation to stage in the search process. A stronger statistical method such as chi square could be run on this data, but chi square would merely report an effect, not directionality or causality. In measuring the effect of search stage and criteria for relevance chosen in that search stage, a statistically significant relationship in terms of frequency of selection could be found. But given the results reported here, a thorough analysis would indicate that the relationship would most likely be due to the subjects' propensity to make more document selections in certain search stages, so that search stage selectivity would appear to be the reason for the effect (more documents are selected by the subject in that search stage), and the importance of relevance criteria to the subject in that search stage would not necessarily be the cause. A modification of the research protocol which asks the subject to provide an integral value to indicate the importance of the relevance criteria to them in that search stage would provide better data for analysis. With this change in the research method, analysis of the results of the experiment would provide a better indication of the relative importance of relevance criteria to the subjects across

different search stages.

Proposed Research Design

The research will be conducted in a naturalistic environment using data collection methods which allow access and usage via the Internet. Subjects will be working on a real information problem: collecting information to write a graded presentation as part of a college class. The following section provides details on the methods to be employed in the main study. In this section, each research question is addressed using the abbreviation RQN where, N is one of the five research questions identified previously.

Approximately 80 subjects will be drawn from a convenience sample of junior and senior business students at an American university. Subjects will be students in a business class, and will be assigned a research project as part of a class assignment. Their progress on gathering information for their assigned research project will be used to collect data for the main study. Subjects will be allowed to choose a research topic from a list of predetermined research topics. Research topics will be of the same level of difficulty for subjects and the structure and rubric for grading the assignment will be the same. Though the research assignment is a required part of the curriculum for the class, the student's participation as a subject providing data for this study is voluntary.

Subjects will not be experts in the subject area of their research topic and will therefore be required to gather information to successfully complete their assignment. Subjects will be given several weeks to complete their task. Data will be collected anonymously using survey instruments integrated into a Web search engine interface which allows subjects to work in naturalistic environment, and to conduct a series of information searches at their own pace.

Subjects will conduct searches and review documents returned from their searches. Using the Web-based search engine, subjects will examine the documents returned, indicate the relevance for the

documents they examined (to address RQ1). Relevance will one of either relevant, not relevant, partially relevant/not sure about relevance. Subjects will also identify a *stage in the information search process* (to address RQ3) by selecting from a predetermined list of search stages, and will indicate the criteria used to make that relevance judgment by selecting from a predetermined list of relevance criteria. Subjects will assign a weight which indicates how important each selected criteria is to them at that point in their search process. All document reviews conducted by the subjects will be stored in a database records for later review. Data collected from the subject's review of documents will include the date the subject conducted their document review, the relevance judgment made, the criteria used to make that relevance judgment, the weight of the criteria (how important it was to their decision), and the subject's stage in the information search process (as selected from a pre-determined list).

The subject's progress for the assignment will be monitored and a project deliverable will be due each week. These will be identified as the *stage in task completion*, with the project outline and abstract due the first week, a detailed outline due the second week, a rough draft of the presentation slides due the third week, and the final presentation slides due the fourth week (to address RQ2).

Associating the date of the document review with the project deliverable required during that time period will provide an association of the relevance criteria used and *stage in task completion*. Associating the subject's selection of stage in the information search process with relevance criteria used will provide an association of relevance criteria choices and user-selected *stage in the information search process*. Similarly, associating relevance judgments made by the subject with relevance criteria choices will allow for an empirical examination of these relationships.

Initial analysis of the data will examine the associations between aggregated results of subject selections of specific criterion for relevance and their progress through the search process (to address RQ5). Additional analysis will examine potential associations between criterion for relevance and relevance choices, specifically partial relevance choices. Analysis will also involve examination of

groupings of criteria and their relationship to progress through the information search process.

Descriptive statistics and other statistical methods will be used to perform the analysis.

Experimental Apparatus

Data will be collected through a Web site which will contain detailed instructions for the

subjects on the use of the site and how to provide information about their searches and relevance

judgments. The site will be accessible from the Internet with any Web browser, thus allowing subjects

to perform the research without intrusive monitoring. Should the subjects require additional help on

using the web site, Web pages will be provided to explain the search stage choices (see Appendix E),

relevance judgment choices (see Appendix D), and criteria for relevance judgment choices (see

Appendix F) to be used by the subjects. Participants will also be provided instruction on using the

research Web site and allowed to practice using the site as part of a tutorial session.

The operation of the site will be similar to the use of a commercial Web search engine such as

Yahoo! with the addition of inputs for search stage, relevance judgments and relevance criteria. The

research Web site will capture data by subject (using an anonymous user ID) for choices of the search

stage (as a category code), the document selected (as a URL and associated ID code) and the relevance

judgment (as a category code).

Search Process

The search stage model (see Table 9 and 10) will be identified by the search phases suggested

by Wilson (1999) which combine the information behaviors of Ellis (1997) with the search process of

Kuhlthau (1993). The terms presented to the subject are shown in "Description Displayed to Subject"

column in Table 3. The term "initiation" is selected because it provides a clearer term than "beginning"

or "starting." The term "browsing" is avoided in this study since this term is now commonly used for all

Web-related searching. Instead, the term "exploration" is used to describe the process of scanning and

gathering information. The remaining terms, "differentiating," "extracting" and "verifying" are from Ellis (1997) and are used in lieu of Kuhlthau's terminology because they provide a more finely grained description of the subject's progression through the search process.

Table 9 - Search Stage Model from Wilson (1999)

Ellis's (1997) Information Behavior	Kuhlthau's (1993) Search Stage	Description Displayed to Subject
starting	initiation	beginning the search process; an initial search
browsing/chaining/monitoring	selection/exploration	browsing, scanning for information
differentiating	selection/exploration	choosing between different areas of focus
extracting	formulation/re-formulation	extracting information to answer the question
verifying	formulation/re-formulation	verifying information that has been gathered previously
ending	presentation	ending the search process

Table 10 - Search Stage Choices Presented to Subject

Search Stage Choice	Description Displayed to Subject
initiation	initial search; start of search process
exploration	scanning for information
differentiating	choosing between different areas of focus
extracting	extracting information
verifying	verifying information that has been gathered previously
ending	completing the search process

Table 11 contains the relevance criteria identified by Barry (1994), Barry and Schamber (1998), and Cool et al. (1993). A number of studies have identified these the criteria and have provided some confirmation as to their consistency across IR tasks (Xu & Chen, 2006; Park, 1993; Schamber, 1994; Schamber & Bateman, 1996). These criteria will be presented to the subjects not as specific criteria, but using the contents of the "description" column in Table 4 (see Appendix C). Subjects will be allowed to choose one or more criterion which they consider appropriate. Subjects will also assign a level of significance to each criterion they select (not currently shown in Appendix C).

41

The criteria selected for this research are a subset of those identified in Table 3. The reasons for these selections are as follows. The 'source' column in Table 3 identifies the source of the relevance criteria: Barry (1994), or Schamber (1994), both Barry and Schamber (1998), or Cool et al. (1993). The table also identifies whether or not the criteria relates to the 'document' or the subject's 'situation.' For this research, only those criteria relating to the document will be used.

All criteria identified as being from both Barry, and Barry and Schamber will be used in this study. All criteria identified in the table as only from Barry's research will be used. Excluded from this study are those criteria which are specific only to Schamber's study and relate to document qualities which were peculiar to her topic (weather reports) and will not apply to technical topics.

Cool et al. (1993) examined a study of students writing a research paper on a technology topic, a task very similar to the one proposed for this study. For this reason, the more detailed, specific document-related criteria for that study are proposed for this study (criteria are identified as "values -- dimensions of judgment," p. 3). Where these criteria duplicate criteria identified in other studies, they have been eliminated.

Table 11 - Relevance Criteria

Criteria	Type	Source*	Used	Description Displayed for Subject
depth/scope/specificity	document	Barry, Cool	Yes	document contains good depth on the topic
accuracy/validity	document	Barry, Cool	Yes	document appears to be accurate
currency	document	Barry, Cool	Yes	information is current, recent, up-to-date
tangibility	document	Barry, Cool	Yes	information relates to real, tangible issues; not esoteric or theoretical
quality of sources	document	Barry, Cool	Yes	source is reputable, trusted, considered expert
accessibility	situation	Barry, Cool	Yes	the effort required to access the information; assumes some cost or effort is involved

Criteria	Type	Source*	Used	Description Displayed for Subject
availability of information	situation	Barry, Cool	Yes	the extent to which the information is available
verification	document	Barry, Cool	Yes	the information is consistent with the body of knowledge the field; the information supports the user's point of view
affectiveness	document	Barry, Cool	Yes	the user's emotional response to the information; pleasure, enjoyment, entertainment
amount of information	document	Cool	Yes	document provides sufficient information
depth	document	Cool	Yes	document covers the topic in good depth
effectiveness of proposed approach	document	Barry	Yes	how effective is the approach proposed
consensus within the field	document	Barry	Yes	how much consensus there is in the field for what is proposed in the document
time constraints	situation	Barry	Yes	how much time is allowed for the task to be completed
background/ experience/ ability to understand	situation	Barry	Yes	expression of concern over the ability to understand a document . (same as 'understandability')
novelty/content novelty/source novelty	document	Barry	Yes	the source or content of the document is new to the subject
geographic proximity	document	Schamber	No	refers to weather information in a geographic location
dynamism	document	Schamber	No	refers to the ability to dynamic manipulate the information in a document
presentation quality	document	Schamber	No	indication that the source of the information could be manipulated in some way
structure	document	Cool	Yes	the structure of the document; how the information is presented/organized
timeliness (age of document)	document	Cool	Yes	is the time frame of the document appropriate; (current where recent information is required; written in a certain time period for historical significance)
understandability	document	Cool	Yes	the document is understandable by the subject (ability to understand)
guidelines	document	Cool	Yes	provides basic direction and structure
ideas	document	Cool	Yes	provides basic ideas and thoughts

Criteria	Type	Source*	Used	Description Displayed for Subject
tips	document	Cool	Yes	provides basic advice and instructions
definitions	document	Cool	Yes	provides basic and/or advanced definitions
connections	document	Cool	Yes	provides links for related topics and subtopics
survey	document	Cool	Yes	provides a good high level overview
history	document	Cool	Yes	provides a good history and background
level of detail	document	Cool	No	provides good depth (similar to scope/depth)
descriptions	document	Cool	Yes	provides explanations and adds clarity
precision	document	Cool	No	the document is written with precision (similar to clarity)
bias	document	Cool	Yes	the document is written with a particular viewpoint
specificity (to topic)	document	Cool	Yes	specific to the topic (topicality, on topic)
authority	document	Cool	Yes	the author or publication has a good reputation in this field

* from Barry (1994, p. 154), Barry and Schamber (1998, p. 226), and Cool et al. (1993, p. 3)

This full list of criteria may be too large for subjects to review. A long list may lead to survey exhaustion, considering that these criteria must be entered each time a document/web page is reviewed. The list could potentially be culled based on the frequency of selection for the criteria revealed in previous experiments.

Research Process

Each subject will perform searches at their own convenience and at their own pace within the parameters of the deliverable due dates. Specifically, the process of searching for information and reporting on results of the search as part of this research study will be as follows.

● Subjects will be asked to sign an informed consent form which explains the purpose of the research and that the information they provide will be treated anonymously

● Subjects will log into a research Web site to conduct their search using a login ID previously assigned, and a personal password they have chosen.

- Subjects will then enter search criteria as if they were using a commercial search engine such as Yahoo! and will execute a search (see Appendix A). (The search results will be returned by the Yahoo! search engine and reformatted to allow user selection of relevance criteria, search stage, and criteria used to judge relevance.)

- The research Web site will generate a search results page with a list of results returned by a commercial search engine for the search query the user entered. For each result returned, the search results page will include an explanation of the result page/document (as returned by the commercial search engine), links to the results page, and links which the user can use to enter relevance information about the page (see Appendix B).

- Subjects will be asked to enter a relevance judgment for the document, a search stage which identifies where the subject was in their search process when they made the judgment, and criteria which were considered by the subject in making that judgment, and the significance of that criteria to their relevance judgment (see Appendix C). Relevance assessments will be one of either *relevant*, *not relevant*, or *partially relevant/ not sure* (from Greisdorf, 2003; Spink et al., 1998). Criteria choices available will be those identified in Table 3. To address issues of order effects, these choices will be randomized in the list of relevance criteria choices displayed to the subjects.

- When the subjects have finished providing information for the documents reviewed on the results page, they will submit the information they have entered to the data collection program on the research Web site which will store the results anonymously for later analysis.

- Subjects may repeat the process outlined above as often as they feel necessary and whenever they wish in order to gather the information they need to complete their assigned report.

- A post-test questionnaire will be used to collect demographic information for reference purposes (Appendix I). This will be associated with a subject, but will be maintained and stored

anonymously (linked to the anonymous subject ID).

Data collected will be stored in a database which will be referenced by an anonymously generated subject ID and a session ID. The database is stored on a secure computer system in a locked room at Rider University. Only faculty has access to the computer system which is under the control of the Computer Information Systems department at that university.

Proposed Analysis

Analysis will involve a combination of descriptive and non-parametric statistical tests using the R system, an open source statistical analysis package similar to SPSS. Frequency counts will be computed for relevance criteria choices, search stage choices and in-group and in-group and cross-group percentages will be computed. Cross-tabs of frequency counts for relevance criteria and search stage choices will be generated to evaluate interactions between the two variables. Cross-tabs will also be generated for relevance criteria choices and relevance judgments made. Regression analysis will be conducted to determine appropriate groupings of relevance criteria choices made by subjects.

Search progress will be analyzed from two perspectives: stage in task completion as the deliverable assigned for a specific week, and the stage in the information search process as identified by the subject. A chi-square test will be used to evaluate associations between search progress and the choice of criteria for relevance, between stage in task completion and choice of criteria for relevance, and between relevance judgments made and choice of criteria for relevance. An aggregation of relevance criteria choices by search stage by subject will also be performed and placed in a cross-tab with search progress. This examination will determine any associations between relevance thresholds (Griesdorf, 2003) and search progress as reported by this sample.

Expectations are that analysis will provide additional insights into the dynamic nature of relevance judgments by examining the interactions between relevance judgments, relevance criteria

choices, and search progress and stage in task completion. Findings of interactions would confirm and extend previous findings of interactions (Vakkari, 2000; Taylor et al., 2006; Wang & White, 1999). Identification of new associations between relevance criteria choices and search stage can provide insights into the users' cognitive state as they progress through the search process and changes to that state as a result of interactions with text. Additionally, analysis of relevance criteria choices and relevance assessments may generate additional insights into the examination of regions of relevance judgments, relevance thresholds and relevance criteria choices. Regression analysis will be used to detect groupings of criteria which will be examined to determine if there are any relevance thresholds in relation to groups of criteria and progress through the information search process.

Limitations of Approach

Additional influences on relevance decisions are known to be user's background or knowledge of the subject domain and search task. The convenience sample for this research is drawn from a pool of undergraduate students who are business majors in a business school at an American university. All students were taking the same course and were given the same assignment. These influences are controlled in this study by drawing from a subject pool whose members have similar backgrounds, experiences, and domain knowledge. Though this aspect of the design of the study attempts to control for variations in domain knowledge, there may still be some variations in knowledge among the subject pool. The choice of this convenience sample also limits the generalizability of the results.

Though subjects are allowed to work in a naturalistic setting, the structure of the study is experimental and not naturalistic. Subjects were given a choice of search topics, but the list of topics was imposed and do not necessarily represent an interest for the subject. To some degree, this represents an imposed query (Gross, 2002) and may impact some of the subject's early searches. The time constraints imposed on the students (4-5 weeks) may also limit the amount of research the subjects can perform.

Task is also known to influence relevance judgments and in this research is treated as a constant since all subject's will be instructed to complete the same task, but variations in the complexity of some assigned subject areas, though controlled and managed as part of the course curriculum for the students used as subjects in the sample, is a limitation.

The Yahoo! search engine was used to generate the search results for the modified search engine used in this research. The reason for this choice was the technical ease in working with Yahoo! search results output. It is a limitation of the study that other search engines such as Google may have provided more robust results based on different search algorithms, and may have provided a more familiar environment for some of the subjects.

Bibliography

Barry, C. L., Schamber, L. (1998). Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing and Management, 34* (2-3), 219-236.

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

Barry, C.L. (1998). Document representations and clues to document relevance. *Journal of the American Society for Information Science*, 49, 1293-1303.

Bateman, J. (1998). Changes in relevance criteria: A longitudinal study. *Proceedings of the ASIS Annual Meeting*, October 25-29, Pittsburgh, PA, *35*, 23-32.

Bates, M. (1989)*.* The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review,* 13(5). 407-424.

Belkin, N. (1982). ASK for information retrieval - part I. *Journal of Documentation, 38*, 61-72.

Belkin, N.J. (1990). The cognitive viewpoint in information science. *Journal of Information Science, 16*, 11--15.

Belkin, N.J. (2000). Helping people find what they don't know. *Communications of. ACM, 43*, 58-61.

Bookstein, A. (1979). Relevance. *Journal of the American Society for Information Science*, *30*(5), 269-273.

Borlund, P. (2003). The concept of relevance in IR. Journal of the American Society for *Information Science and Technology*, *54*, 913-925.

Borlund, P. & Ingwersen, P. (1998). Measures of Relative Relevance and Ranked Half-Life: Performance Indicators for Interactive IR. *SIGIR '98: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* August 24-28 1998, Melbourne, Australia, ACM, 324-331.

Cool, C., Belkin, N. J., Frieder, O., & Kantor, P. (1993). Characteristics of texts affecting relevance

judgments. In M. E. Williams (Ed.), *Proceedings of the fourteenth national online meeting* (pp. 77–83). Medford, NJ: Learned Information, Inc.

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

Cosijn, E.&.Ingwersen, P. (2000). Dimensions of Relevance. *Information Processing and Management, 36* (4), 533-550.

Crystal, A., and Greenberg, J. (2006). Relevance Criteria identified by Health Information Users During Web Searches. *Journal of the American Society for Information Science and Technology.57* (10), 1368-1382.

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

Dervin, B. (1983). *An overview of Sense-Making research: Concepts, methods, and results to date*. Paper presented at the annual meeting of the International Communication Association, Dallas, TX.

Dervin, B., & Nilan, M. (1986). Information needs and uses. *Annual Review of Information Science and Technology* (ARIST) (Vol. 21, pp 3-33). White Plains, NY: Knowledge Industry Publications.

Ellis, D. (1997). Modeling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation, 53* (4), 384-403.

Greisdorf, H. (2003). Relevance thresholds: a multi-stage predictive model of how users evaluate information. *Information Processing & Management*, 39 (3),403--423.

Gross, M. (2002). Integrating the Imposed Query into the Evaluation of Reference Service: A Dichotomous Analysis of User Ratings. *Library & Information Science Research*, *24* (3), 251-263.

Gross, M. (2000). The Imposed Query and Information Services for Children. *Journal of Youth Services in Libraries*, *13* (3), 10-17.

Hansen, P. (1999). User interface design for IR interaction: A task-oriented approach. In: Aparc, T. et al. (Eds.) [CoLIS 3]: 191-205.

Harter, S. P. (1992). Psychological Relevance and Information Science. *Journal of the American Society for Information Science,* 43 (9), 602-615.

Hirsh, S.G. (1999). Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science. 50* (14), 1265-1283.

Hjørland, B. (2000). Relevance Research: The missing perspective(s): non-relevance and epistemological relevance. *Journal of the American Society for Information Science,* 2000*, 51,* 209-211.

Ingwersen, P. & Jarvelin, K. (2005). Information retrieval in context: IRiX. ACMSIGIR Forum, ACM Press, New York, NY, USA, Vol. 39, p. 31--39.

Kuhlthau, C. (1993). Seeking Meaning: A process approach to library and information services. Ablex Publishing Company**:** Norwood, NJ.

Kuhlthau, C.C. (1991). Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science, 42* (5), 361-371.

Langley, P., & Rogers, S. (2005). *An extended theory of human problem solving*. *Proceedings of the Twenty-Seventh Annual Meeting of the Cognitive Science Society*. Stresa, Italy.

Li, Y. (2008). Relationships among work tasks, search tasks, and interactive information searching behavior.

Maglaughlin, K.L., & Sonnenwald, D. H. (2002). User perspectives on relevance criteria: A comparison among relevant, partially relevant, and not-relevant Judgments. *Journal of the American Society of Information Science and Technology, 53* (5), 327-342.

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

Park, K., T. (1993). The Nature of Relevance in Information Retrieval: An Empirical Study. *Library Quarterly*, 63, 318-351.

Rees, A. M., and Schultz, D. G. (1967). A Field Experimental Approach to the Study of Relevance Assessments in Relation to Document Searching, 2 vols. Western Reserve University, School of Library Science, Center for Documentation and Communication Research, Cleveland, OH.

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. (2006). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II. in Danuta Nitecki, Eileen Abels (Eds.), *Advances in Librarianship*, Vol. 30., Academic Press.

Sarcevic, T. (1996). Relevance reconsidered. *Proceedings of the Second Conference on Conceptions of Library and Information Science (CoLIS 2)*, Oct. 14-17, pp. 201-218. Copenhagen: Denmark.

Schamber, L. (1991). Users' criteria for evaluation in a multimedia environment. In: Griffiths, Jos,-Marie, ed. *ASIS '91: Proceedings of the American Society for Information Science (ASIS) 54th Annual Meeting, 28*, 126-133.

Schamber, L. & Eisenberg, Michael B & Nilan, Michael S (1990). A re-examination of relevance: toward a dynamic, situational definition, *Information Processing and Management*, 26 (6),755-776.

Schamber, L., & Bateman, J. (1996). User criteria in relevance evaluation: Toward development of a measurement scale. *Proceedings of the 59th Annual Meeting of the American Society for Information Science, 33*, 218-225.

Spink, A., Greisdorf, H. & Bateman, J. (1998). From highly relevant to not relevant: examining

different regions of relevance. *Information Processing and Management, 34* (5), 599-621.

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 and Management* (2-3),
237-256.

Tang, R..S., & Solomon, P. (2001). Use of relevance criteria across stages of document evaluation: On
the complementarity of experimental and naturalistic studies.
Journal of the American Society for Information Science and Technology, 52 (8), 676-685.

Taylor, A. R., Cool, C., Belkin, N. J. & Amadio, W. J. (2007). Relationships between categories of
relevance criteria and stage in task completion. *Information Processing and Management* ,
43,1071-1084

Vakkari, P. (1999). Task complexity, problem structure and information actions: Integrating studies on
information seeking and retrieval. *Information Processing and Management*, *35*, 819-837.

Vakkari, P. (2001). Changes in Search Tactics and Relevance Judgments when Preparing a Research
Proposal A Summary of the Findings of a Longitudinal Study. *Information Retrieval, 4* (3 -
4),295-310.

Vakkari, P. and Hakala (2001). Changes in relevance criteria and problem stages in task performance.
Journal of Documentation 56, (5), 540-562.

Vakkari, P.. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of
Documentation*, 56, 540-562.

Vickery, B.C.. (1959). Subject analysis for information retrieval. In (ed.), *Proceedings of the*

    International Conference on Scientific Information, National Science Foundation.

Wang, P. & Soergel, D. (1998). A Cognitive Model of Document Use during a Research Project. Study

    I. Document Selection. *Journal of the American Society of Information Science and*

    Technology, 49 (2)*,* 115 – 133.

Wang, P.&.White, M. D. (1999). A cognitive model of document use during a research project. Study

    II: Decisions at the reading and writing stages. *Journal of the American Society for Information*

    Science, 50 (2), 98-114.

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

Wilson, T.D. (1984). The cognitive approach to information-seeking behaviour and information use.

    Seminar on the Psychological Aspects of Information Searching, Copenhagen, 14-18

    November, 1983 and published in *Social Science Information Studies*, 4**,** 1984, 197-204

Wilson, T.D. (1999). Models in information behavior research. *Journal of Documentation, 55*, 249-270.

Xu, Y. (2007). Relevance judgment in epistemic and hedonic information sciences. *Journal of the*

    American Society for Information Science and Technology, 58 (2), 179-189.

Xu, Y., & Chen, Z. (2006). Relevance judgment: What do information consumers consider beyond

    topicality? *Journal of the American Society for Information Science and Technology, 57*(7),

    961-973.

Appendix A - Search Engine Interface

Appendix B - Search Results Page

* Note: relevance judgment choices are presented in a drop-down list which presents a mutually exclusive choice of *relevant*, *not relevant*, and *partially relevant/unsure about relevance*.

Appendix C - Search Stage and Criteria for Relevance Selection
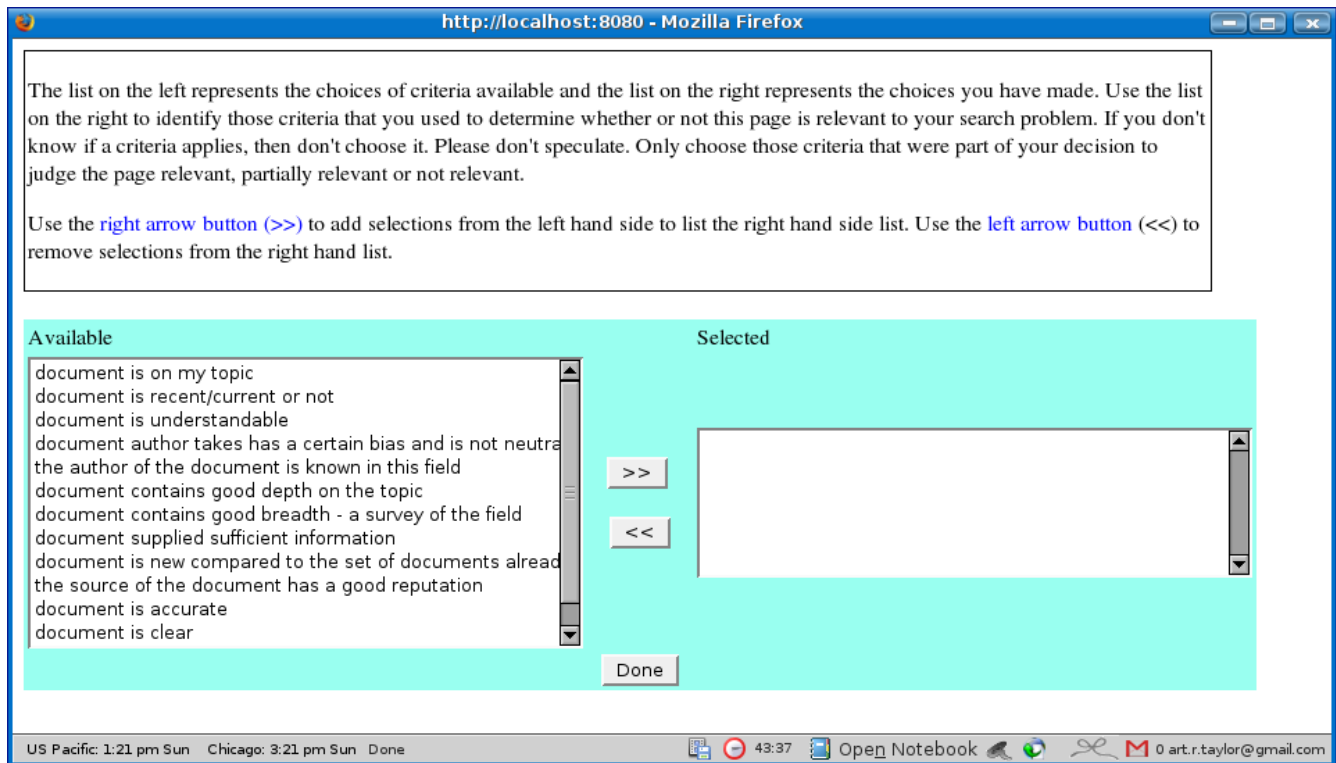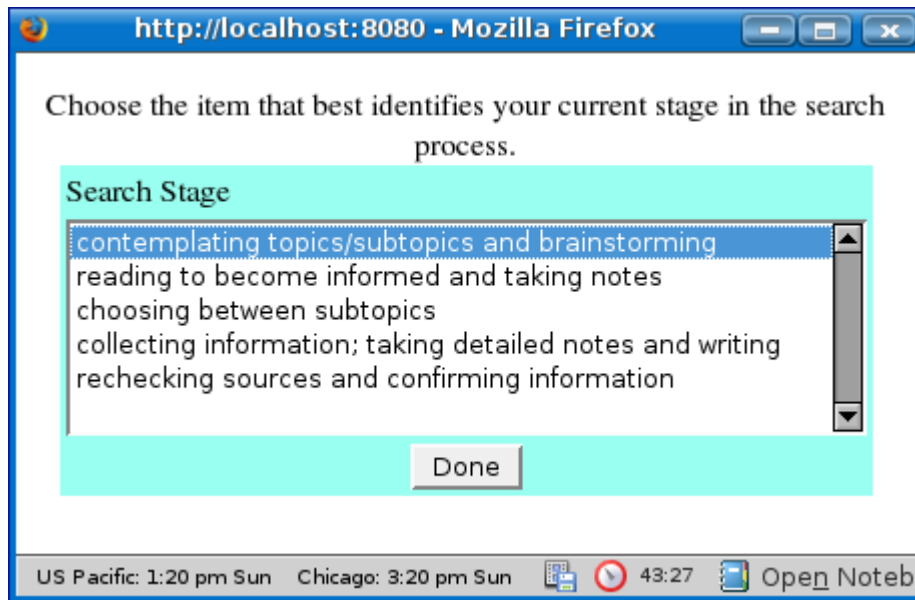
Appendix D - Relevance Judgment Help

US Pacific: 2:06 pm Sun   Chicago: 4:06 pm Sun   Done

Appendix E - Search Stage Help

Appendix F - Criteria for Relevance Judgment Help

http://serenity.rider.edu:8080 - criteria - Mozilla Firefox

criteria

Criteria for Relevance Judgment

These are the criteria you used to judge the page *relevant, partially relevant* or *not relevant*. You should only choose criteria that affected your decision. Don't speculate - if it didn't come to mind as you read the document, then you should not choose that criteria. You should choose relevance criteria for **all** documents judged, this includes relevant, partially relevant and not relevant documents. In the case of not relevant documents, you should choose criteria that led you to reject the page. The following tables contain more detailed descriptions of the criteria provided on the search engine results page.

Criteria for Relevance Judgments

Description	Explanation
document is recent or current	for this field, this is a fairly recent document
document author takes has a certain bias and is not neutral	the author has taken a particular point of view, a specific slant
the author of the document is known in this field	the author of the document has a reputation for being accurate, correct and providing useful information
document is new compared to the set of documents already collected	relative to documents already retrieved, this document provides new information or perspective
document is on my topic	the topic of this document matches or nearly matches the topic of my search problem
document is accurate	the facts in the document appear to be accurate
the source of the document has a good reputation	the source of this document (magazine name, journal, university) is known to be a quality source
the structure of the document is good	the structure of the document helps make the content understandable
document is understandable	the document is understandable to you
document contains good depth on the topic	the depth of coverage within the field is good
document contains good breadth - a survey of the field	the breadth of coverage across subtopics is good; a good survey of the field

US Pacific: 2:07 pm Sun   Chicago: 4:07 pm Sun   Done

1:00:40   Open Notebook   0 art.r.taylor@gmail.com

Appendix G - Research Topics Assigned to Subjects

Computer Security: Making Computer Technology Accessible and Secure
Computer Security: Making Desktop Systems Secure
Computer Security: Preventing Computer Fraud
E-Commerce: After the Internet Bubble
E-Commerce: How to Put Your Company on the Web
Internet Business Models
ERP Systems: The Future
Customer Resource Management (CRM) Systems: Current Status
Does IT Matter: What Role Will IT Take in the Future?
New Technologies: Can Linux be Mainstream ?
New Technologies: the Future of WiFi
Microsoft: Dealing with the 500 Pound Gorilla
Ethics and the Information Age: Is It Really Stealing if It's Digital ?
Distributed Computing
Grid Computing
Group Collaboration with Computers
Computer Aided Design (CAD) Systems
Supply Chain Management with Computers
Privacy and Computers
Decision Support Systems
Implementing Enterprise Resource Planning (ERP) Systems
Alternatives to ERP Systems
The Current State of Artificial Intelligence and Expert Systems
Systems Design and Development
Enterprise Portals and Application Integration
Open Source Software on the Desktop: Current Status
ERP: Implementation Issues

Appendix H - Search System Pilot
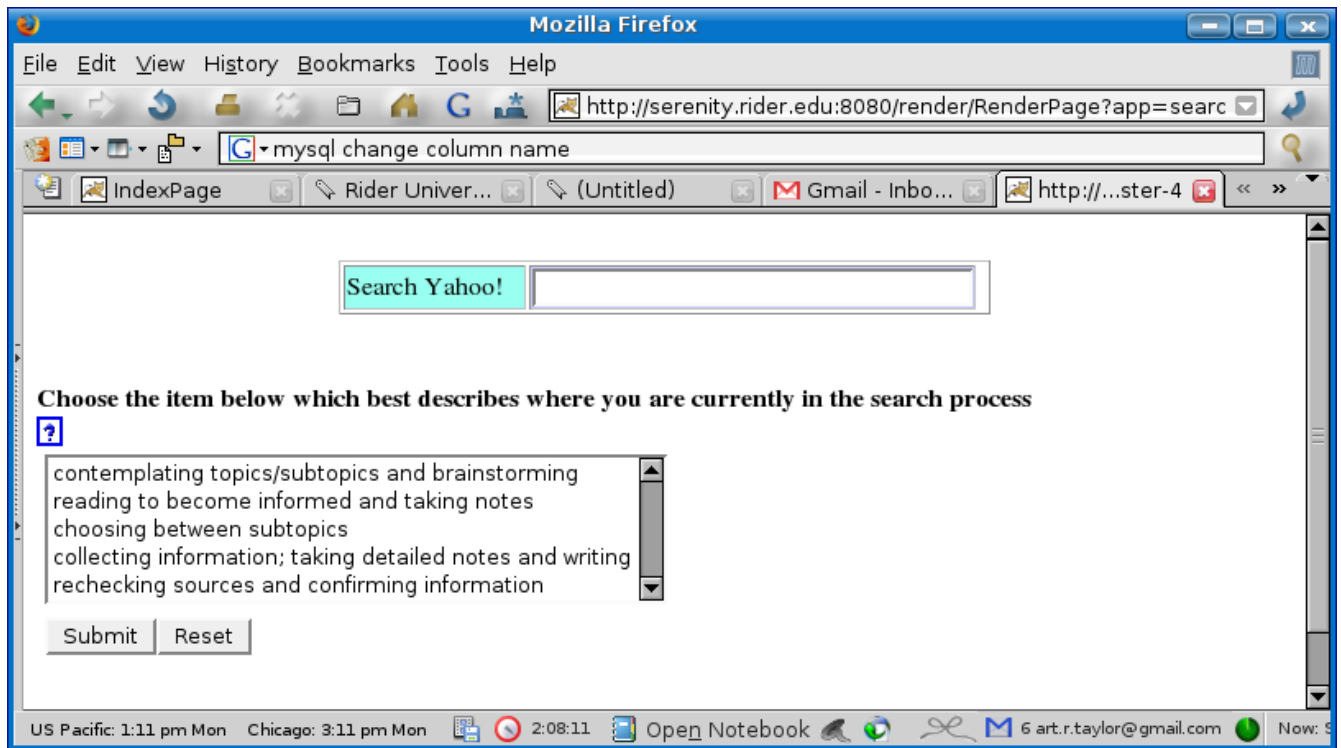
Appendix I - Collection of Demographic Information

Question	Text
1	How frequently do you perform online searches? 1=not very often; 7 = very often)?
2	At this point in time what is your highest level of education? A.) High school, B.) College,, C.) Graduate
3	What is your age? _____
4	What is your gender? M __ F ___
5	Is English your primary language? Y___ N ___