RELEVANCE CRITERION CHOICES IN RELATION TO SEARCH PROGRESS

BY

ARTHUR TAYLOR

A Dissertation 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

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2009

ABSTRACT OF THE DISSERTATION

RELEVANCE CRITERION CHOICES IN RELATION TO SEARCH PROGRESS

BY ARTHUR TAYLOR

Dissertation Director:

Professor Tefko Saracevic

Relevance as an information science concept is at the center of human information behavior. Relevance judgments occur within an information search process, where time, context and situation can impact relevance judgments. The determination of relevance is dependent on a number of factors and variables which include the criteria used to make relevance judgments. Research has shown that relevance judgments are dynamic, varying among user's evaluating the same document, and varying by user for the same document over the course of the information search process. Research has suggested that the criteria used to make these relevance judgments may also be dynamic. Understanding which relevance criteria are chosen and when they are chosen during the information search process can provide important information about the dynamic relevance judgment process and inform the development of information retrieval systems.

The purpose of this exploratory research is to examine the importance of the criteria used by searchers to make relevance judgments over the course of an information

search process. The goal is to determine if users' choices of criteria, and the importance of those criteria change over the course of an information search process.

This research encompasses three separate studies which examined a subject's relevance judgment and the criteria used to make that judgment over the course of an information search process. Subjects were asked to search for information, evaluate documents, and then make relevance judgments for those documents. They were then asked about their relevance judgment, where they were in their search process when they made that judgment, and which criteria were used to make that judgment. Statistical analysis was used to examine these results. Findings include consistent selections of criteria across three distinct studies, and a statistically significant relationship between criteria choices and stage in the information search process. Sets of criteria choices were also examined and statistically significant relationships between sets of criteria and stage in the information search process were also detected.

ACKNOWLEDGEMENTS

The dissertation process is long and arduous and could not be completed without the help and support of numerous individuals.

First, I would like to thank my dissertation committee chair Tefko Saracevic for his help and support. His deep knowledge of the subject area and his keen insights into the research process have been a tremendous asset to me during this process. Most of all I would like to thank him for his commitment to me and my dissertation. His contributions to this effort could not have been greater, and I am privileged to have had him as a dissertation chair.

I would like to express my thanks to my committee members, Marie Radford, Xiangmin Zhang, and Bill Amadio. Their contributions have been significant, and their academic and personal support have been constant. Their participation in this process is deeply appreciated.

Joan Chabrak has been a tremendous help to me throughout my duration at SCILS, helping maintain my focus on the requirements of the program. I am grateful for her encouragement and support..

Last but in no way least, I would like to express my appreciation to my wife Carolyn. She has supported me throughout this effort, believing in me, never doubting, always supportive, and proving to me yet again that I am lucky to have her in my life.

Table of Contents

List of Tables

List of Figures

Chapter 1 - 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. 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 user selection of relevance criteria and how the importance of those selections to the user may change over the course of the information search process (ISP). A dynamic relevance judgment process implies 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 also changes.

1.1 Statement and Significance of Problem

A user with an information need may iterate through multiple information search sessions, retrieving 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 be a required criteria since the document must be on the topic of the search, but other criteria may have increased or decreased importance as the users' subject area

knowledge (cognitive state) changes. Current search engines rarely recognize criteria beyond topic, do not recognize any progression through an information search process, and provide no facility to adjust to changes in the cognitive state of users. Previous information science research has provided little guidance on 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 (Schamber, Eisenberg, & Nilan,1990; Vakkari, 2000; Taylor, Cool, Belkin, & Amadio, 2006; Wang & White, 1999) and would inform the design of information search systems. Improved search system design could extend basic topical search queries with additional criteria, and could adapt information retrieval (IR) processing to the users' cognitive state changes as they progress through the ISP.

The research detailed here was conducted over a period of three years in three distinct studies referred to as Study 1, Study 2, and Study 3. The purpose of the research was to examine the selection of criteria used by subjects to make relevance judgments, and determine whether or not the importance of the criteria used to make relevance judgments changes over the course of the information search process. The importance of criteria selections was evaluated through frequency of criteria selection, and weights assigned by users to relevance judgments and criteria selections.

1.2 Variables and Research Questions

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.

The studies detailed here examined the relationship between relevance judgments, the criteria used to make those judgments, and progress through the information search process (ISP). Subjects were presented with a research problem either as a question which they needed to answer, or as a research assignment. This created an information need which required them to conduct information searches and gather information. As the subjects gathered information, data was captured on the relevance judgments of the subjects, the criteria subjects use to make those judgments, a weight assigned to the relevance criteria (Study 3) or the relevance judgment (Study 1), and the subject's progress through the ISP. The variables examined in these studies were 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;
- the *stage in task completion* which is operationalized as the point in time when subjects were required to produce a project deliverable for a multi-week research project;
- the subjects' *relevance criteria* choices which are operationalized as the subjects' choices of criteria which were critical in making their relevance judgment. Criteria were chosen from a list of predetermined criteria presented to each subject as they evaluated a document. The level of importance for the criteria choice was based on frequency of selection in a search stage, and in Study 1 and Study 3, was additionally based on filtering of data using weights selected by the subject;

- 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. In each study an information problem was assigned to the subject.

Using these variables, the following research questions were examined in these studies.

1. Does the user's choice of some *relevance criterion* change in relation to *relevance judgments*?
2. Does the importance of some *relevance criterion* change in relation to *stage in task completion* as indicated by the frequency of criterion selection, and/or a weight indicating importance as assigned by the user?
3. Does the importance of some *relevance criterion* change in relation to a user-identified *stage in the search process* as indicated by the frequency of criterion selection, and/or a weight indicating importance as assigned by the user?
4. Are there sets of *relevance criteria* choices which change in importance in relation to a user-identified *stage in the search process* as indicated by the frequency of criterion selection, and/or a weight indicating importance as assigned by the user?

1.3 Research Synopsis

The research detailed here involved three studies, one short-term study which subjects completed in a 1-2 hour session, and two longitudinal studies where subjects recorded information about their search over a 4-5 week period. Methods for the studies differed and are explained in detail with the description of the study in the associated chapter. A synopsis of each study follows.

Study 1 involved 39 subjects, a convenience sample of undergraduate students from an American university. Subjects volunteered for the research. Research was conducted over a 1-2 hour session in a computer laboratory on campus. Subjects were assigned a research question and recorded a relevance judgment for each document they considered relevant, the search stage they were in when they made that relevance judgment, and the criteria used to make that relevance judgment. Relevance judgments were recorded on an interval scale, using a value from 1 to 10. This was interpreted as a weight indicating the strength or level of importance for that relevance judgment.

Study 2 involved 82 subjects, a convenience sample of undergraduate students from an American university. Subjects were students in a class where they were assigned a research project to complete in a 4 to 5 week time frame. Research for the project was conducted using a modified search engine where subjects recorded information about the documents/web pages they reviewed. Information was collected using a modified online search engine. Information collected included a relevance judgment (relevant, not relevant, partially relevant), and the criteria they used to make the relevance judgment.

Study 3 involved 53 subjects, a convenience sample of undergraduate students from an American university. Subjects were students in a class where they were assigned a research project to complete in a 4 to 5 week time frame. Subjects used a modified online search engine to perform research for the research project. The search engine recorded information about their search and the documents reviewed. Information reported by the subjects included a relevance judgment (relevant, not relevant, partially relevant) and the criteria they used to make the relevance judgment. Subjects also provided a weight indicating the importance of the each relevance criterion they chose.

Chapter 2 - Literature Review

2.1 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, 2007a, 2007b) 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 research into the criteria user's identify when making their relevance judgment. These clues represent artifacts of the search process, and the criteria used by subjects are the properties which describe these clues. The importance of these criteria changes with task, progress in task over time, and varies by some categorization or class of user. The author emphasizes that searchers use the same criteria, but assign different weights to these criteria (Saracevic, 2007b).

Cosijn and Ingwersen (2000) built on the work of Saracevic (1996) and others to develop a revised table of attributes and manifestations of relevance (Cosijn 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).

Cosijn 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. Cosijn 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. 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.

2.2 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.

Schamber (1991) conducted similar relevance criteria research with 30 users in three different occupational fields. 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 (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.

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.

Barry and Schamber (1998) later combined the data collected from the Barry (1994) study with Schamber (1991). 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 2.1 which represent a cross-section of the attributes and manifestations of relevance as identified by Cosijn 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 2.1 - 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

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.

Table 2.2 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).

Table: 2.2 Criteria Identified in Previous Research

Criteria	**Type**	**Source***	**Used**	**Description Displayed for Subject**
depth/scope/specificity	document	Barry, Cool	Yes	document contains good depth, good coverage of 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

Criteria	Type	Source*	Used	Description Displayed for Subject
				involved
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 (see depth/scope/specificity)
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

Criteria	Type	Source*	Used	Description Displayed for Subject
				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
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

Criteria	Type	Source*	Used	Description Displayed for Subject
specificity (to topic)	document	Cool	Yes	specific to the topic (topicality, on topic; also depth/scope/specificity)
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)

2.3 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.

2.4 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, 1993, 2004). 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 was 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 study which examined 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) performed a study in which 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").

Table 2.3 - Methodological Comparison of Studies

Study	Criteria Choices and ISP Stage *	Relevance Judgment Criteria	Statistics Used	Criteria Sets Used**	ISP Stages	Collection of Criteria for Relevance
Cuadra and Katter (1967)	No	Range	Descriptive and parametric	No	n/a	questionnaire
Rees and Schultz (1967)	No	Range	Descriptive	No	3	Simulated research project; questionaire
Bateman (1998)	No	High relevance only	Descriptive	No	6	content analysis of user comments
Tang and Solomon (1998)	Yes	Relevant only	Descriptive	No	2	subject marked text in document
Tang and Solomon (2001)	Yes	Relevant only	Descriptive	No	2	specific criteria rated on form
Wang and White (1999)	No	Relevant and Not relevant	Descriptive	No	3	structured interviews with subjects; content analysis
Vakkari (2000); Vakkari and Hakala (2000)	Yes	Relevant, Partially relevant, not relevant	Descriptive	No	6	semi-structured interview and questionnaire ; content analysis
Taylor et al (2007)	Yes	Relevant, Not relevant,	Descriptive and parametric	No	5	content analysis of user

Study	Criteria Choices and ISP Stage *	Relevance Judgment Criteria	Statistics Used	Criteria Sets Used**	ISP Stages	Collection of Criteria for Relevance
		Can't Tell				comments

* were search stage interactions with criteria choices examined
** were search stage interactions examined with groups of criteria

2.5 Task and Situation Influences on Information Seeking Behavior

A number of factors or intervening variables have been demonstrated to have an impact on information seeking behavior, and the relevance judgment process. 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. In the studies conducted as part of this research, all subjects were assigned consistent, similar work task. Since all subjects were performing the same work task, the influence of the work task was controlled with the subject pools used. Work tasks do, however, vary and any attempts to generalize the results of this study must consider differences in the work task.

2.6 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), the research did not examine selection of criteria for relevance judgment in relation to progress through the ISP. With the exception of Tang & Solomon (2001) and Taylor et al. (2007), sample sizes were too small to allow for stronger statistical analysis beyond descriptive statistics. Though some of these studies suggest that subjects use more than one criterion to make their relevance judgments, none specifically examined or analyzed sets of criterion 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).

Examination of the criteria used by subjects, and the importance of those criteria to subjects in relation to search stage progress requires a consistent and broad set of criteria choices, and a consistent search process model. With the exception of Taylor et al (2007), the studies referenced here used a small set of criteria choices and a limited number of search stages. The use of small samples, and the use of qualitative analysis, or limited descriptive statistic analysis reduces the generalizability of the findings. The

findings reported by these studies, while useful, still leave uncertainties about the use of relevance. Questions persist about which relevance criteria are chosen and when, and the importance of the criteria used, and whether or not the importance of those criteria change over the course of the information search process. A deeper understanding of the dynamic nature of the relevance judgment process requires an understanding of the interactions between relevance criteria choices and the information search process.

2.7 Justification for Research

To advance understanding of the dynamic relevance judgment process, this study examined these interactions during the relevance judgment process directly, and used stronger statistical methods to provide additional clarity and depth. This research examined 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 added to the statistical strength of this study. Subjects were 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, this research performed statistical analysis to determine whether subjects consistently used groups of criteria 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 examined a number of these aspects of relevance dynamics in a single study, allowing various interactions to be examined with a single subject pool. This examined groups of relevance criteria selections in relation to the ISP stage, a relationship which had not been examined. These studies examined relevance criteria interactions in depth, using a more detailed level of

criteria choices (examining more criteria) and used a sufficiently large sample size to provide additional statistical strength to the results.

Chapter 3 - Research Model and Methodological Framework

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.

3.1 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 information retrieval (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 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

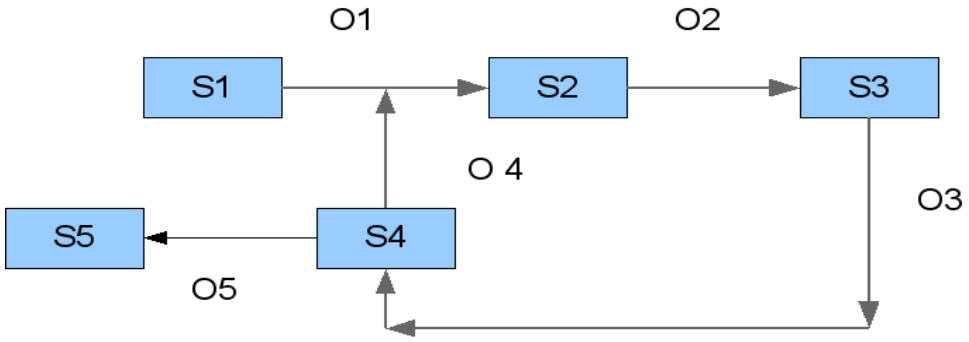
search episodes. Over time (a series of search episodes) the subject's cognitive state changes as the subject's ASK changes. These models of information need and their concepts have been studied and refined (Schamber et al. 1990; Bruce, 1994; Barry, 1994), in turn laying the groundwork for elements of other theories such as polyrepresentation (Ingwersen, 1996), and more recent multiple search session models (Spink, Wilson, Ford, Foster, and Ellis, 2002; Lin and Belkin, 2004).

3.2 Research 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 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 a manifestation of 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.

Figure 3.1 presents this model using a knowledge state diagram to identify the states specific to a single information search episode. In this diagram, the character 'O' represents an *operation* in the process, and the character 'S' represents a cognitive *state*

during the 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 as documents or document representations, and the subject absorbs the results of the search either by reading or skimming the documents and/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 absorption of information, the user's knowledge of the subject area will 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 – read document/ document representation and absorb information
O3 – judge relevance using various criteria
O4 – reformulate search
O5 – complete search

Figure 3.1: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 in operation O3. Following this relevance judgment, 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.

Figure 3.2 shows the subset of operations from the model in Figure 3.1 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 the subject evaluates the document. This model represents as single iteration within a larger information gathering process, which will likely require multiple iterations to complete. These multiple iterations themselves exist within a work task, with multiple iterations of search episodes as shown in Figure 3.2 taking place as part of this work task.

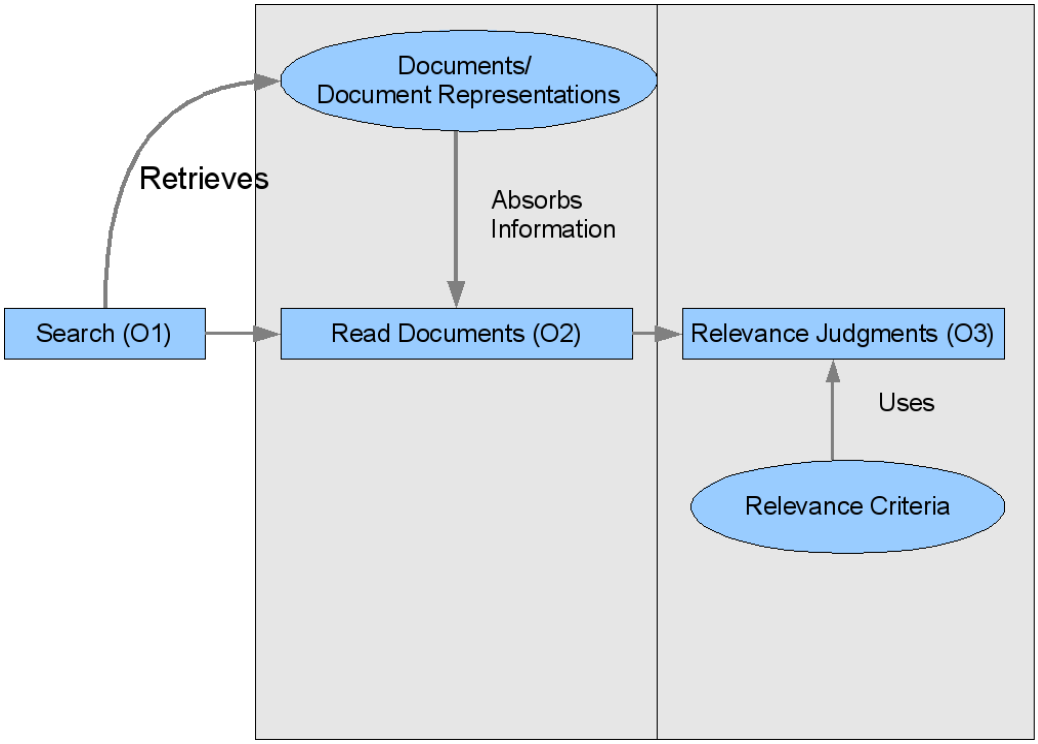


Figure 3.2: Search Episode within the Information Search Process

As this model indicates, the state of the user's cognitive model 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 Figure 3.1). 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 change. The goal of this research is to examine these criteria choices in relation to progress through the information search process. These criteria choices depend on a number of variables including *topic* and *work task*, and the individual's background as reflected in the user'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 with similar knowledge, background, and search skills.

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 longer time frame since the longer time frame would allow more time (more cycles) for the individual to absorb the information being reviewed. Study 2 and Study 3 were longitudinal studies examining relevance judgments over a 4-5 week time period.

3.3 Relevance Criteria

A number of previous studies have identified criteria used to judge relevance. Table 3.1 contains the relevance criteria identified by Barry (1994), Barry and Schamber (1998), and Cool et al. (1993). These studies have identified criteria and have provided some confirmation as to their consistency across IR tasks (Xu & Chen, 2006; Park, 1993; Schamber, 1994; Schamber & Bateman, 1996).

The 'source' column in Table 3.1 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.' The criteria identified as 'situation' represent intervening variables

for the relevance judgment process which the research methods in these studies have sought to control. These criteria were not examined in these studies.

Table 3.1 - Relevance Criteria Identified in Previous Studies

Criteria	Type	Source*	Used	Description
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
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 sufficient detail (similar to depth/scope)
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

Criteria	Type	Source*	Used	Description
				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
tips	document	Cool	Yes	provides basic advice and instructions

Criteria	Type	Source*	Used	Description
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)

The criteria list provided to the subjects was a subset of the 'document' characteristic criteria in Table 3.1. The reasons for reducing this list are as follows. Since subjects were evaluating this list for each document they reported, it could quickly become tedious to review a long list of criteria and survey exhaustion could result. A number of criteria in the list reflect similar document characteristics and are thus redundant. Those criteria specific only to Schamber's study relate to document qualities specific to the search topic she used (weather reports) and did not apply to the topics assigned to subjects in these studies. Many of the criteria identified in Cool's study had similarities to those identified in Barry and Schamber's work as identified in the table, and in the interest of being concise and keeping the list of criteria presented to the

subjects short, they were excluded from the list presented to the subjects. Specific lists of criteria used in each study are presented in the methods section of the chapter in which the study is discussed.

3.4 Search Stage

Various search stage models have been proposed and studied. The ISP stage references used in these studies were a combination of Kuhlthau's (1993) ISP stages and Ellis's (1997) search patterns. Wilson's (1999) study which combined the information behaviors of Ellis (1997) with the search process of Kuhlthau (1993) informed the process.

For Study 1, the ISP stages of Kuhlthau were merged with the search patterns of Ellis. Table 3.2 presents the search stages used in Study 1. In this search stage model, the first two ISP stages of Kuhlthau ('task initiation' and 'topic selection') were not used since the search task and topic were selected for the subjects and presented as a list of possible topics. Additionally, the ISP stages developed by Kuhlthau were based on research conducted using mediated library searches. The process of online searching as conducted in this study did not involve the use of mediators and involved activities that were not covered in detail by Kuhlthau's studies. To address this issue, Ellis patterns of 'browsing' and 'extracting' were used to provide additional depth and are search behaviors more amenable to the online searching.

Table 3.2- Merging of ISP Stages - Study 1

ISP Stages	**ISP Stages***	**Search Patterns****
	1 - Initiation	
	2 - Selection	
Becoming informed on the topic.	3 prefocus exploration	
Learning about the topic.	3 prefocus exploration	
Trying to *focus* on the topic/subtopic.	4 - focus formulation	

Defining and *extending* focus.	5 - information collection - supporting focus	
Browsing for information on the focus I've identified.		Browsing
Extracting useful information.		Extracting
Verifying information retrieved.	6 - search closure/presentation	
Completion and presentation of information.	6 - search closure/presentation	Ending

* Kuhlthau (1993); ** Ellis (1997)

This model was revised for Study 2 and Study 3. Building from the consolidating work of Wilson (1999), the synthesis presented in Table 3.3 was used to develop the model in Table 3.4. The 'ending' search stage was not used since the completion of the assigned research project by subjects was considered an implied ending or completion of the search process.

Table 3.3 - Revised 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 3.4 - Search Stage Model - Study 2 and Study 3

Search Stage Term	**Description**	**Behaviors***
initiation	initial search; start of search process	exploring and thinking about the topic/subtopics and information needed
exploration	scanning for information	reading to learn about the topic/subtopics

Search Stage Term	Description	Behaviors*
differentiating	choosing between different areas of focus	identifying areas of interest and focus; choosing one or more subtopics
extracting	extracting information	collecting information; taking detailed notes and preparing (writing) the presentation
verifying	verifying information that has been gathered previously	verifying information I have gathered; checking and evaluating sources (web pages) used
ending	completing the search process	

- from Kuhlthau (1993) and Wilson (1999)

3.5 General Experimental Design

To examine the impact of the progression through the information search process on the relevance judgment process, specifically the criteria used to make relevance judgments, three distinct studies were conducted. The studies were conducted in sequence over a period of three years. The results of each study informed the research design of the next, resulting in various design and procedure changes as they studies progressed. Table 3.5 details the studies and the research questions they addressed.

Table 3.5: General Research Design

Study	Research Questions	General Design
1	1,3	Short-duration search session, college undergraduates, monitored in a lab, single question assigned to subjects, subjects selected search stage and criteria
2	1,2,3,4	longitudinal study, college undergraduates, naturalistic environment, subjects assigned different topics for a research assignment
3	1,2,3,4	longitudinal study, college undergraduates, naturalistic environment, subjects assigned different topics for a research assignment, subjects assigned weights of importance to relevance criteria chosen

All studies sought to examine the selection of relevance criteria in relation to time. To analyze the data collected, and to add relevance to the projects in general, a search stage model as described in this chapter was used with minor modifications in each study. Data collection involved asking the subject (through online surveys) which search stage they were in, what their relevance judgment was, and which criteria were important in making that relevance judgment. Analysis involved preparation of descriptive statistics, analysis of variance, and then a cross-study analysis to examine results across studies.

Study 1 addressed research question 1 and 3. This study was conducted over a short period of time (1 to 2 hours) in a computer laboratory on a college campus. Subjects were undergraduate students who were assigned a question which they needed to answer by searching for information.. Since the research was a short duration study, there was no meaningful method for identifying a stage in task completion.

Results of Study 1 and concern that a short search session was not long enough to capture shifts in the subject's preference for relevance criteria led to the development of a multi-week assignment for students. This longer duration assignment was used in Study 2 and Study 3. Study 2 and Study 3 also sought to allow the subject to work in a naturalistic environment by using a Web site accessible from the Internet for data collection. Study 2 and Study 3 addressed research questions 1,2,3 and 4. Study 3 also had subjects assign a weight of importance to the criteria being selected.

Chapter 4 - Research Study 1

The first study conducted examined relevance criteria selections in relation to progress through the information search process. The study addressed research questions 1 and 3. The study used a convenience sample of undergraduate students at an American university, examined a short information search process (1 to 2 hours), and involved monitoring subjects in a laboratory setting. The methods used in this study and evaluation of the results laid the groundwork for Study 2 and 3. The methods are explained in more detail in the following section.

4.1 Methods

Subjects for Study 1 were a convenience sample from a pool of undergraduate students at an American university. Subjects volunteered for the research. All subjects who volunteered were directed to report to a computer laboratory at an assigned time to participate in the research. Once in the laboratory, subjects were assigned to a computer workstation and given the research question shown below. All subjects were assigned the same research question as shown below.

Question Assigned to Subjects:

Consider that you have been assigned the following question as part of an open book, open Internet exam. Conduct a search for documents which you would find useful in answering these questions. Attempt to find at least ten documents which you would find useful.

> *Compare and contrast the benefits of using a fixed exchange rate versus a flexible exchange rate for international transactions.*

Subjects worked alone to gather information to answer the research question. Subjects used an online information search service that allowed them to find journal articles on their assigned subject. Subjects could review document representations (descriptions of the journal article), or could review the actual journal article.

Subjects were told to record document relevance, search stage, and criteria used to judge document relevance for each document reviewed. Subjects recorded this information about their search using online data collection instruments as detailed in the "Data Collection" section below.

4.1.1 Search Process

Subjects selected a search stage from a list of search stages available. The list presented to subjects for selection in this study is a synthesis of several published and validated information search stage models as detailed in chapter 3, Table 3.2. The list presented to subjects in Study 1 is shown in Table 4.1.

Table 4.1: Search Stage - Study 1

Search Stage	**Description Presented to Subjects**
becoming informed	Becoming informed on the topic.
learning	Learning about the topic.
focusing	Trying to focus on the topic/subtopic.
defining	Defining and extending focus.
browsing	Browsing for information on the focus identified.
extracting	Extracting useful information.
verifying	Verifying information retrieved.
presentation	Completion and presentation of information.

4.1.2 Relevance Criteria

Subjects indicated document relevance on a scale of 1 to 10, with 1 being least relevant and 10 being most relevant (see Appendix N). Subjects selected relevance criteria using the list presented in Table 4.2. This list was based on the model presented in chapter 3, Table 3.1. The number of criteria was reduced for clarity and to avoid survey exhaustion on the part of the subjects. The list presented to subjects in Study 1 is shown in Table 4.2. (The criteria of "depth" and "scope" were combined in Study 1, but were split into separate criteria in Study 2 and Study 3.)

Table 4.2: Relevance Criteria for Study 1

Criteria
Amount of information
Specificity
Clarity of presentation
Ability to Understand
Depth/Scope
Precision of Document
Recency of Document Publication
Interest in Topic
Instructional
Authority of Author
Bias of Author

4.1.3 Data Collection

A convenience sample of 39 subjects was drawn from the student population of undergraduate students at an American university. The subject pool was a mix of approximately 20 percent paid subjects and 80 percent unpaid subjects. Subjects were 42% female, 58% male.

In the experiment, subjects were asked to retrieve at least 10 documents to solve an assigned research questions. Subjects were instructed to retrieve and report on

documents they thought would be useful to answer their assigned question. The relevance assessment questions, search stage questions and search question used in the study are listed in Appendix N, O and P respectively. For each document selected, subjects recorded the relevance criteria which they used to assess the document, the search stage they were in when they evaluated the document, and whether or not they considered the document relevant to answering the search question they had been assigned. Relevance was captured on an interval scale from 1 to 10, with one being the least relevant and 10 being the most relevant.

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. A pre-test survey asked the subject questions about their experience with search engines, and a post-test survey asked the subject about their satisfaction with the search process. The data collected from these two surveys was not used in the analysis performed for Study 1.

All subjects were assigned the same search task as shown previously, and performed the task while being monitored in a computer lab at the university where the study was conducted. Monitors in the computer lab did not interfere with the subjects as they performed their search and recorded information. Subjects worked alone on their search problem and took between 45 minutes and two hours to complete their research session.

Subjects searched for information using the ABI/Inform online database to find documents to help them answer their assigned question. The ABI/Inform database indexes a combination of trade journals and scholarly journals using a search engine

interface similar to Google. If they chose to, subjects could print the documents and review them during their session. After reviewing a set of ten documents, subjects were asked to complete a questionnaire on each document. In that questionnaire they identified document relevance, where they were in the search process, and the criteria which were useful in making that relevance judgment. All questionnaires were completed online by the subjects during the research session.

4.2 Results of Data Collection

The data collected from the research sessions consisted of 383 records each representing a relevant document that a subject used in his/her search process, and corresponding relevance criteria and ISP stage selections. Documents could be used in more than one stage of the search process, and some subjects did report on the same document in more than one stage. When each relevance judgment was joined with a search stage, the 383 rows of relevance judgments generated 795 separate *judgment-stage* reports. Subjects were not required to report particular ISP stages and were not required to report ISP stages in a specific order. An examination of the percentage of relevance criteria category selections as a percentage of total selections in Table 4.3 indicates no particular criteria dominated the users' selection of criteria. The results shown in Table 4.4 shows that with the exception of the "presentation" search stage, search stages were selected somewhat consistently, with subjects showing a slight preference for the "focusing," "browsing," and "extracting" search stage over other search stages. Because of the low selection counts for the "presentation" stage, and the difficulty of using some statistical analysis methods with low counts, the results for the selection of the "presentation" search stage were not used in the analysis.

Table 4.3 - Relevance Criteria Selections by Users as Percent of Total

Criteria	Pct
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	99.99%

Table 4.4 - Search Stage Counts*

Search Stage	Count	Percent
becoming informed	87	11.05%
learning	103	13.09%
focusing	126	16.01%
defining	80	10.17%
browsing	126	16.01%
extracting	135	17.15%
verifying	83	10.55%
completion	47	5.97%
Total		100.00%

* Percentage of all relevance judgments reported by subjects

4.2.1 Statistical Analysis of Cross-Tabulations

The following section contains cross tabulations of relevance criteria choices and information search process stage. An analysis of variance test was used to examine the values in these tables to determine if a relationship exists between the column values (the categories) and the values in the rows (the cases). These statistical tests determine the probability that the variables represented in the row and column are related, with the value of 'p' representing the probability that the relationship is due to chance and not a valid effect. A lower 'p' value indicates that the relationship between the two values is more likely due to a relationship. When cell values are low, the statistical strength of

analysis of variance tests is reduced. With Chi Square tests a Yates Correction can be applied to improve the validity of results with low cell values in the results table. The Yates Correction was applied to the calculations performed in this section to increase the validity of the results when the frequency counts in cells was low.

4.2.2 Relevance Judgment Groups

The research methods for Study 1 had subjects indicate the degree of relevance when they made a relevance judgment. The degree of relevance was selected as a value between 1 and 10, with 1 being the least relevant and 10 being the most relevant. As part of data filtering process, these scores were folded into three distinctive bands of relevance: *low relevance* interpreted as a value of 3 or less assigned to the degree of relevance weight; *high relevance* interpreted as a value of 8 or greater assigned to the degree of relevance weight; and *partial relevance* was interpreted as a value between 4 and 7 assigned to the degree of relevance weight.

4.2.3 Partial Relevance

Analysis involved looking for statistically significant variance in the frequency of selection across search stages. Table 4.5 shows frequency counts for partial (a value between 4 and 7) relevance judgments by search stage in relation to relevance criteria used. Examination of these results using Chi Square did not reveal statistically significant relationships.

Table 4.5 : Frequency Counts for Partial Relevance Selections

Criteria	becoming	browsing	defining	extracting	focusing	learning	verifying
understand	49	70	49	63	75	66	43
amt_info	40	50	34	61	51	50	26
authority	28	57	33	57	47	44	37
bias	39	63	36	68	54	51	40
clarity	48	67	40	66	71	58	45
depth_scope	51	69	41	79	69	59	46
instructional	44	73	46	68	44	57	47
precis	49	70	48	71	68	56	38
recency	33	49	22	52	47	43	37
specificity	41	60	42	63	63	60	37
topic	54	77	43	76	82	70	36
Total	476	705	434	724	671	614	432

4.2.4 High Relevance Judgments

Table 4.6 contains a cross-tabulation of the criteria selections across all search stages for *high relevance* document selections. Taken as a whole, these results do not demonstrate a statistically significant association between relevance criteria selection and search stage. Additional analysis, however, identified other associations.

Table 4.6: Frequency Counts for High Relevance Selections

Criteria	Becoming	Browsing	Defining	Extracting	Focusing	Learning	Verifying
ability understand	9	23	10	27	12	10	13
amt_info	13	33	16	26	23	15	22
authority	11	19	9	21	15	13	10
bias	5	11	5	8	8	7	9
clarity	11	22	7	15	10	10	11
depth_scope	9	16	10	14	19	13	12
instructional	3	11	5	13	0	2	9
precision	7	20	7	20	19	10	13
recency	14	33	19	28	16	14	19
specificity	13	29	12	28	21	15	19
topic	4	11	4	11	9	3	13
Total	99	228	104	211	152	112	150

The search stage counts in Table 4.4 show that subjects were more likely to select some search stages over others, a tendency which could skew analysis of variance results. Table 4.7 and related Figure 4.1 examine the subset of criteria which varies in frequency

of selection across search stages. This figure shows that in terms of frequency of selection for high relevance, there is a shift in the importance of some criteria across search stages. Most selections appear to peak during the "browsing" stage, and an examination of the total row in Table 4.6 indicates that this is the most commonly selected search stage. Since subjects appeared to have a preference for that search stage in general (as shown in Table 4.4), thus increasing the possibility that selection of a particular criteria in the "browsing" stage may have had more to do with a subject's preference for the "browsing" stage in general, than a preference for a particular relevance criteria in a particular search stage. Analysis of the relationship between criteria selections in other search stages would provide a stronger indication that search stage impacts criteria selection.

To provide greater evidence for the selectivity of criteria choices varying over search stage, an additional data filtering mechanism was applied by selecting search stages which are approximately equal in frequency of document selection: "learning," "browsing," "extracting," and "focusing." With these stage selections being roughly equal, variations in category selections between search stages are more likely to be due to preference for relevance criteria than for search stage.

The Chi Square analysis of variance formula was applied to the results shown in Table 4.7. Analysis of the data determined that there was no statistical relationship when a Chi Square test was performed across all search stages, but an examination of the relationship between specific search stages revealed a statistical relationship between the "focusing" and "browsing" stage for the criteria reported in Table 4.8. These results indicate that for this sample, the criteria of "ability to understand", "clarity" and

"recency" are selected more in the "browsing" stage than in the "focusing" stage, and this frequency of selection has the statistical significance shown in Table 4.8.

Table 4.7: High Relevance - Criteria and Search Stage Subset

Criteria	Learning	Focusing	Browsing	Extracting
ability understand	10	12	23	27
amt_info	15	23	33	26
authority	13	15	19	21
bias	7	8	11	8
clarity	10	10	22	15
instructional	2	6	11	13
precision	10	19	20	20
recency	14	16	33	28
specificity	15	21	29	28
Total	96	130	201	186

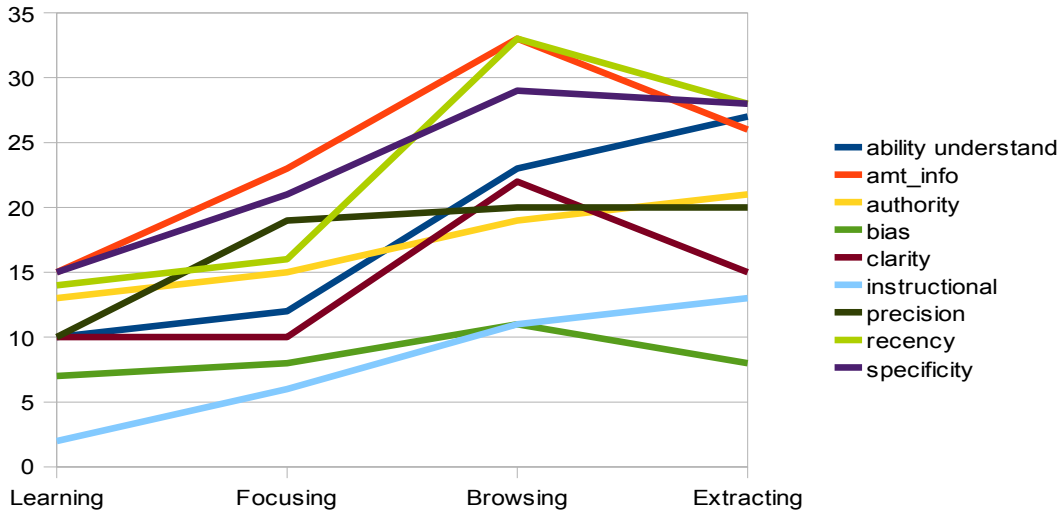


Figure 4.1: Criteria Which Vary over Search Stage

Table 4.8: Criteria Chi Square Values*

Criteria	χ^2	P-value
ability to understand	3.5	$p < .05$
clarity	4.5	$p < .05$
recency	5.9	$p < .05$

* "focusing" and "browsing" search stages

4.2.5 Low Relevance Judgments

Table 4.9 shows the frequency counts for relevance criteria selections with *low relevance* (a selection of 3 or less on a scale of 1 to 10) across search stages. Taken as a

whole, these results do not produce a statistically significant relationship. Additional analysis was performed using specific search stages with similar levels of selection by subjects did detect relationships.

Table 4.9: Frequency Counts - Low Relevance

Criteria	Becoming	Learning	Focusing	Defining	Browsing	Extracting	Verifying
ability understand	1	0	11	5	5	4	5
amt_info	11	8	14	14	20	14	12
authority	26	25	41	21	31	23	16
bias	32	30	43	24	37	34	23
clarity	2	0	8	8	8	6	5
depth_scope	4	7	14	7	16	11	7
instructional	18	15	51	11	26	16	11
precision	9	8	11	5	14	6	7
recency	18	21	35	17	29	20	10
specificity	8	6	12	13	15	10	9
Total	129	120	240	125	201	144	105

The selection of more low relevance criteria selections as subjects progress through the search process would suggest a *decrease* in importance for those criteria as measured by frequency of selection. Conversely, a selection of fewer low relevance criteria would suggest that the criteria is increasing in significance as subjects progress through the search process. Table 4.10 identifies those criteria-stage results which appear to vary across search stages. Figure 4.3 provides a graphical representation of this data. Various relationships were analyzed, and Table 4.11 provides a list of those statistical relationships. Note that criteria such as 'ability to understand,' 'instructional,' 'precision,' and 'recency' appear to have fewer low relevance criteria selections as subjects progress through the search process, possibly indicating these criteria are increasing in importance to subjects.

Table: 4.10 Criteria Which Vary over Search Stage (Low Relevance)

Criteria	Learning	Focusing	Browsing	Extracting
ability understand	0	11	5	4
amount info	8	14	20	14
authority	25	41	31	23
bias	30	43	37	34
clarity	0	8	8	6
depth/scope	7	14	16	11
instructional	15	51	26	16
precision	8	11	14	6
recency	21	35	29	20
specificity	6	12	15	10
Total	120	240	201	144

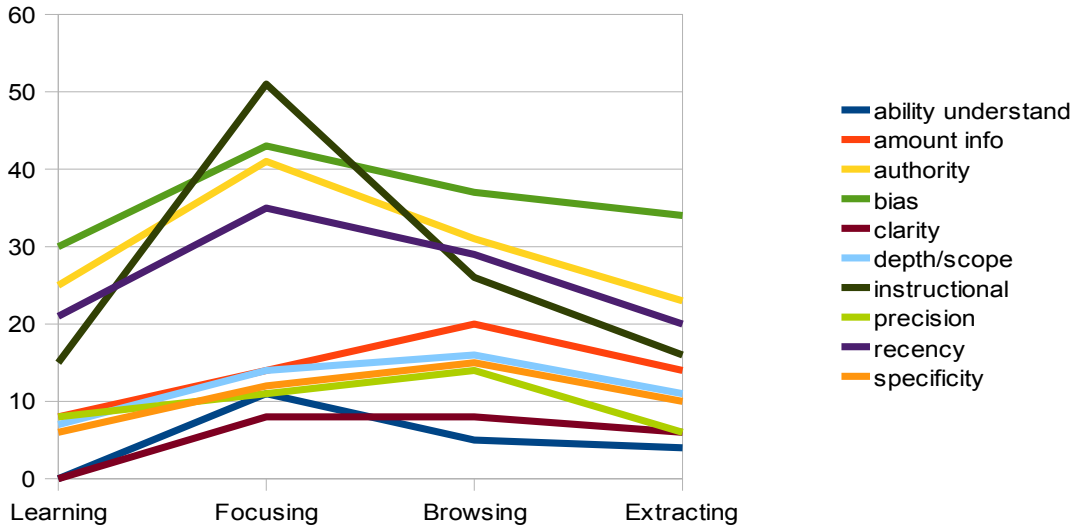


Figure 4.2: Frequency Counts of Selections for Highly Relevant Selections

As shown in Table 4.11, there were statistically significant associations for the "ability to understand" and the "amount of information" criteria moving from the "learning" stage to the "focusing" stage. These statistical results indicate subjects were more likely to select criteria using a low rating in the "focusing" stage than in the "learning" stage.

Conversely, in evaluating the progress from the *focusing* stage to the *extracting* stage, results are found which indicate a statistically significant decrease in the selection

of criteria using a low relevance ranking. As shown in Table 4.12, the criteria of "recency," "instructional," and "authority" all showed a decrease in selection, an indication that these criteria may have increased in significance (since fewer lower relevant selections were made).

Table 4.11 : Criteria Categories with Statistical Significance*

Criteria	χ^2	P - value
ability understand	11	$p < .001$
amount info	11	$p < .001$
authority	3.88	$p < .10$
instructional	19.64	$p < .10$
recency	3.5	$p < .10$

* analyzed for "focusing" and "browsing" search stages

Table 4.12 : Criteria Categories with Statistical Significance

Criteria	χ^2	P - value
recency	4.09	$p < .10$
instructional	18.28	$p < .001$
authority	5.06	$p < .10$

* analyzed for "focusing" and "extracting" search stages

4.2.6 Single Search Stage Per Relevance Selection Analysis

In this study subjects were allowed to select more than one search stage for each relevance judgment. As shown in Table 4.13, a number of subjects reported being in more than one search stage as they judged document relevance. This could indicate that some subjects were confused by the search stage selection and did not discriminate in the selection of search stage, or it could represent a valid reporting of search stage behavior in lieu of a search stage (as in Ellis, 1997). To strengthen these results, analysis was done with subjects who selected one search stage per relevance judgment.

Table 4.13: Search Stage Selection Counts Per Document Judged

Search Stage Selection	Documents Judged	Percent
one search stage	196	51.17%
greater than one search stage	164	42.82%
zero search stages*	23	6.01%
Total	383	100.00%

* zero search stage selections were not included in cross-tabulations of search stage and criteria selected

Table 4.14 lists the search stages selected by subjects for single search stage selection with a degree of relevance value of 5 or greater. These results vary from those in Table 4.4, reflecting different search stage selectivity.

Table 4.14: Single Search Stage Selection *

Search Stage	Count	Pct
becoming informed	15	7.28%
learning	20	9.71%
focusing	32	15.53%
defining extending	21	10.19%
browsing	45	21.84%
extracting	40	19.42%
verifying	17	8.25%
Total	190	92.22%

* exclusive of the "completion" search stage which was not analyzed

The frequency counts for relevance criteria selections for single search stage selections for relevance importance values greater than 5 is shown in Table 4.15. These results are shown graphically in Figure 4.3 which indicates the variation in selection is fairly consistent for most criteria used. However, for the criteria of "clarity," "recency," "depth," and "scope" there is statistically significant variation in moving from the "focusing" stage to the "extracting" search stage as shown in Table 4.16, an indication these criteria were more important to subjects in later ISP stages.

Table 4.15: Criteria Stage Counts for Single Search Stage Selections

Criteria	Becoming	Learning	Focusing	Defining	Browsing	Extracting	Verifying
ability understand	7	12	19	15	30	32	13
amt_info	2	12	18	12	26	27	13
authority	3	8	7	7	21	17	7
bias	2	5	6	6	14	11	5
clarity	7	12	20	9	26	32	15
depth_scope	7	10	15	12	22	27	12
instructional	4	11	11	6	16	27	12
precision	3	14	19	10	24	35	10
recency	9	9	8	7	19	20	9
specificity	7	9	17	11	26	31	10
Total	51	102	140	95	224	259	106

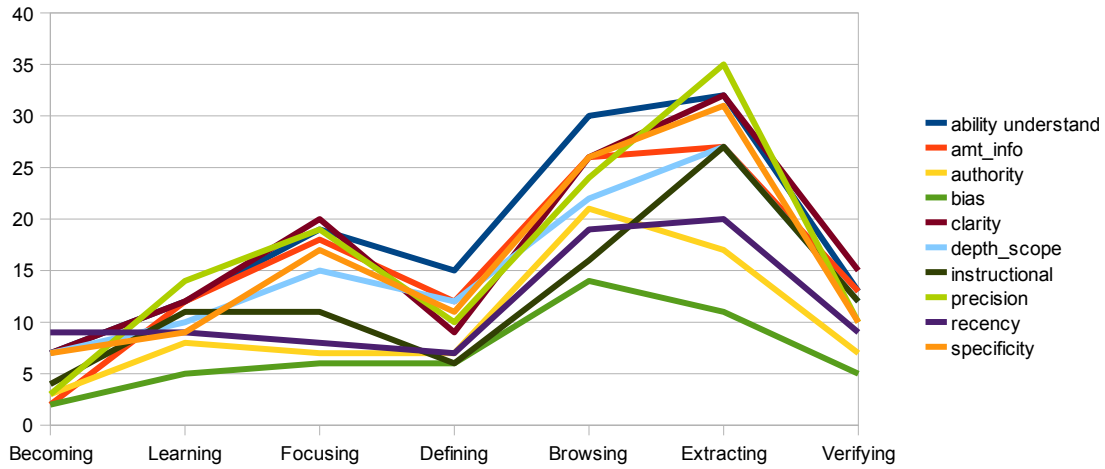


Figure 4.3: Figure 4.3: Criteria Stage Counts for Single Search Stage Selections

Table 4.16: Table Criteria Categories with Statistical Significance*

Criteria	χ^2	**P-value**
clarity	4.08, df = 1	$p < 0.04$
precision	3.3, df = 1	$p < 01$
recency	5.1429, df = 1	$p < .001$
depth/scope	3.43, df = 1	$p < .10$

* results for the "focusing" and "extracting" search stages

4.2.7 Comparison of Results

Table 4.17 compares the results of relationships of high relevance, low relevance and single search stage selections. The criteria of "recency," "instructional," and "authority" demonstrated statistical variations at various levels of significance in either the high and low relevance analysis, strengthening the case for the underlying relationship between relevance criteria selection and search stage. The criteria of "recency" also demonstrated statistically significant results in the analysis of single search stage, further strengthening the case for the relationship of search stage and relevance criteria selections.

Table 4.17: Comparison of Statistically Significant Results

Criteria	High	Low*	Single Search Stage
ability to understand	x		
clarity	x		x
amount of information		x	
recency	x	x	x
instructional	x	x	
authority		x	
depth/scope			x
precision			x

* a decrease in low relevance ratings - see table 4.12

4.3 Discussion

The data collected in this study provide empirical evidence of the relationship of relevance judgment choices to relevance judgment (research question1), and of the relationship of relevance criteria choices to search stage (research question 3). Statistically significant results demonstrate the relationship of relevance criteria choices to search stage, indicating that as subjects progress through the search process, their relevance criteria choices may change, with some criteria being more important in some search stages than other criteria. These changes occur as subjects progress through an individual search episode and evaluate documents and make relevance judgments.

4.3.1 Major Findings

Specifically, the major findings based on these results are as follows.

1. Subjects demonstrated a preference for the criteria of "ability to understand," "clarity," "depth/scope," "precision," and "specificity" for relevant documents over all search stages (research question 1) as shown in Table 4.3.

2. Analysis found the statistical relationships shown in Table 4.17, suggesting that these criteria increased in importance to searchers as they progressed through the information search process (research question 3).

4.3.2 Detailed Discussion

Based on these statistically significant results, it appears that based on frequency of selection, the criteria of "recency,", "instructional" increase in importance as the user progresses through the search process. This tendency was demonstrated with both the low relevance and high relevance analysis. It is possible that since the question posed to the subjects involved a topic somewhat unfamiliar with the subjects, the criteria of "instructional" was important. The selection of "authority" of the source as a criteria with low relevance also provides some indication that the subjects were evaluating who was providing the information in making their relevance decision. The criteria of "recency" also appears to be a consistently selected criteria for subjects in this study, and increases in importance as subjects progress through the ISP, even though the question did not require current information.

Subjects were assigned a question which asked them to "compare and contrast" two concepts (see Appendix P). This should have led to some use of "bias" as part of the criteria for relevance, but though selection of that criteria selection does fluctuate over several search stages, this variation was not statistically significant. The criteria of "clarity," "ability to understand," "depth/scope," and "precision" also increase in importance as subjects progress through the search process. Subjects appear to be more concerned with being able to understand and interpret the material and to find documents with "depth/scope" in later stages, and this study finds this tendency to be statistically significant.

4.4 Implications for Study 2 and Study 3

It is possible that the short duration study of the ISP in Study 1 may not have allowed 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 in Study 1, 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 might therefore be appropriate to provide subjects with a smaller set of relevance criteria from which to choose. The design of methods for Study 2 and Study 3 reflect these observations.

Chapter 5 - Research Study 2

Study 1 provided a basic examination of relevance criteria selections during the information search process. The methods for that study compressed the search process into a short time frame and placed subjects in a monitored lab environment. The results of Study 1 and the experience of conducting the research informed the design of Study 2 and 3. Based on the results of Study 1, Study 2 used a naturalistic search environment and allowed a 5 week time period for subjects to gather information. This methodology allowed a more detailed analysis of the subject's progression through the search process and their selection of relevance criteria. The use of interim deliverables due on a weekly basis encouraged subjects to work more consistently over the period of time alloted for the study, and allowed the examination of relevance criteria choices in relation to stage in task completion. Analysis of Study 1 data indicated subjects had a tendency to select criteria in groups throughout the search process. Study 2 therefore included analysis of selection of groups of criteria across search stages.

Study 2 continued examination of the relevance criteria choices by subjects as they progress through the information search process and proceed to complete their work task. Study 2 added to this examination of progress through a work task, and examination of groups of relevance criteria. The data collected in this study is applicable to analysis of all research questions. The following section describes the research methods in detail.

5.1 Methods

Subjects for Study 1 worked on a real information problem in a naturalistic environment: collecting information to write a graded presentation (research project) as part of a college class (see assignment description in Appendix A). The research assignment allowed subjects to work for several weeks to search for and gather

information for the assignment. Interim deliverables encouraged subjects to work in a consistent manner and provided the opportunity to track the subject's progress in relation to a stage in task completion. The following section provides details on the methods for this study. Each research question is addressed using the abbreviation RQN where, N is one of the four research questions for these studies.

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

Subjects conducted searches and reviewed documents returned from their searches. Using the Web-based search engine, subjects indicated the relevance for the documents they examined (RQ1). Relevance was one of either relevant, not relevant, partially relevant/not sure about relevance. Subjects also identified a *stage in the information search process* (RQ3) by selecting from a predetermined list of search stages, and indicated the criteria used to make that relevance judgment by selecting from a predetermined list of relevance criteria.

The subject's progress for the assignment was monitored and a project deliverable was due approximately each week. These interim deliverables are identified in the analysis for this study 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 fifth week. Associating the date when the criteria used to judge relevance for the document was

stored with the project deliverable required during that time period provides an opportunity to correlate the relevance criteria used with *stage in task completion* (as associated with the project deliverable data). Associating the subject's selection of stage in the information search process with relevance criteria used also provides an association of relevance criteria choices and user-selected *stage in the information search process*.

Data was collected through a Web site which contained detailed instructions for the subjects on the use of the site and how to provide information about their searches and relevance judgments. The site was accessible from the Internet with any Web browser, thus allowing subjects to perform the research without intrusive monitoring. The operation of the site was 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. If the subjects required additional help on using the web site, Web pages could be reviewed to explain the search stage choices (see Appendix J), relevance judgment choices (see Appendix J), and criteria for relevance judgment choices (see Appendix K) used by the subjects. Participants were also provided instruction on using the research Web site and allowed to practice using the site as part of a tutorial session.

5.1.1 Search Process

The search stages presented to subjects in Study 2 were based on Kuhlthau's (1993) ISP stages, and Ellis's (1997) search patterns as synthesized by Wilson (1999), these stages are summarized in Table 3.3. The 'ending' stage was not used in this research since the completion of the research project by the subjects implies the search process is complete. Table 5.1 shows the research stages used in this section, and the description displayed to the subjects.

Table 5.1 : Search Process - Study 2

Search Stage Term	Description Displayed
initiation	contemplating topics/subtopics and brainstorming
exploration	reading to become informed and taking notes
differentiating	choosing between subtopics
extracting	collecting information; taking detailed notes and writing
verifying	rechecking sources and confirming information

5.1.2 Relevance Criteria

The relevance criteria list displayed to subjects is shown in Table 5.2. These criteria are an annotated version of the list of criteria presented in Table 3.1, reduced for clarity and to avoid survey exhaustion on the part of the subjects. This list is based on criteria identified by Barry (1994), Barry and Schamber (1998), and Cool et al. (1993).

In Study 2 and Study 3, topic was presented as a criteria choice to the subjects but was not part of the analysis, since all documents judged relevant are assumed to be on topic.

Table 5.2: Relevance Criteria Descriptions Displayed - Study 1

Term	Description	Explanation
topic	document is on my topic	the topic of this document matches or nearly matches the topic I have chosen
instructions	document contains basic advice and instructions (tips)	the document contains tips and advice that improve your understanding of the subject
understandability	document is easy to understand; the technical information is easier to read and interpret	the technical information in the document is presented in such a way that it is easy to understand
history	document contains a history and/or background of the topic	the document contains a history and background on the subject
guidelines	document contains basic guidelines and directions	the document contains basic guidelines and directions
novelty	the content of the document adds new information to what	relative to documents already retrieved, this document provides

Term	Description	Explanation
	I already have	new information or perspective
affectiveness	document is enjoyable	the document is fun to read
source	the document is from a source (website, journal) which has a good reputation in this area	the source of this document (magazine name, journal, university) is known to be a quality source
authority	the author of the document is considered an expert in this field	the author of the document has a reputation for being accurate, correct and providing useful information
recency	document is up to date and contains current information	for this field, this is a fairly recent document
definitions	document contains basic and/ or advanced definitions	the document contains definitions which improve my understanding of the subject
descriptions	document contains good descriptions and explanations	the document contains descriptions and explanations that improve my understanding of the subject
breadth	document covers many topics/subtopics in this subject area	the breadth of coverage across subtopics is good; a good survey of the field
structure	the structure of the document makes it easier to read and understand	the structure of the document makes it more understandable
time	document is useful because of time constraints	the document was selected because I have little time left to search for other documents
accuracy	document seems to have accurate information about my topic	the facts in the document appear to be accurate to the best of my knowledge
bias	document author takes a stand and has a specific opinion (bias); the author is not neutral	the author approaches the subject with a bias
depth	document contains good depth on the topic	the depth of coverage within the field is good; provides sufficient detail on the topic

5.1.3 Data Collection Process

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

Each subject performed 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 was as follows.

- Subjects were asked to sign an informed consent form which explained the purpose of the research and that the information they provide would be treated anonymously.
- Subjects logged into the research Web site to conduct their search using a login ID previously assigned, and a personal password they had chosen.
- Subjects entered search query terms as if they were using a commercial search engine such as Yahoo! and executed a search (see Appendix D). The search results were 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 generated 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 included an explanation of the result page/document (as returned by the commercial search engine), links to the results page, and links which the user used to enter relevance information about the page (see Appendix E).
- Subjects were 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 (see Appendix F). Relevance assessments were one of either *relevant*, *not relevant*, or *partially relevant/ not sure* (from Greisdorf, 2003; Spink et al., 1998). Criteria choices available to the subjects were those identified in Table 3.1 and related paragraphs. To address issues of order effects, these choices were randomized in the list of relevance criteria choices displayed to the subjects.
- When the subjects finished providing information for the documents reviewed on the results page, they submitted the information they entered to the data collection program on the research Web site which stored the results anonymously for later analysis.
- Subjects repeated the process outlined above as often as they felt necessary and whenever they wished in order to gather the information they needed to complete their assigned report.
- A post-test questionnaire was used to collect background information for reference purposes (Appendix M). This information was associated with a subject,

but was maintained and stored anonymously (linked to the anonymous subject ID).

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

5.2 Study 2 - Pilot Study

A pilot study was conducted to test the methods and data collection instruments for Study 2. This study 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 used in this study were the same as those detailed for Study 2 with two exceptions. In the 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 B, D, and G for images of the data collection pages). Also, the pilot study had a single deliverable: an assigned research project due three weeks after assignment.

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 5.3. Subjects appeared to have a lack of preference for the search stages identified as "differentiating" and "verifying." It also appears from this data that the search stage of "extracting" represents a consistent choice, and the stage of "exploration" (with the description of "scanning for information") was chosen much less frequently than "extracting."

Table 5.3 - Pilot Study 1 Relevance Criteria Choices

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

It is possible that subjects were progressing through multiple stages of the ISP as they examined documents retrieved by their search, but since the search stage was selected once at the start of the search process, the subject had no way to record changes in search stage with the method used for this study.

To increase the likelihood of subjects providing a more consistent reporting of their search progress, interim deliverables were added to the methods for Study 2. These interim deliverables encouraged 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 provided an opportunity to analyze criteria selections and relevance judgments in relation to *a stage in task completion*. To encourage a more varied and accurate selection of search stages, the data collection instrument was also changed to allow a subject to select a search stage with the evaluation of each document/web site, instead of selection of search stage once at the start of the search session.

5.3 Results of Data Collection

A convenience sample of 82 subjects from a population of undergraduate students at an American university participated in the study in Fall of 2007. Subjects examined and reported on a total of 758 distinct documents found on the Web.

Sixty-four of the 82 subjects provided background information in the form of a survey. These results indicated that approximately 36% of the subjects reported searching for information at least once a day, and approximately 41% reported searching for information several times a day. Approximately 64% of the subjects were male, and 35% were female. Most subjects (77%) spoke English as their primary language.

Table 5.4 provides a count of Web documents evaluated by subjects in relation to the stage in task completion. Analysis of criteria for this study did not include the criteria of "topicality." Given the parameters of the subject's assignment, the selection of documents which are on the assigned topic is assumed.

Table 5.4 - Documents Assessed by Deliverable Due

Deliverable	**Count**	**Percent**
Abstract	81	10.69%
Detailed Outline	225	29.68%
Rough Draft	187	24.67%
Final Presentation	265	34.96%
Total	758	100.00%

Table 5.5 reports the total number of relevance judgments made in each search stage. Since a number of documents were selected in multiple stages by subjects, the document count varies across these tables. Table 5.4 only counts those documents once, Table 5.5 however, counts all incidents of document and search stage and therefor has a higher total count. Table 5.6 compares the search stage reported by the subject to the stage in task completion as determined by the point in time when the subject made their relevance judgment.

A progression through the search process which is consistent with the assigned deliverables would have subjects reporting early search stages ("initiation," "differentiation," "exploration,") during the preparation of the "project abstract" and "detailed outline" deliverable as they learn about the topic and do the their initial

research, and then reporting later search stages ("extracting," "verifying") during the preparation of the "rough draft" and "detailed outline" deliverable. However the results shown in Table 5.6 indicate that subjects are reporting early search stages throughout the duration of the assignment. The "initiation" stage is reported early in the search process during the preparation of the "project abstract" deliverable, and much later near the end of the duration for the assignment, during the preparation of the "final draft" deliverable. This would suggests that within the time frame of the document deliverable, subjects are repeating the search process as needed in an iterative fashion.

Subjects also appeared to demonstrate a preference for selecting some search stages over others. The "initiation," "extracting," and "exploration" stages are most commonly selected. The stage of "differentiating" was reported much less often than other stages, and the stage of "verifying" was reported even less.

Table 5.5 - Documents Selected by Stage in the Search Process*

Search Stage	Count	Percentage
initiation	182	18.88%
exploration	197	20.44%
differentiating	71	7.37%
extracting	427	44.29%
verifying	87	9.02%
Total	964	100.00%

* documents may be selected in multiple stages

Table 5.6 : Search Stage and Deliverable Comparison

Search Stage	Abstract	Detailed Outline	Rough Draft	Final Presentation
initiation	258	498	137	109
exploration	125	505	84	156
differentiating	41	208	24	113
extracting	113	702	555	652
verifying	11	54	43	196
Total	548	1967	843	1226

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 (research question 3). The analysis for the importance of criteria to subjects in this study is based on frequency of selection and the assertion that a relevance criteria that is more commonly selected in a search stage is more important to the subject in that search stage than other criteria.

Table 5.7 lists relevance criteria and search stage reported for all relevance judgments as a percentage of the total reported for the search stage. Since there were a large number of criteria provided and reported by the subjects, this table has been simplified to list only those criteria which changed most as the subject moved through the search process. These results show some variability in the selection of search stage and criteria used to judge relevance, and provides some indication that the importance of these criteria to the subjects varies by search stage. Further analysis is required to determine whether or not these variations are statistically significant. The analysis presented here involved stronger statistical methods, specifically an analysis of variance which examined frequency count changes across search stages (RQ3), and stage in task completion (RQ4). To provide data for analysis of research question 4, results were filtered to determine which groups of criteria were selected by subjects while evaluating document relevance.

Table 5.7: Criteria Choices as a Percentage Selected for a Search Stage

Criteria	initiation	differentiating	exploration	extracting	verifying
accuracy	10.31%	9.17%	11.05%	10.39%	10.94%
affectiveness	0.57%	4.37%	2.25%	2.77%	3.52%
amount of information	8.97%	12.23%	8.61%	10.77%	7.42%
authority	1.91%	3.49%	2.43%	2.65%	3.91%
bias	1.91%	1.75%	1.12%	1.94%	3.13%
breadth	5.92%	5.68%	6.93%	6.00%	5.86%
depth	8.78%	12.23%	8.99%	9.03%	7.42%
novelty	1.15%	3.93%	1.12%	2.06%	2.73%
recency	8.40%	7.86%	7.87%	6.97%	6.25%
source quality	4.77%	3.49%	2.06%	4.06%	3.13%
structure	6.30%	5.68%	7.49%	7.55%	8.20%
time constraints	0.57%	1.31%	0.94%	1.29%	2.34%
understandability	13.17%	12.66%	13.67%	13.55%	13.28%

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

Table 5.8 shows the type of relevance judgment made by search stage. This indicates that subjects had a general preference for judging or reporting relevant document judgments as opposed to not relevant or partially relevant. Partially relevant documents were selected most often during the initiation stage, and not relevant documents were primarily selected in the initiation and exploration stages -- early stages of the ISP. Continuing the focus on these relevance judgment types, the following sections analyze these results based on judgment types.

Table 5.8: Relevance Judgment by Search Stage*

Search Stage	not relevant	partially relevant	relevant	Total
initiation	11.09%	18.36%	70.55%	100.00%
differentiating	4.80%	12.66%	82.53%	99.99%
exploration	10.71%	12.03%	77.26%	100.00%
extracting	7.10%	4.07%	88.83%	100.00%
verifying	5.49%	13.73%	80.78%	100.00%

* as percentage of relevance judgments made during a search stage

5.3.1 Partial Relevance Judgments

In Table 5.9, the frequency counts for the selection of criteria for partially relevant documents is shown. Based on these results, the criteria of "ability to understand," "accuracy," and "recency" appear to be important to subjects during this search stage.

These results demonstrate that subjects showed a preference for some criteria over others when making a partial relevance judgment, and provide empirical results for the analysis of research question 1. Further examination of the selection of these criteria in relation to search stage will provide additional insights.

Table 5.9: Frequency Counts for Criteria Selection - Partial Relevance *

Criteria	Count	Percentage	Rank
accuracy	34	11.85%	2
affectiveness	8	2.79%	
amount of information	22	7.67%	4
authority	3	1.05%	
bias	6	2.09%	
breadth	16	5.57%	
depth	15	5.23%	
novelty	3	1.05%	
recency	24	8.36%	3
source quality	8	2.79%	
structure	19	6.62%	5
time constraints	4	1.39%	
understandability	52	18.12%	1
Total	214	74.58%	

* topic not listed

The frequency counts for criteria judgments within each search stage are shown in Table 5.10. The results in this table, taken as a whole, indicate a statistically significant relationship ($\chi^2 = 62.96$, $p < .10$) between search stage and relevance criteria selection.

Table 5.10: Frequency Counts for Partial Relevance across Search Stage

Search Stage	accuracy	affectiveness	amt	auth	bias	breadth	depth	novelty	recency	source	struct	time	underst
initiation	13	1	4	1	1	4	2	0	12	3	6	0	18
differentiating	3	2	7	1	1		1	2	2	1	1		5
exploration	7	1	4		2	4	5	1	3	1	5	2	12
extracting	6	0	5		0	5	6	0	5	3	4	2	13
verifying	5	4	2	1	2	3	1	0	2	0	3	0	4
Total	34	8	22	3	6	0	15	3	24	8	19	4	52

As in previous studies, subjects showed a preference for some search stages over others based on frequency of selection, as shown in Table 5.11 and shown graphically in Figure 5.1. These results show that subjects are more inclined to select documents as partially relevant in *early* search stages ("initiation," "exploration," and "differentiating")

where they select 66 % of all partially relevant documents. Only a third (33 %) of all partially relevant documents are selected in *later* search stages ("extracting," and "verifying"). It would therefore be logical to test for an association between early and late stages.

Table 5.11: Partial Relevance Selections by Search Stage - Study 1

Search Stage	Count	Pct
initiation	96	33.45%
exploration	64	22.30%
differentiating	29	10.10%
extracting	63	21.95%
verifying	35	12.20%
Total	287	100.00%

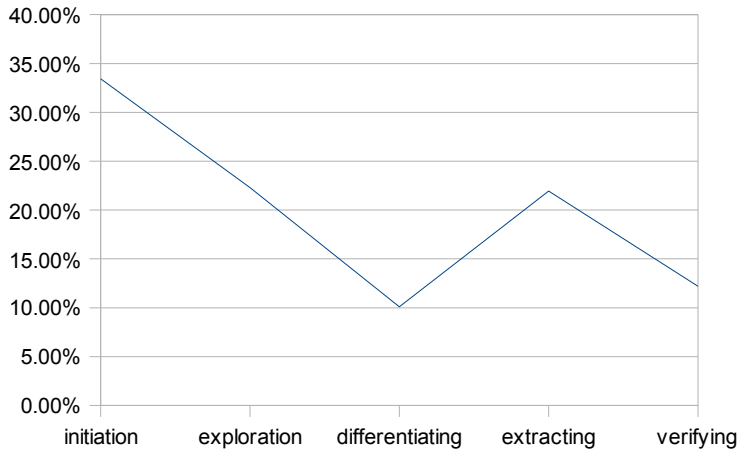


Figure 5.1: Partial Relevance Judgments by Search Stage

The document counts for partial relevance judgments were small, so a limited amount of analysis of variance could be done. When the results in Table 5.10 were analyzed, associations were found for the search stages of "initiation" and "exploration" and the criteria of "recency" (χ^2=5.4, df=1, $p < .05$), and for the criteria of "recency" and the search stages of "differentiation" and "verifying" (χ^2=3.56, $p < .10$). These results suggest the criteria of "recency" increases in importance over the course of the information search process for documents judged partially relevant These results suggest

that "recency" increases in importance over the course of the information search process for documents judged partially relevant. .

5.3.2 Relevant Judgments

Subjects in this study were more likely to judge documents relevant rather than partially relevant or not relevant as shown in Table 5.12. Table 5.13 show the frequency counts for relevance criteria selections for relevant documents over all search stages. These results provide evidence for the analysis of research question 1, and demonstrated that subjects showed preference for some relevance criteria over others. As with partial relevance judgments, the criteria of "ability to understand" and "accuracy" were important to subjects in judging document relevance. Additional analysis provided some indication of how these preferences changed over the course of the search process.

Table 5.12: Frequency Counts/Percentages for Document Relevance Judgments

Judgment	Count	Percentage
not relevant	250	8.08%
partially relevant	287	9.28%
relevant	2551	82.48%
Total	3088	99.84%

Table 5.13: Frequency Counts for Criteria Selections - Relevant Document Judgments*

Criteria	Count	Percentage	Rank
accuracy	280	10.98%	2
affectiveness	58	2.27%	
amount of information	242	9.49%	3
authority	76	2.98%	
bias	50	1.96%	
breadth	151	5.92%	
depth	231	9.06%	4
novelty	53	2.08%	
recency	196	7.68%	5
source quality	105	4.12%	
structure	185	7.25%	
time constraints	31	1.22%	
understandability	355	13.92%	1
Total	2013	78.93%	

* the criteria of "topic" is not listed

The analysis of the frequency of selection of criteria for documents judged to be relevant was based on the contingency table shown in Table 5.14. Viewing these results as a percentage of selections as shown for a criteria across search stages provides some insight into which criteria the subjects considered important in a particular search stage, and how these preferences changed over the course of the search process as shown in Table 5.15 and Figure 5.2. In examining the rankings shown across all stages in the ISP, the criteria of "ability to understand" and "accuracy" were consistently the top two selections within a search stage. Some variation existed in the rank of 3 through 5 for the criteria of "recency," "depth," and "structure." Based on frequency of selection, subjects demonstrated varying degrees of preference for these criteria as they progressed through these stages of the search process. A statistical analysis of variance test determined that several of these shifts in preference were statistically significant as shown in Table 5.16. These results provide a statistically strong indication of a relationship between relevance criteria choices and progress through the information search process.

Table 5.14: Frequency Count of Criteria Selections for all Search Stages - Relevant Documents

Search Stage	accur	affect	amt	auth	bias	breadth	depth	novelty	recency	source	struct	time	underst
initiation	38	2	35	8	9	23	34	5	30	22	24	3	51
exploration	49	11	30	13	4	30	35	5	38	9	31	3	56
differentiate	18	7	19	7	2	13	24	7	15	7	12	2	23
extracting	152	33	142	39	30	77	120	29	100	59	103	17	196
verifying	23	5	16	9	5	8	18	7	13	8	15	6	29
Total	280	58	242	76	50	151	231	53	196	105	185	31	355

Table 5.15: Percentage of Criteria Selections within a Search Stage - Relevant Documents

Criteria	initiation	exploration	differentiating	extracting	verifying
accuracy	10.30% (2)	11.92% (2)	9.52% (4)	11.05% (2)	11.17% (2)
affectiveness	0.54%	2.68%	3.70%	2.40%	2.43%
amt info	9.49% (3)	7.30%	10.05% (3)	10.32% (3)	7.77% (4)
authority	2.17%	3.16%	3.70%	2.83%	4.37%
bias	2.44%	0.97%	1.06%	2.18%	2.43%
breadth	6.23%	7.30%	6.88%	5.60%	3.88%
depth	9.21% (4)	8.52% (4)	12.70% (1)	8.72% (4)	8.74% (3)
novelty	1.36%	1.22%	3.70%	2.11%	3.40%
recency	8.13% (5)	9.25% (3)	7.94% (5)	7.27%	6.31%
source qual.	5.96%	2.19%	3.70%	4.29%	3.88%
structure	6.50%	7.54% (5)	6.35%	7.49% (5)	7.28%
time const.	0.81%	0.73%	1.06%	1.24%	2.91% (5)
understand	13.82% (1)	13.63% (1)	12.17% (2)	14.24% (1)	14.08% (1)

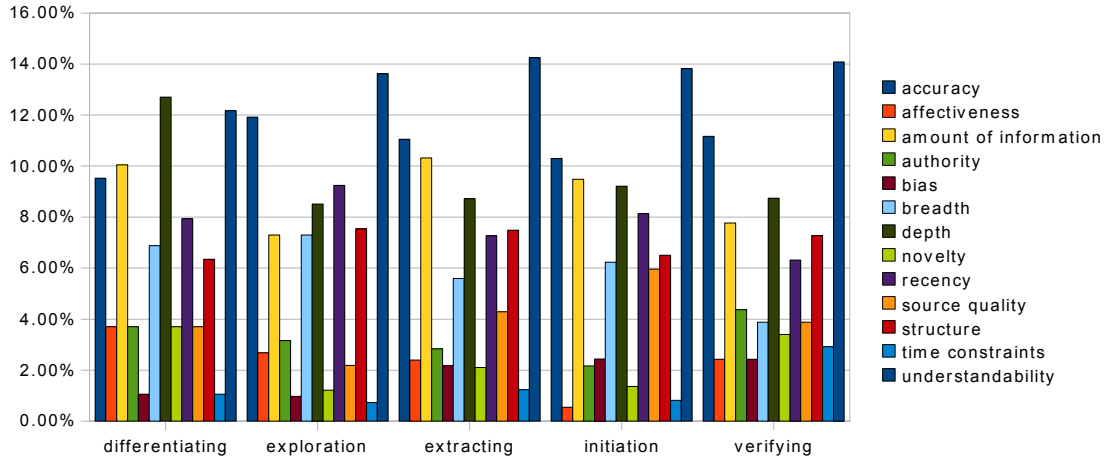


Figure 5.2: Percentage of Criteria Selections within a Search Stage

Table 5.16: χ2 for Criteria Selections across All Search Stages - Relevant Documents*

Criteria	χ2
accuracy	216.46
affectiveness	53.03
amount	231.26
authoritative	47.95
bias	52.6
breadth	100.36
depth	151.71
novelty	40.3
recency	128.95
source	93.05
structure	153.24
time	24.97
ability to understand	286.17

* $p < .001$

Results indicated that since there was an unequal selection of documents across search stages as shown in Table 5.5, reflecting a tendency of subjects to select some search stages over others. Therefore folding of search stage results into combined categories would mitigate this tendency and strengthen the statistical results of this analysis. The filtering process for the analysis combined the "initiation," "exploration" and "differentiating" stages. As the results in Table 5.17 show, this combination provides a frequency count which is 70 % of the frequency count for the "extracting" stage, providing a basis for an approximately equal comparison between *early* search stage relevance judgments ("initiation" through "exploration") and *later* search stage judgments ("extracting").

Table 5.17:Relevant Judgments by Search Stage

Search Stage	Count
initiation	369
exploration	411
differentiating	189
extracting	1376
verifying	206
Total	2551

Table 5.18 lists the frequency counts for those criteria which demonstrated a statistically significant association between *early* search stage selections and a *later* search stage (extracting). Figure 5.3 provides a graphical presentation of this data. A dichotomous analysis of variance test between early and later search stages generated the statistical results shown in Table 5.19.

Table 5.18: Select Frequency Counts for Relevant Documents

Stage	accuracy	amt of info	bias	depth	novelty	source	structure	time	understand
Early	105	84	15	93	17	38	67	8	130
extract	152	142	30	120	29	59	103	17	196

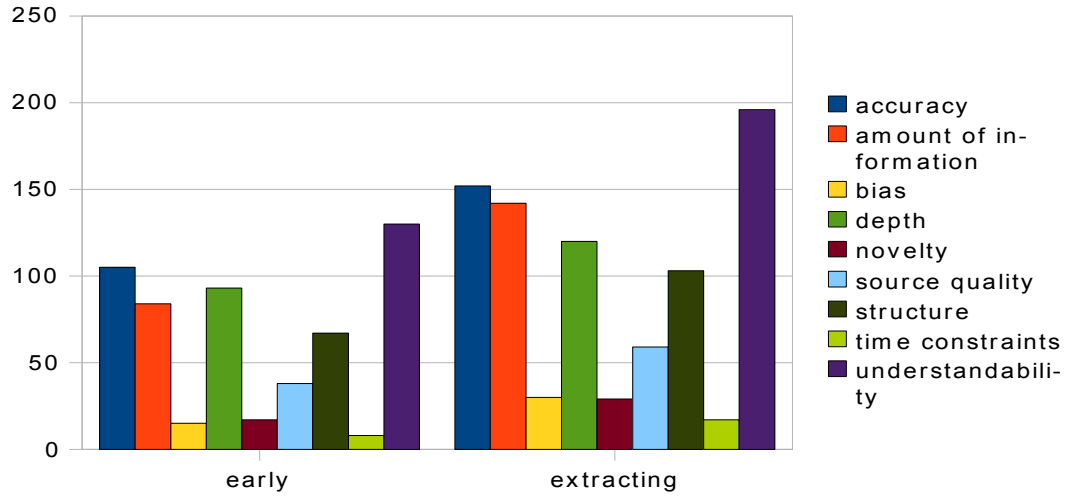


Figure 5.3: Select Frequency Counts for Relevant Documents

Table 5.19: $\chi 2$ Values for Statistically Significant Associations - Early Stages vs. Extracting*

Stage	accuracy	amt of info	bias	depth	novelty	source	structure	time	underst
χ^2	8.6	14.89	5	3.42	3.13	4.55	7.62	3.24	13.35
p value	p < .005	p <.001	p <.05	p <.10	p <.10	p <.05	p <.05	p <.10	p <.001

5.3.3 Not Relevant Judgments

Table 5.20 lists the frequency counts for documents judged not relevant across the search stages used. Comparing these results with the frequency counts for relevant document judgments in Table 5.17, it is clear that these judgments are a small part of the judgments made in any particular search stage as shown in Table 5.21.

Table 5.20: Frequency Counts by Search Stage for Not Relevant Judgments

Search Stage	Count
initiation	58
exploration	57
differentiating	11
extracting	110
verifying	14
Total	250

Table 5.21: Criteria Selections for Not Relevant Documents versus Relevant Selections by Search Stage

Search Stage	Not	Relevant	Percent Selected*
initiation	58	524	9.97%
exploration	57	534	9.64%
differentiating	11	229	4.58%
extracting	110	1550	6.63%
verifying	14	256	5.19%
Total	250	3093	

* as a percentage of "relevant" and "not relevant"

Table 5.22 shows the selection of relevance criteria for not relevant document judgments. These results provide empirical results for analysis of research question 1. The "amount of information" the document provides, and the "depth" of the document were frequently selected criteria for not relevant judgments. No statistically significant associations were found in analyzing these results for changes in frequency of selection across search stage.

Table 5.22: Criteria Counts as a Percentage of All Not Relevant Judgments

Criteria	Count	Percentage	Rank
accuracy	9	5.42%	
affectiveness	11	6.63%	5
amount of info	42	25.30%	1
authority	3	1.81%	
bias	2	1.20%	
breadth	22	13.25%	3
depth	35	21.08%	2
novelty	4	2.41%	
recency	8	4.82%	
source quality	2	1.20%	
structure	20	12.05%	4
time constraints	2	1.20%	
understand	6	3.61%	
Total	166	100.00%	

5.4 Stage in Task Completion Analysis

An examination of the subject's selection of criteria in reference to the project deliverable due provides an opportunity to examine criteria selections in relation to stage in task completion as posed by research question 2. This analysis provided similar insights in the subject's relevance decision as detailed below. As shown in Table 5.4, the number of documents evaluated for the Detailed Outline, Rough Draft, and Final Presentation deliverables are roughly equal. A comparison of the selections made in these stages of task completion provides a sample where the statistically significant changes across these stages is more likely due to the preference for criteria within a stage, than for the selection of documents in that stage. The following sections detail the results of analyzing the frequency counts of documents evaluated for these deliverables.

5.4.1 Partial Relevance Judgments Selections by Deliverable Due

Table 5.23 lists the frequency counts for partial relevance selections for the preparation of the Detailed Outline deliverable and the Rough Draft deliverable. Figure 5.4 provides a graphical presentation of this data. These presentations indicate some

degree of variability in terms of frequency of criteria selection during the preparation of these deliverables.

Table 5.23: Comparison of Criteria Code Selections for Detailed Outline and Final Presentation - Partial Relevance

Criteria	Detailed Outline	Final Presentation
accuracy	14	9
affectiveness	1	6
amount of information	10	7
authority	1	1
bias	1	4
breadth	6	5
depth	9	4
novelty	3	0
recency	19	4
source quality	2	4
structure	10	5
time constraints	0	2
understandability	24	13
Total	100	64

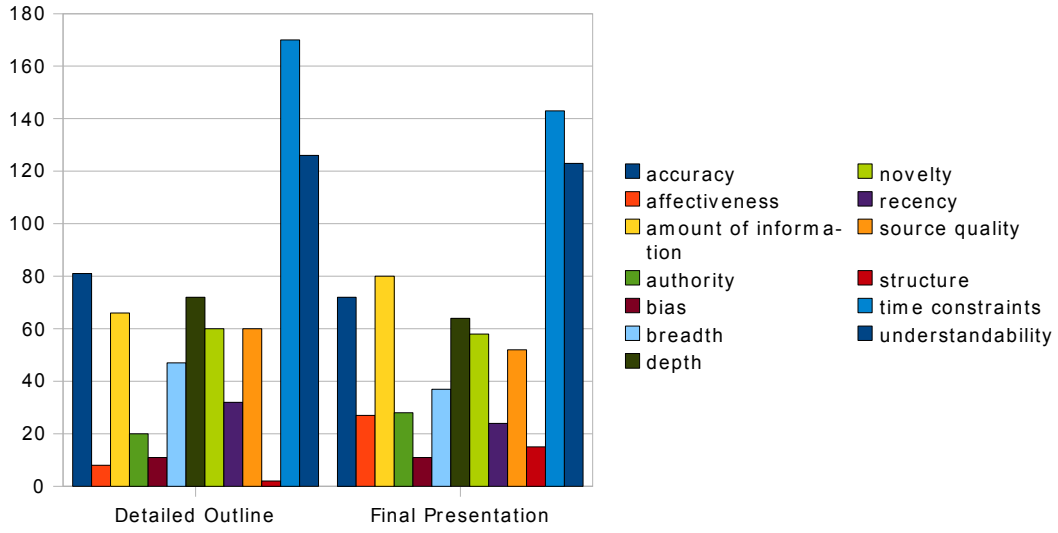


Figure 5.4: Comparison of Criteria Code Selections for Detailed Outline and Final Presentation for Partial Relevance

In analyzing this data as a whole, a statistical relationship is found for all criteria across all deliverables (χ^2=53.79, p < .10). This indicates a relationship between criteria and and stage in task completion. Analyzing these results using a dichotomous analysis reveals relationships between the "Detailed Outline" and "Final Presentation" deliverables for the criteria of "recency" (χ^2=9.78, p< .005) and "ability to understand" (χ^2=3.27, p<.10). Since these criteria choice counts decrease in value over the dates of these two deliverables, this association is an indication that partially relevance judgments for these criteria are less important when moving from the detailed outline stage in task completion to the final presentation stage in task completion.

5.4.2 Relevant Judgments Selections by Stage in Task Completion

Table 5.24 lists the frequency counts for relevance criteria choices made by subjects during the preparation of project deliverables for relevant documents. Figure 5.5 removes some criteria for clarity and provides a graphical presentation of this data. An analysis across the deliverables of Detailed Outline, Rough Draft, and Final Presentation found several associations at various levels of significance as shown in Table 5.25. As with the analysis of changes in criteria selection for search stage, the criteria of "ability to understand" and "amount of information" were important to subjects and increased in importance as the subject progresses through stages in task completion.

Table 5.24: Frequency Counts of Relevance Criteria Choices for Project Deliverables - Relevant Documents

Criteria	Detailed Outline	Rough Draft	Final Presentation
accuracy	81	64	72
affectiveness	8	12	27
amount of information	66	43	80
authority	20	13	28
bias	11	14	11
breadth	47	28	37
depth	72	55	64
novelty	14	8	22
recency	60	38	58
source quality	32	24	24
structure	60	34	52
time constraints	2	10	15
topic	170	105	143
understandability	98	67	107
Total	741	515	740

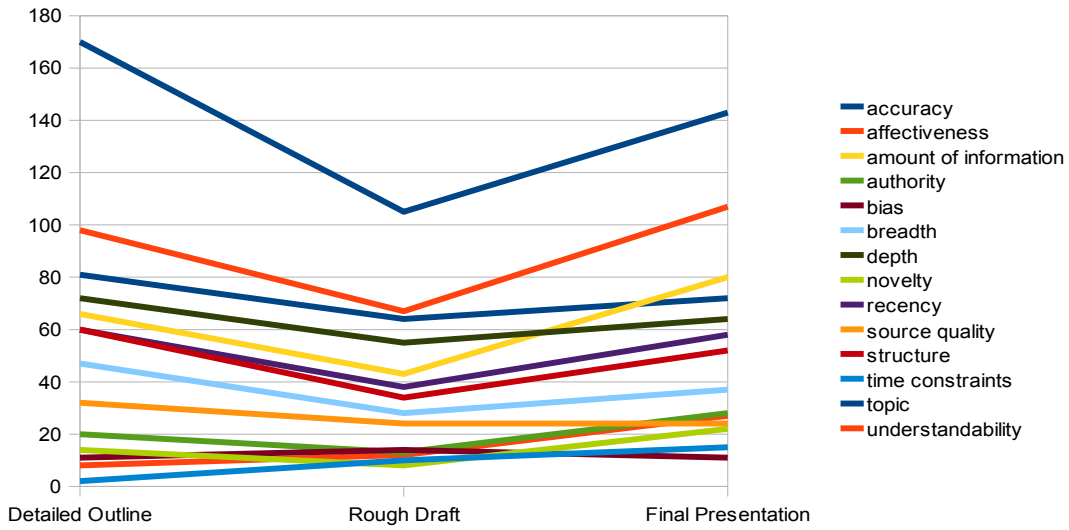


Figure 5.5: Frequency Counts of Select Relevance Criteria Choices for Project Deliverables

Table 5.25: Statistically Significant Associations across the Detailed Outline, Rough Draft, & Final Presentation

Criteria	χ^2	p
affectiveness	12.81	$p < .001$
amount of information	11.08	$p < .001$
authority	5.54	$p < .10$
breadth	4.84	$p < .10$
novelty	6.73	$p < .05$
recency	5.69	$p < .10$
structure	7.29	$p < .05$
understandability	9.71	$p < .05$

5.5 Analysis of Grouped Relevance Criteria Selections

Analysis of single criteria selections by subjects showed statistical variations over search stage and stage in task completion. The criteria of "amount of information," "depth," and "accuracy" increased in importance over later search stages and stage in task completion. However, subjects generally used more than one criteria to evaluate the relevance of a document. Statistical analysis of pairwise groups can provide some indication of which groups of criteria subjects are using to judge relevance, providing empirical results for analysis of research question 4. To determine groupings of criteria, factor analysis was conducted on the data collected to create several correlation matrices. These matrices revealed several groupings, which were examined for partial relevance judgments for all search stages, and for relevant document judgments for specific search stages.

5.5.1 Grouped Relevance Criteria for Partially Relevant Documents across All Search Stages

Table T-1 in Appendix R presents a Pearson correlation matrix from relevance judgments of "partially relevant" across all search stages. To create this matrix, if a subject selected one of the relevance criteria listed while making a relevance judgment of partially relevant, then a value of '1' was used, otherwise a value of '0' was used. The

correlation coefficient indicates the degree with which two variables covary, with a value of 1 being perfect correlation. Given the large number of criteria choices available for subjects to choose, and the number of search stages, the correlation coefficient generated from these results is rarely above .30, with a few returning results as high as .50. These values provide some measure of the correlation and thus indicate the paired groupings preferred by subjects. The correlation matrix was being used as a filter process for the analysis of the relationship between criteria groups and search stage. For this reason, the selection of a threshold value for selection was arbitrary and was only used to provide a smaller, more manageable set of criteria groups to examine using analysis of variance. After reviewing the results and determining the correlation coefficient commonly ranged between .05 and .30, the value of .15 was selected as a filter threshold since it appeared this would allow for the selection of a reasonable number of criteria sets to examine further.

The results in Table 5.26 lists the criteria with coefficients greater than or equal to .15 for partial relevance judgments. These results indicate that with the criteria of document "structure," subjects showed a tendency to identify "authority" and "accuracy" as document characteristics (criteria) which affected their relevance judgment. With the criteria of "source," the "novelty" and the "accuracy" of the document were often selected as a group, in addition to the "breadth" of information in the document. The criteria of "recency" and "novelty" were also selected as a group, in addition to the criteria of "breadth" and "depth."

Table 5.26: Correlation Coefficients >= .15 for Partial Relevance Judgments

Criteria	Coefficient
structure - authority	0.226
structure - accuracy	0.149
source - breadth	0.183
source - novelty	0.174
source - accuracy	0.211
recency - novelty	0.282
breadth - depth	0.129
depth - understanding	0.196
depth - amount	0.345

5.5.2 Grouped Relevance Criteria Selections for Relevant Document Judgments

Table T-2 in Appendix R shows a correlation matrix which examined relevance judgments of "relevant" across all search stages. This matrix generated a number of correlations equal to .15 or above listed in Table 5.27. The "ability to understand" was a frequently selected criteria with relevant documents as demonstrated in Table 5.15. With relevant documents, subjects grouped the "ability to understand" (understandability) the document with the criteria of "structure," "source," "depth," "breadth," and "amount of information." The criteria of "structure" grouped with "ability to understand," "depth," "breadth," "amount of information," and "accuracy." The criteria of "source" grouped with "authority," "amount of information" and "accuracy." Some groupings hint at similar concepts: "breadth" and "depth," "amount of information" and "depth." Other groupings hint at the characteristics of the subject's informational need: "recency" and "authority," and "novelty" and "authority." Examining these groupings by search stage will provide a more focused perspective on these relationships. That analysis is performed in the following sections.

Table 5.27: Correlation Coefficients > .15 - Relevant - All Stages

Criteria	Correlation
ability to understand - structure	.150

Criteria	Correlation
ability to understand - source	.155
ability to understand - depth	.155
ability to understand - breadth	.152
ability to understand - amount of information	.170
ability to understand - accuracy	.242
structure - ability to understand	.148
structure - depth	.174
structure - breadth	.158
structure - amount	.164
structure - accuracy	.186
source - authority	.157
source - amount	.165
source - accuracy	.187
recency - authority	.157
novelty - authority	.145
breadth - depth	.179
amount of information - breadth	.207
accuracy - breadth	.201
amount of information - depth	.289
accuracy - depth	.197

Generating a correlation matrix for relevance criteria used to judge documents relevant in the "initiation" stage identified several criteria with a correlation coefficient > .15 as shown in Table 5.28. The "ability to understand," a consistent selection with relevant documents, grouped with the criteria of "structure," "source," "recency," "depth," "breadth," "amount," and "accuracy." The criteria of "structure" grouped with "depth," "amount," and "accuracy." The criteria of "breadth" and "amount of information," and "depth" and "amount of information" had a very high coefficient suggesting a strong tendency for subjects to select these criteria together. The criteria of "source" grouped with a number of criteria including "authority."

Table 5.28: Correlation Coefficients > .15 - Relevant - Initiation Stage

Criteria	Correlation
ability to understand - structure	0.208
ability to understand - source	0.368
ability to understand - recency	0.364
ability to understand - depth	0.153
ability to understand - breadth	0.377
ability to understand - amount	0.363
ability to understand - accuracy	0.402
structure - depth	0.270
structure - amount	0.322
structure - accuracy	0.223
source - recency	0.155
source - novelty	0.176
source - depth	0.149
source - breadth	0.471
source - authority	0.168
source - amount	0.277
source - affectiveness	0.223
source - accuracy	0.385
recency - depth	0.179
recency - affectiveness	0.220
novelty - depth	0.216
depth - authority	0.174
depth - amount of information	0.352
depth - affectiveness	0.192
depth - accuracy	0.294
breadth - amount of information	0.483
breadth - affectiveness	0.237
breadth - accuracy	0.273
authority - affectiveness	0.485
amount of information - affectiveness	0.175
amount of information - accuracy	0.294

In examining the groupings of relevance criteria during the "exploration" stage, a number of relationships were suggested as shown in Table 5.29. Evaluation of this search stage produced fewer groupings with coefficients over a value of .15, an indication that

subject selections of criteria were more varied during this stage. Based on these results, subjects consistently grouped the criteria of "ability to understand" with the criteria of "structure," "authority," "amount of information," and "accuracy." Subjects also grouped the criteria of "source" with the criteria of "authority" and "accuracy." The categories of "authority" and "depth" had a very high correlation coefficient, indicating a strong tendency for subjects to select these categories together.

In examining groupings of criteria selections for the "extracting" stage, more groupings are found for "depth," "breadth," and "recency" in this stage. These criteria are listed in Table 5.30. Analysis revealed fewer groupings for the criteria of "ability to understand," and more groupings for "recency."

Table 5.29: Coefficient >= 0.15 - Relevant - Exploration

Criteria	Correlation
ability to understand - structure	0.227
ability to understand - authority	0.166
ability to understand - amount of information	0.245
ability to understand - accuracy	0.170
structure - breadth	0.245
structure - amount of information	0.159
structure - accuracy	0.152
source - authority	0.210
source - accuracy	0.159
breadth - accuracy	0.207
authority - amount	0.179
authority - depth	0.404

Table 5.30: Coefficient >= .15 - Relevant - Extracting

Criteria	Correlation
ability to understand - authority	0.201
ability to understand - accuracy	0.224
structure - recency	0.223
structure - accuracy	0.168
source - accuracy	0.158
recency - structure	0.169
recency - source	0.223
recency - breadth	0.176
recency - authority	0.266
depth- breadth	0.207
depth - authority	0.168
depth - amount	0.263
breadth - authority	0.189
breadth - amount of information	0.201
breadth - accuracy	0.197
accuracy - authority	0.151

In examining the "verifying" stage as shown in Table 5.31, the criteria of "ability to understand" and "amount of information" generate a high correlation coefficient. The same is true for "recency" and "breadth," and "accuracy" and "depth." The criteria of "recency" and "breadth" had a very high covariation coefficient, indicating that subjects had a strong tendency to select these criteria together. The additional groupings for the criteria of "structure" at this stage indicate the subjects had more interest in structure in groups with other criteria.

Table 5.31: Coefficient >= .15 - Relevant - Verifying

Criteria	Correlation
ability to understand - amount of information	0.309
structure - recency	0.155
structure - depth	0.285
structure - breadth	0.220
structure - amount	0.164
structure s- accuracy	0.246
source - recency	0.145
source - novelty	0.159
source - authority	0.339
source - amount	0.201
recency - novelty	0.182
recency - accuracy	0.156
recency - breadth	0.504
novelty - accuracy	0.241
amount of information - ability to understand	0.309
amount of information - structure	0.164
accuracy - depth	0.370

5.5.3 Grouped Relevance Criteria in Relation to Search Stage Progress

The results presented in previous sections provided an indication that subjects did show a tendency to select criteria in groups. An analysis of the frequency counts for specific groups of relevance criteria (Appendix S) identified several statistically significant changes in selection as shown in Table 5.32 (research question 4). These changes were found in comparing an early search stage (initiation + exploration) and a later search stage (extracting). This combination provides an approximately equal base of document selection for analysis, strengthening the case the subject is showing a preference for relevance criteria, not search stage.

Table 5.32: Significant Groupings - Early Search Stage to Later Search Stage

Criteria	χ^2	p
understandability + accuracy	5.12	< .05
amount of information + depth	9.31	< .001
amount of information + breadth	6.21	< .05
source + accuracy	2.4	=0.12
understandability + accuracy	14.91	< .001
understandability + amount of info.	8.58	< .001
source + amount of information	5.89	< .05
structure + amount of information	6.37	< .05
structure + accuracy	9.49	< .01
understandability + breadth	3.35	< .01
source + understandability	2.97	< .10
recency + source	5.77	< .05
recency + structure	12.94	< .001
recency + breadth	5.67	< .05
recency + authority	12.5	< .001
depth + amount	9.31	< .001
depth + authority	5.12	< .05
depth + breadth	8.07	< .01
breadth + amount	6.21	< .05
breadth + accuracy	3.2	< .10
accuracy + authority	8.26	< .001

5.6 Discussion

Study 1 examined the selection of relevance criteria across search stages and demonstrated that relevance criteria selections had a statistically significant relationship to search stage selection. As subjects progressed through the information search process, their preference for relevance criteria changed.

As detailed in this chapter, Study 2 expanded on the methods of Study 1 and collected additional data for the examination and analysis of the relevance criteria

selections. Study 2 methods refined the search task and used data collection methods that allowed subjects to work in a naturalistic environment over a period of several weeks. These methodological changes included interim deliverables to the research protocol, allowing the examination of relevance criteria selection in relation to stage in task completion. Analysis for Study 2 also examined the relationship of groups of relevance criteria selections to search stage. Findings include detection of several relationships between relevance criteria and search stage, relevance criteria and stage in task completion, and groups of relevance criteria.

5.6.1 Major Findings

The following were the major findings of this study.

1. Subjects demonstrated a preference for "accuracy," "recency," "amount of information," and "understandability" for partially relevant documents as shown in Table 5.9 (RQ1).
2. For partially relevant documents, the criteria of "recency" demonstrated a statistical relationship with search stage, decreasing in selection for partial relevance as the subject progressed through the search process (RQ3).
3. Subjects demonstrated a preference for the criteria of "understandability," "accuracy," "amount of information," "depth," and "recency" across all search stages for relevant documents as shown in Table 5.16 (RQ1).
4. The subjects selection of the criteria of "accuracy," "affectiveness," amount of information," authoritative," "bias," "breadth," "depth," "novelty," "recency," "source," "structure," "time," and "ability to understand" demonstrated a statistically significant relationship with search stage selection for all relevance judgments, indicating the subjects preference for these criteria fluctuated over the

course of the search process, increasing and decreasing, and these increases and decreases were statistically significant (RQ3).

5. Subject selections for the selection of the criteria of "affectiveness" and "amount of information" showed a statistically significant relationship with stage in task completion as shown in Table 5.25 (RQ2).
6. Subjects selections for a number of groups of relevance criteria demonstrated several statistical relationships to search stage at various levels of significance as shown in Table 5.32 (RQ4).

5.6.2 Detailed Discussion

Evaluation of the relevance judgment process in relation to search stage reveals changes in the type of relevance judgment as subjects progress through the search process. As the results in Table 5.11 and Figure 5.1 indicate, there is a continuous decrease in partial relevance judgments as subjects progressed through the search process. Evaluation of partial relevance judgments by deliverable due is shown in Table 5.23. Figure 5.4 does not show the same consistent rate of decline. Partial relevance judgments increase during the time period when the first and second deliverable are due, and then decrease for the preparation of the rough draft, and increase again in the preparation of the final deliverable. The frequency count for the selection of partially relevant documents increases for the final deliverable and is actually slightly higher than the count of the selection of partially relevant documents for the first deliverable.

These results add statistical strength to previous findings (Tang and Solomon, 1998; Spink et al, 1998) which suggested that searchers are more confident about their selections as they progress through the search process, and make fewer partial relevance judgments and more judgments of documents as being relevant to their information

search. The results concerning stage in task completion however, suggest a discontinuity with search stage progression and provides some confirmation of the iterative nature of the search process (search episode) as detailed in Chapter 3. Evidence of this is shown in Table 5.6 where the stages of "initiation" and "exploration" are reported during the preparation of the Rough Draft and Final Presentation stages in task completion. Subjects appear to be repeating the search process in later stages of task completion, and at the same time repeating the behaviors and progressing through the cognitive changes associated with the search process. If this is the case, then an analysis of the 'stage in task completion' relationships with relevance criteria should not demonstrate the same associations as criteria choices and information search process stage.

5.6.2.1 Partially Relevant Documents and Search Stage

As subjects progress through the search process, they make fewer partial relevance decisions and more decisions that a document is relevant (see Table 5.8). As stated previously, this is an indication that subjects are more focused in their search process and making more relevant or not relevant decisions. The results in Table 5.10 show that the criteria of "accuracy," "recency" and "understandability" are most important in the "initiation" search stage when most partial relevance judgments are made. These criteria decrease in importance in later stages, and for the criteria of "recency," this was a statistically significant decrease. Since these results are for partially relevant documents, and since the frequency count decreased, this finding suggests that partially relevant judgments based on "recency" are less important to subjects at this stage.

5.6.2.2 Relevant Documents and Search Stage

Relevant document judgments represent the majority of document relevance judgments throughout the search process, with a tendency to select more relevant

documents later in the search process (see Table 5.8). The results in Table 5.15 and Figure 5.2 indicate that the criteria of "understandability" ("ability to understand") is consistently selected throughout all search stages with relevant document judgments. With the exception of the "differentiating" search stage, it is the most common selection. The second most common selection varies somewhat, but "accuracy" appears to be a common selection in most search stages. The criteria of "amount of information" is selected most often in the extracting stage.

These findings are an indication that subjects are consistently interested in the "accuracy" of information, and must be able to "understand" what they are selecting. Looking at the third or fourth most frequent selection provides some indication of the how the level of importance of criteria varies across search stages. This analysis reveals that subjects are interested in "recency," "structure [of the document]," "depth," and "breadth."

Examining the changes in selection across search stages provides insights into the cognitive changes of subjects as they progress through the search process. The criteria of "accuracy," "amount of information," "bias," "depth," "novelty," "source," "structure," "time," and "understandability" (ability to understand) all demonstrated a statistical variation in comparing an *early* search process stage (initiation, exploration, differentiating) to the *later* stage of "extracting" as shown in Table 5.18. This adds statistical strength to the finding that subjects are interested in a larger number of criteria in later stages as they become more discerning in their document selection.

Table 5.15 shows that the tendency to select "depth" peaks in the "differentiating" stage. The criteria of "source" [quality] is important in the "initiation" stage, and then less important in some later stages, but then appears to increase in importance in the

"extracting" stage. The criteria of "amount of information" increases in early stages and peaks in the "extracting" stage. The criteria of "accuracy" also increases in later stages. The "ability to understand" the document appears to be important to subjects throughout the search process with only slight changes in preference across some search stages. These results provide some indication that subjects are more discerning in making relevance judgments in later search stages, and the criteria that interests them in most in later stages are "source" [quality], "accuracy," "amount of information," and "ability to understand."

It is interesting to note the relative importance of some of the criteria in the "verifying" stage. In the process of completing their research project, the criteria of "structure," "amount of information," and "accuracy" based on frequency of selection, appear to be important to subjects. Though "structure" is not always the most frequently selected criteria, it is consistently selected at 6 or 7 percent of all criteria selected. This finding suggests that subjects using the web for searching must contend with various "document structures" which have an impact on their relevance judgment. Perhaps some structures make it easier to find information, and other structures make it difficult. Ultimately this structure, like the breadth and depth of the document, influences the subject's relevance decision.

5.6.2.3 Non-Relevant Judgments and Search Stage

There were no statistically significant changes found in examining "not relevant" document assessments. These judgments were a small portion of the overall document judgments, varying between 5 and 11 percent of the relevant document judgments across all search stages as shown in Table 5.21. It appears that subjects in this sample were reluctant to report "not relevant" judgments, or few "not relevant" documents were

reviewed. Based on the data collected here, it is not possible to determine which is the case.

The criteria which were consistently important in making "not relevant" judgments are shown in Table 5.22. These results show that subjects making "not relevant" decisions commonly base those decisions on the "amount of information," and the "depth" of the information in the document. The "structure" of the document is also important in these decisions. In contrast to "relevant" document decisions, the "ability to understand" does not rank high among the list of criteria commonly selected in "not relevant" decisions. This finding is an indication that subjects rejecting a document have made the decision to do so based on other criteria and have not evaluated the document's understandability.

5.6.2.4 Partially Relevant Documents and Deliverable Due

The subjects in this study reported a variety of search stages during the stages of task completion, demonstrating an iterative progression through the search process as they worked to complete their research task. Table 5.6 shows that the "initiation" stage is reported by subjects during the preparation of the rough draft and final presentation, the two final stages of the research project assigned to the subjects. This indicates that subjects appear to be repeating the information search process within stages in task completion. Examining the interaction of criteria selections and task completion stage should therefor provide results different from those of search stage interactions.

In examining partial relevance judgments during the preparation of the detailed outline deliverable, the criteria of "accuracy," "ability to understand," and "recency" are important based on frequency of selection. Moving to the final presentation, there is a statistically significant decrease in the selection of all criteria for all stages in task completion. When examined discretely, the decrease in importance is statistically significant for "recency" and "ability to understand." This adds clarity to the finding that subjects are more discerning in later stages and make fewer partial relevance judgments based on all criteria, and specifically based on the criteria of "recency" and "ability to understand."

5.6.2.5 Relevant Documents and Deliverable Due

Examination of the criteria used to make relevant document selections across stage in task completion provides evidence of significant variability in the selection of criteria. The criteria of "ability to understand," "accuracy," "structure," "recency," "depth," and "amount of information" were all important criteria for the subjects (see Table 5.24). These results are consistent with search stage - criteria analysis interactions reported earlier. In moving across all stages in task completion (as identified by deliverables), the criteria identified in Table 5.25 all demonstrate statistically significant changes as the subject progressed towards completion of their work task. The importance of the criteria of "novelty" adds statistical strength to previous findings by Vakkari (2001) and others that subjects look for new sources later in the search process. These statistical results demonstrate that "breadth," "amount of information," "structure," "authority," and "ability to understand" increase in importance as subjects progress through the search process. These results strengthen the results shown in Table 5.18 which identified statistically significant variations in the selection of many of the same criteria. Since all

the criteria listed in Table 5.25 increased in frequency of selection, these results indicate that as subjects moved from the detailed outline, rough draft, to the preparation of the final presentation, these criteria increased in importance based on frequency of selection. Subjects were looking for new documents (documents they did not currently have), documents which could be understood, and were recent. (Since the topics assigned were related to computer technology which changes rapidly, the selection of current material was important.) The "structure" of the document was also important to subjects. It appears that for this sample, some web documents have a better "structure" which make them easier to gather information and understand their topic better.

5.6.2.6 Grouped Relevance Criteria - Partially Relevant Documents

Previous findings established the increase in importance of single criteria selections during both search stage progression and progression through the work task. Analysis for research question 4 demonstrated that subjects were more likely to select multiple criteria for document selection, and had some tendency to select these criteria in groups. A correlation matrix was used as a filter to determine the groups of criteria subjects selected during the search process and provide empirical results to answer research question 4. These results revealed a number of groupings for partially relevant document selections as shown in Table 5.26. These choices provide some indication that subjects do not base their relevance decision on a single criteria, but instead some combination of criteria are used. In examining partial relevance over all search stages, it appeared that subjects were basing their decision based on a group of criteria which included "structure," "authority", and "accuracy." The quality of the source was also important to subjects as grouped with "breadth," "novelty", and "accuracy." The criteria

of "depth" was grouped with "understanding" (ability to understand) and "amount of information."

Subjects also chose "recency" and "novelty" in pairs, suggesting that at some point in the search process, both the "recency" of the document and the fact that the document is new are important to the subject. The "depth" of the document and the "amount of information" also grouped with a very high coefficient, possibly indicating that these are similar concepts to the subject.

Since partial relevance decisions are more likely to be made early in the search process rather than later, these findings provide some indication of the groups of criteria used by subjects earlier in the search process. Subjects are looking for documents with a structure that makes it easier to understand, and authoritative, accurate documents, from current quality sources.

5.6.2.7 Grouped Relevance Criteria - Relevant Documents

Examining the criteria choices groups for relevant document judgments over all search stages reveals a more varied set of criteria groups as shown in Table 5.27. The groups of "depth" and "breadth," "amount of information" and "breadth," and "amount of information" and "depth" suggest that these are similar concepts to the subject. Other groups of the criteria of "structure," "source," and "ability to understand" suggest these document characteristics are important to subjects in groups. An analysis by search stage was used to provide additional insights into how these groups relate to the subject's progress through the information search process.

5.6.2.8 Grouped Relevance Criteria - Initiation Search Stage

Analysis for criteria groups selected with relevant document judgments during the "initiation" stage revealed a number of groups as shown in Table 5.27. The ability to

understand a document, the structure of the document (which relates to understandability), and various characteristics of the source of the document appear to have an impact on the relevance decision. The criteria of "source" [quality], "breadth," "amount of information," and "accuracy" all grouped with "ability to understand" with a high coefficient, indicating a stronger tendency for subjects to select these groups, and indicating that these criteria are important to subjects in combination with the ability to understand a document. This finding provides some indication that subjects care about the depth and amount of information in the document, and as would be expected, the ability to understand that information. Subjects also demonstrated a preference for the criteria of "structure" (easy to understand) and "amount of information."

The source quality in combination with accuracy, and breadth were also important to subjects. As subjects began their search they wanted to find documents from a source with a good reputation, and they wanted the document to be accurate, and have sufficient depth to help them in their "exploration" of the topic.

Subjects also selected "depth" and "amount of information," and "breadth" and "amount of information," suggesting that these criteria relate to similar concepts for the subject. Results also indicated that the criteria of "authority" and "affectiveness" had a high correlation coefficient, indicating the subjects wanted a document which was both "enjoyable," and from an authoritative source.

5.6.2.9 Grouped Relevance Criteria - Exploration Search Stage

In the exploration stage, subjects are "reading to learn about the topic." The user behavior in this stage is similar to that of the "initiation" stage, so it is not unexpected that the results summarized in Table 5.29 are similar to those of the previous search stage. Subjects were interested in the "ability to understand" a document, and the "structure" of

the document (related to understandability) was important. The authority of the source continues to be important, and grouped closely with the "depth" of the document.

5.6.2.10 Grouped Relevance Criteria - Extracting Search Stage

In the "extracting" stage, the subject is retrieving information and preparing their report. Results reveal a slightly different mix of criteria groupings as shown in Table 5.30. A large number of criteria selections are made in this stage, and there appears to be more variability in what the subjects were selecting. The "ability to understand" in combination with authority and accuracy are important, but other criteria such as "recency" and "depth" and "breadth" have more groupings with higher coefficients than in previous search stages. The criteria of "recency" groups with "source" and "authority," indicating that subjects are looking for recent documents from a reputable, authoritative source. The "depth" and "breadth" of the document continue to be important. This is an indication that in this stage, subjects prefer a recent and authoritative document, with depth and breadth of coverage on their topic.

5.6.2.11 Grouped Relevance Criteria - Verifying Search Stage

In the verifying stage, where subjects check and validate sources, a number of groupings are found. The "recency" of the documents demonstrates a strong grouping with the "breadth" of the document, suggesting that subjects are looking for a document with current information and sufficient depth. They are also looking for documents which are both quality sources and authoritative. The "amount of information," and the "ability to understand" the document are both important. The 'accuracy" of the document together with the depth of the document are also important to subjects. It is interesting to note that document "structure" appears to be more important at this stage based on the number of groupings. Various criteria such as "depth," "source," "recency," "breadth," "amount of

information," and "accuracy" are important to subjects in this stage. This may be a reflection of the difficulty of document structure with web documents, where many documents are pages with links to information sources which lack the depth and breadth subjects are seeking.

5.6.2.12 Grouped Relevance Criteria in Relation to Search Stage Progress

Previous analysis established various pairwise groupings for relevance judgment types during search stages. Additional analysis was performed to determine whether these criteria groups showed a statistically significant increase in importance across search stages as shown in Table 5.32. As with previous analysis, subjects appear to be focusing more on accuracy and understandability earlier in the search process (see Table 5.15), but their interest changes in later stages as indicated by the preference for depth, breadth, and amount of information while still retaining their interest in the accuracy of the document. Previous analysis did not indicate that subjects showed a significant interest in recency of the document in later stages, but when combined with the criteria of authority, breadth, and structure, a significant increase in preference is found. This indicates that the currency of the document is of interest to the subject (given the technical nature of the assignment in this study, that is expected since computer technology changes quickly), and the interest in that criteria is closely related to the authority of the source of the document, and the amount of information and breadth of coverage in that document.

Chapter 6 - Research Study 3

Study 1 laid the groundwork for the examination and exploration of the selection of relevance criteria across search stage. Study 2 expanded on Study 1 and examined the selection of criteria across search stage, and stage in task completion. The examination of groups of relevance criteria in relation to search stage was also part of Study 2.

The goal of Study 3 was to collect additional data to explore the relationship between the selection of relevance criteria and progress through the information search process. Data was collected to examine relevance criteria selections across search stage and stage in task completion, and groups of relevance criteria were analyzed. The data collected was relevant to all research questions.

There were methodological differences from Study 2 in the selection of relevance criteria. In Study 3, subjects identified a weight to indicate the importance of the relevance criteria, and identified the most important/ least important relevance criteria. Other methodological differences are detailed below.

6.1 Summary Methodological Differences from Study 2

The methods and analysis for this study were the same as those for Study 2 with the following changes.

The criteria selection available to the subjects was modified from the criteria used in Study 2 by adding additional criteria related to technical documents ("tips," "descriptions," "guidelines," and "history" from Table 3.1), and eliminating the criteria of "amount of information" as a duplicate of the criteria of "depth" and "breadth" (see Appendix T).

The protocol for criteria selection was changed. In this study, subjects were required to indicate the level of importance for a criteria selected (one of "not very

important," "slightly important," "somewhat important," "important," or "very important"), and were required to indicate which criteria was "most important," and which criteria was "least important" for a particular document (see Appendix H).

6.2 Detailed Explanation of Methods

The research was conducted in a naturalistic environment where subjects worked on a real information problem. The following section provides details on the methods employed in this study. In this section, each research question is addressed using the abbreviation RQN where, N is one of the four research questions identified previously.

Fifty-three subjects, were drawn from a convenience sample of junior and senior business students at an American university. Subjects were undergraduate students at a university and were not experts in their assigned subjects. A research project was assigned and their progress on gathering information for their research was used for data collection (see Appendix A). Subjects chose their own research topic from a list of topics provided, and topics were of the same level of difficulty for subjects. The student's participation as a subject was voluntary. Data was collected anonymously using a Web application, thus allowing subjects to work in a naturalistic environment.

As subjects conducted their search, they made *relevance judgments* (relevant, not relevant, partially relevant/not sure about relevance), identified a *stage in the information search process,* and selected relevance criteria from a predetermined list of relevance criteria (see Table 3.1). Subjects in Study 3 assigned a weight indicating the importance of each relevance criterion selected. The data collected was relevant for all research questions.

As data was collected, a time-stamp recorded the date and time the data was stored. When the date the data was stored is examined relative to the interim project

deliverables required from the subjects, the results can provide an analysis of relevance judgments and relevance criteria choices relative to *stage in task completion* (for RQ2). The deliverables due for the assigned research project were 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.

Data was collected using a Web site accessible from the Internet, allowing subjects to conduct their searches in a naturalistic setting. The website used an interface similar to that of commercial search engines such as Yahoo! but with required inputs for search stage, relevance judgment, relevance criteria, and relevance criteria weight. Subjects in Study 3 also indicated the least and most important criteria of all criteria selected. Subjects were provided brief training sessions on the use of the site, and help pages were available for those subjects who needed them (see Appendix J and Appendix K).

6.2.1 Search Process

The search stages presented to subjects in Study 2 were based on Kuhlthau's (1993) ISP stages, and Ellis's (1997) search patterns as synthesized by Wilson (1999), as summarized in Table 3.4 and Table 5.1.

6.2.2 Relevance Criteria

Study 3 used the relevance criteria identified for Study 2 as shown in Table 3.1 and Table 5.2. These criteria combined the relevance criteria identified by Barry (1994), Barry and Schamber (1998) and have been found to be consistent selections by subjects across studies. This study also included several of the criteria of Cool et al. (1993); these

technology specific criteria were not used in Study 2, but were included in this study to allow subjects to choose more technology specific criteria. Subjects in this study also assigned a level of significance to each criterion they selected. This level of significance was input as one of 'not very important,' 'slightly important,' 'somewhat important,' 'important,' 'very important.' Subjects also indicated which criteria were most important and which were least important in the evaluation of the document.

6.2.3 Data Collection Process

Study 3 was conducted over two 6 1/2 week semesters in the summer of 2008. The convenience sample used was comprised of undergraduate students at an American University.

The process of searching for information and reporting on results of the search as part of this research study was as follows.

- Subjects logged into the research Web site to conduct their search using a login ID previously assigned, and a personal password they had chosen.
- Subjects entered search query terms as if they were using a commercial search engine such as Yahoo! and executed a search (see Appendix D).
- The research Web site generated 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 included an explanation of the result page/document (as returned by the commercial search engine), links to the results page, and links which the user used to enter relevance information about the page (see Appendix E).
- Subjects were 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. Relevance assessments were one of either *relevant*, *not relevant*, or *partially relevant/ not sure* (from Greisdorf, 2003; Spink et al., 1998).

- Subjects selected one or more relevance criteria which were considered by the subject in making that judgment. For each relevance criteria chosen, subjects indicated the significance of that criteria in making the relevance judgment. For each document judged, subjects indicated which criteria was *most* important and which criteria was *least* important in making that judgment (see Appendix H).
- Results of the data collection process were submitted to the data collection program on the research Web site which stored the results anonymously for later analysis.
- Subjects repeated the process outlined above as often as they felt necessary to gather the information they needed to complete their assigned report.
- A post-test questionnaire was used to collect demographic information for reference purposes (Appendix M).

6.3 Results of Data Collection

Study 3 data was collected in two 6 week sessions, one in June and one in July, and then pooled for analysis. A total of 657 documents were reviewed by the 53 subjects who participated in the research.

Approximately 66% of the 53 subjects responded to a survey used to gather background information on the subjects. The results of this survey indicated that 46% of the subjects were frequent users of search engines, performing searches several times a day. Based on the survey, most subjects were female (74%). Approximately 40% of the subjects participating did not report English as their primary language.

The count of documents reviewed by subjects by search stage is listed in Table 6.1. Subjects in this study selected a consistent number of documents through the "exploration" stage, but expanded the number of documents selected in the "verifying" stage. As in previous studies, subjects appeared to iterate through the search process stages throughout the duration of the study as shown in Table 6.2. For example, the subjects reported being in the "initiation" search stage late in the search process during the preparation of the Final Presentation deliverable.

Table 6.1: Unique Documents Selected by Search Stage

Search Stage	Count
initiation	113
differentiating	128
exploration	140
extracting	250
verifying	90
Total	721

Table 6.2: Search Stage Selections by Stage in Task Completion (Deliverable Due)

Search Stage	Abstract	Detailed Outline	Rough Draft	Final Presentation
initiation	250	54	110	241
differentiating	294	29	37	85
exploration	199	315	329	196
extracting	213	123	308	602
verifying	84	20	88	68
Total	1040	541	872	1192

When subjects selected a criteria, they were asked to indicate the level of importance for that criteria. The majority of subject selections (73%) was at the level of "somewhat important" or higher as shown in Table 6.3. Subjects also tended to select documents with the higher levels of importance later in the search process as shown in Table 6.4 and 6.5. Since the purpose of this study was to examine the criteria which were important to subjects, the focus of the analysis was on the selections with the higher levels of importance.

Table 6.3: Criteria Code Level of Importance - Frequency of Selection

Criteria Code Weight	Count	Percentage
not very important	124	11.31%
slightly important	171	15.60%
somewhat important	192	17.52%
important	387	35.31%
very important	222	20.26%
Total	1096	100.00%

Table 6.4: Criteria Code Level of Importance Frequency of Selection by Search Stage

Search Stage	Not very impt	slightly impt	somewhat impt	important	very important
initiation	71	62	48	109	75
differentiating	30	37	79	146	160
exploration	22	54	78	216	147
extracting	42	71	169	448	283
verifying	17	52	63	122	58
Total	182	276	437	1041	723

Table 6.5 and Figure 6.1 relate criteria selections with level of importance selections by subjects and provides a further indication that subjects had a tendency to select criteria using higher levels of importance. As can be seen in Figure 6.1, the most common level of importance for criteria is "important" with the exception of "accuracy," and "authority" where "very important" is the most common selection.

Table 6.5: Criteria Code by Importance Weight - All Relevance Types

Criteria	not very impt	slightly impt	somewhat impt	important	very impt
accuracy	6	8	10	85	94
advertisement	24	8	12	8	9
affectiveness	36	21	41	32	12
authority	3	7	16	27	29
bias	8	8	11	22	8
breadth	4	12	15	58	36
definitions	4	12	22	40	18
depth	6	23	29	88	45
descriptions	1	10	11	75	29
guidelines	2	8	19	31	10
history	4	4	20	29	16
novelty	4	19	28	59	54
recency	8	44	29	80	57
source	7	10	16	49	31
structure	7	20	35	57	14
time	6	8	12	21	14
tips	14	9	22	31	6
understandability	15	28	42	70	31
Total	159	259	390	862	513

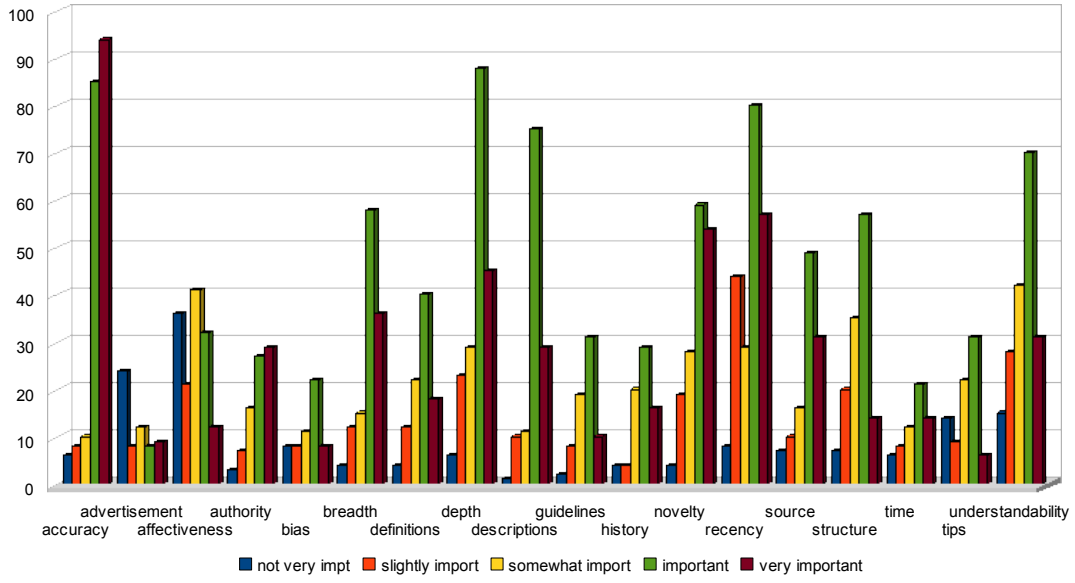


Figure 6.1: Criteria Level of Importance

As in Study 2, subjects had a tendency to select most documents during the "extracting" stage (see Table 5.5). This trend is apparent when selection counts are

evaluated with all relevance types as shown in Table 6.6, and is more pronounced with relevant document types as shown in Table 6.7 and Figure 6.2. But the trend is not apparent when partially relevant document types are evaluated. Subjects tended to select partially relevant documents at a consistent rate throughout the search process as shown in Table 6.8 and Figure 6.3.

Table 6.6: Critieria Selected by Stage in Search Process - All Judgment Types

Search Stage	Count	Percentage
initiation	365	13.73%
differentiating	452	17.00%
exploration	517	19.44%
extracting	1013	38.10%
verifying	312	11.73%
Total	2659	100.00%

Table 6.7: Criteria Selected by Stage in Search Process - Relevant Document Types

Search Stage	Count	Percentage
initiation	179	10.04%
differentiating	282	15.82%
exploration	293	16.44%
extracting	811	45.51%
verifying	217	12.18%
Total	1782	99.99%

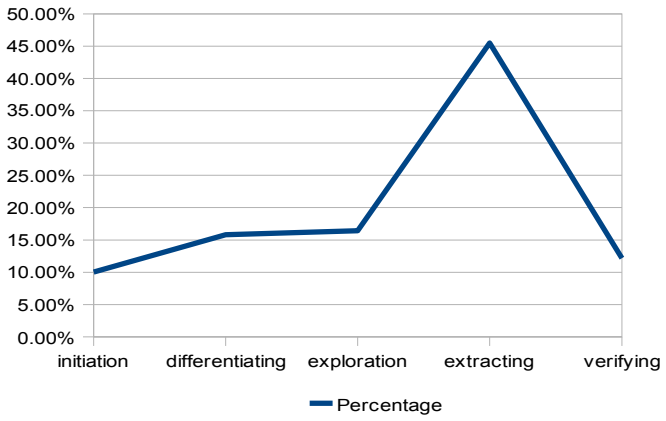


Figure 6.2: Documents Selected by Stage in Search Process - Relevant Document Types

Table 6.8: Documents Selected by Stage in Search Process - Partially Relevant Documents

Search Stage	Count	Percentage
initiation	15	11.11%
exploration	34	25.19%
differentiating	31	22.96%
extracting	36	26.67%
verifying	19	14.07%
Total	135	100.00%

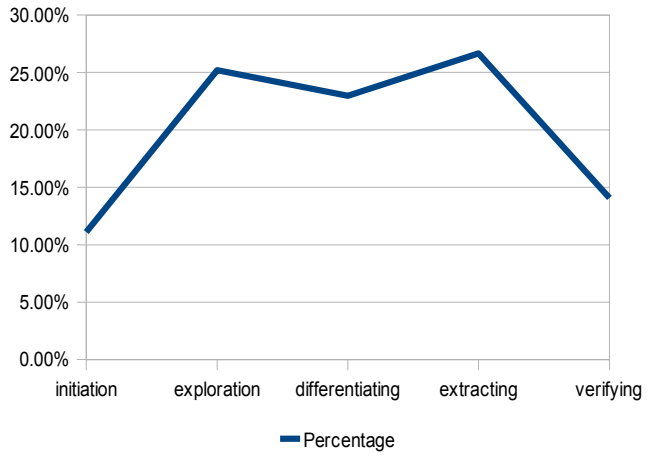


Figure 6.3: Documents Selected by Stage in Search Process - Partially Relevant Documents

When examining the weights assigned to the selection of criteria differ for high and (subjects indicated they were "important" or "very important") low relevance (subjects indicated they were "not very important" or "slightly important") a pattern emerges as shown in Table 6.9. Subjects were more likely to select a high level of importance in the "exploration" and "extracting" stage, and less likely to select a high level of importance in the "initiation" and "verifying" stage.

Table 6.9: Search Stage Selections - High and Low Levels of Importance

Search Stage	High	Pct High	Low	Pct Low
initiation	123	68.72%	56	31.28%
differentiating	214	75.89%	68	24.11%
exploration	234	79.86%	59	20.14%
extracting	597	73.61%	214	26.39%
verifying	143	65.90%	74	34.10%

Table 6.10 lists the selection of relevance criteria across all search stages. This provides some indication of the importance of relevance criteria based on frequency of selection. In this sample, subjects preferred the criteria of "accuracy" and "depth," followed by "novelty," "recency," and "understandability."

Table 6.10: Criteria Selections in all Search Stages

Criteria	Count	Percentage
accuracy	150	11.69%
depth	118	9.20%
novelty	111	8.65%
recency	110	8.57%
understandability	104	8.11%
descriptions	97	7.56%
structure	83	6.47%
breadth	75	5.85%
source	72	5.61%
definitions	65	5.07%
affectiveness	59	4.60%
authority	56	4.36%
history	43	3.35%
guidelines	41	3.20%
tips	34	2.65%
time	32	2.49%
bias	25	1.95%
advertisement	8	0.62%
Total	1283	100.00%

These results reveal consistent selectivity for certain criteria, but they do not examine the relationship of search stage to criteria selection. The results in Table 6.11 show the percentage of selections for a criteria within a search stage, and provides a rank for those selections. These results indicate that there is variation in the selection of

criteria within search stage, with criteria of "accuracy" and "depth" consistently ranking high in the list.

Table 6.11: Criteria Selection by Search Stage - All Relevance Types

Criteria	initiation	differentiating	exploration	extracting	verifying
accuracy	7.12%(3)	8.63%(1)	7.74%(1)	7.80%(2)	6.09%(4)
advertisement	3.84%(5)	1.99%	2.32%	2.17%	1.28%
affectiveness	3.84%(5)	7.52%	4.84%	5.33%	4.81%
authority	3.56%	3.10%	2.13%	2.67%	5.45%(5)
bias	3.29%	1.77%	1.74%	1.97%	2.56%
breadth	5.21%(4)	6.42%(4)	5.61%	3.85%	2.24%
definitions	2.19%	3.54%	4.26%	3.75%	3.85%
depth	7.12%(3)	8.41%(2)	7.54%(2)	6.22%(4)	7.69%(3)
descriptions	2.19%	4.42%	4.06%	6.02%	5.13%
guidelines	1.37%	2.21%	4.06%	2.96%	1.28%
history	4.11%	3.10%	4.06%	1.78%	1.60%
novelty	4.66%(4)	6.42%(5)	6.19%(4)	6.52%(5)	6.41%
recency	10.96% (1)	6.86%(3)	6.77%(3)	7.90%(1)	10.26%(1)
source	4.11%	3.10%	4.06%	4.54%	5.45%(5)
structure	4.66%(4)	4.20%	4.06%	6.12%	4.49%
time	2.47%	2.88%	2.13%	2.17%	1.92%
tips	2.74%	2.43%	3.87%	3.16%	2.88%
understandability	7.95%(2)	4.65%	6.00%(5)	7.40%(3)	9.29%(2)

These results provide some evidence of which criteria were consistently important to subjects across search stages, but additional analysis is required to provide a better understanding of the relationship between criteria and search stage, or stage in task completion. As part of that analysis, the changes of frequency of selection for specific relevance criteria will be examined using descriptive statistics, and using more rigorous analysis of variance methods.

Table 6.12 lists the type of relevance judgments made by subjects. While relevant document judgments represent the bulk of judgments made by subjects, other categories of relevance represent approximately 1/3 of judgments. Since these judgments represent distinct cognitive states, they should be examined separately to develop a clearer

understanding of the subject's behavior during that stage. This will be done in the following sections.

Table 6.12: Relevance Judgment Type Selected

Relevance	Count	Percentage
not relevant	271	10.18%
partially relevant	604	22.70%
relevant	1786	67.12%
Total	2661	100.00%

6.3.1 Partial Relevance Judgments

Table 6.13 lists the criteria selected for documents judged "partially relevant" by search stage as a percentage of criteria judgments within that search stage, and reporting all importance weights. These results indicate that subjects have a preference for some criteria over other criteria, and these preferences change depending on the search stage. Some changes in preference appear to be significant, for example "ability to understand" (understandability) is selected at a rate of 12.40% as a percentage of selections in the "initiation" stage, and is only selected at a rate of 2.40% as a percentage of criteria selected in the "differentiating" stage. The criteria of "depth" is selected at a rate of 10.58% as a percentage of criteria selections in the "initiation" stage, but is selected at a rate of 6.52% as a percentage in the "extracting" stage.

Table 6.13: Partial Relevance Analysis - All Relevance Weights

Criteria	initiation	differentiating	exploration	extracting	verifying
accuracy	5.43%	8.11%(1)	5.07%(4)	6.90%(3)	7.50%(4)
advertisement	5.43%	3.60%	4.35%	5.52%(4)	1.25%
affectiveness	1.55%	8.11%(1)	5.80%(3)	2.76%	12.50%(2)
authority	3.10%	5.41%(5)	2.17%	0.00%	3.75%
bias	2.33%	3.60%	2.90%	3.45%	5.00%(5)
breadth	4.65%(4)	7.21%(2)	1.45%	2.76%(5)	1.25%
definitions	3.10%	2.70%	3.62%	4.14%	3.75%
depth	10.85%(3)	3.60%	6.52%(2)	3.45%	5.00%(5)
descriptions	1.55%	5.41%(5)	3.62%	4.14%(5)	5.00%(5)
guidelines	0.00%	2.70%	4.35%	4.14%(5)	0.00%
history	4.65%	0.90%	1.45%	2.07%	1.25%
novelty	3.10%	6.31%(4)	7.97%(1)	5.52%(4)	5.00%(5)
recency	14.73%(1)	8.11%(1)	7.97%(1)	12.41%(1)	10.00%(3)
source	3.88%	3.60%	4.35%	5.52%(4)	3.75%
structure	2.33%	1.80%	5.07%(5)	5.52%(4)	3.75%
time	1.55%	4.50%	3.62%	2.07%	1.25%
tips	1.55%	6.31%(3)	6.52%(2)	3.45%	3.75%
understandability	12.40%(2)	2.70%	5.07%(4)	8.28%(2)	13.75%(1)

These results provide some indication of an association between criteria selection and search stage for partial relevance judgments, but additional analysis is needed to establish stronger statistical significance. Table 6.14 lists the frequency counts for specific criteria for relevance judgments for documents identified by subjects as "important" and "very important."

Since the goal of this analysis is to test for variance in preferences for criteria within a search stage, selecting search stages where a roughly equivalent number of documents were selected would increase the likelihood that strong statistical associations were due to preferences for criteria and search stage, and not just a general preference for search stage. The "exploration" and "extracting" stages provide roughly equal numbers of document evaluations and were therefore used to perform the variance test whose results are reported in Table 6.15 These results indicate that these criteria become more

important to subjects moving from the "exploration" to the "extracting" stage and that this change is statistically significant for most of the criteria.

Table 6.14: Criteria Selections for Partial Relevance Judgments*

Stage	accuracy	affective	authority	descrip	guides	hist	novelty	recency	source	struct	time	understand
initiation	14	6	6	6	1	3	9	9	5	9	2	7
different	26	3	4	14	2	5	20	12	8	10	5	10
exploration	29	6	4	15	6	9	16	14	9	5	3	15
extracting	65	14	21	49	17	13	42	45	29	29	10	37
verifying	12	2	9	7	3	2	7	14	12	3	2	8
Total	146	31	44	91	29	32	94	94	63	56	22	77

* criteria selected as "important," and "very important"

Table 6.15: Comparison of Exploration and Extracting Stage Selections for Partial Relevance Selections

Criteria	χ^2	p
accuracy	13.79	< .001
affectiveness	3.2	< .10
authority	11.56	< .001
descriptions	18.06	< .001
guidelines	5.26	< .05
novelty	11.66	< .001
recency	16.29	< .001
source	10.53	< .001
structure	16.94	< .001
time	3.77	< .10
understandability	9.31	< .001

6.3.2 Relevant Judgments

Table 6.16 lists the selection for criteria which were given a level of importance of "somewhat important," "important," and "very important" for documents judged to be relevant. Table 6.18 lists the percentage of each criteria selection within that search stage.

These results indicate the criteria of "accuracy" was consistently considered important to subjects across all search stages. The criteria of "authority," "structure,"

"recency," and "guidelines" appeared to increase in importance to subjects as they progressed through the search process. These results provide some indication of the changes in the selection of criteria across search stages. A statistical variance analysis of the frequency counts as a whole is just outside of the standard threshold for statistical significance.

Table 6.16: Criteria Selection by Search Stage - Relevant Documents *

Criteria	initiation	differentiating	exploration	extracting	verifying
accuracy	14	26	30	67	13
advertisement	1	3	0	2	2
affectiveness	6	7	11	33	2
authority	7	7	5	27	10
bias	4	3	5	11	2
breadth	6	16	18	31	4
definitions	3	11	16	26	9
depth	6	28	14	51	19
descriptions	6	14	16	52	9
guidelines	2	4	9	22	4
history	7	9	11	13	3
novelty	11	20	18	51	11
recency	11	14	14	51	20
source	6	10	11	33	12
structure	9	15	10	43	6
time	3	6	3	16	4
tips	5	1	6	17	5
understandability	9	13	16	53	13
Total	116	207	213	599	148

* ranked with a level of importance of "somewhat important" or higher

When criteria are evaluated individually over all search stages, a number of statistically significant results are found as shown in Table 6.17. These results provide some indication of variance, but the tendency to select more documents in general in the "extracting" stage provides a partial explanation for the variance found. An examination of a more equivalent set of selections would strengthen these findings.

Table 6.17: Variance Analysis for Criteria Across All Search Stages - Relevant Documents

Criteria	χ^2
accuracy	64.33
affectiveness	51.09
authority	29
bias(1)	10
breadth	31.2
definitions	22.92
depth	50.56
descriptions	71.77
guidelines	32.29
history(2)	6.88
novelty	49.68
recency	49.73
source	31.47
structure	55.01
time	18.94
understandability	63.5

* df=4, $p < .001$ except (1) $p < .10$, (2) $p=.1422$

In examining the results for changes in preference for criteria based on frequency counts, a comparison of *early* results (initiating + differentiating) versus *late* (extracting) was used to provide an even base of document selections for relevant documents. When early selections are compared with late selections, a number of statistically significant associations are found as shown in Table 6.19. These results further strengthen the association between criteria selection and search stage with this sample.

Table 6.18: Criteria Selection by Search Stage - Relevant Documents *

Criteria	initiation	differentiating	exploration	extracting	verifying
accuracy	10.00%(1)	10.32%(2)	11.19%(1)	9.22%(1)	6.99%(3)
advertisement	0.71%	1.19%	0.00%	0.28%	1.08%
affectiveness	4.29%(5)	2.78%	4.10%	4.54%(5)	1.08%
authority	5.00%(4)	2.78%	1.87%	3.71%	5.38%
bias	2.86%	1.19%	1.87%	1.51%	1.08%
breadth	4.29%(5)	6.35%(4)	6.72%(2)	4.26%	2.15%
definitions	2.14%	4.37%	5.97%(3)	3.58%	4.84%
depth	4.29%(5)	11.11% (1)	5.22%(4)	7.02%(3)	10.22%(2)
descriptions	4.29%(5)	5.56%	5.97%(3)	7.15%(2)	4.84%
guidelines	1.43%	1.59%	3.36%	3.03%	2.15%
history	5.00%(4)	3.57%	4.10%(5)	1.79%	1.61%
novelty	7.86%(2)	7.94%(3)	6.72%(2)	7.02%(3)	5.91%(5)
recency	7.86%(2)	5.56%(5)	5.22%(4)	7.02%(3)	10.75%(1)
source	4.29%(4)	3.97%	4.10%	4.54%(5)	6.45%(4)
structure	6.43%(3)	5.95%	3.73%	5.91%(4)	3.23%
time	2.14%	2.38%	1.12%	2.20%	2.15%
tips	3.57%(5)	0.40%	2.24%	2.34%	2.69%
understand	6.43%(3)	5.16%	5.97%(3)	7.29%	6.99%(3)

* ranked with a level of importance of "somewhat important" or higher

Table 6.19: Variance Analysis for Criteria Across *Early* versus *Late* Search Stages - Relevant Documents*

Criteria	χ^2
accuracy	6.8131
affectiveness	8.6957
authority(1)	4.122
definitions(2)	3.6
depth	3.4
descriptions	14.22
guidelines	9.1429
novelty(1)	4.878
recency	8.8947
source(1)	5.898
structure(1)	5.3881
understandability	12.8133

* df=1, $p < .001$, except (1) $p < .10$, (2) $p < .05$

Table 6.20 lists the percentage of documents identified as 'most important' by the subjects in each search stage. Based on frequency counts, the criteria of "accuracy" appears to be consistently selected as "most important," but there are numerous shifts for second and third rankings across search stages. When compared to the selections for relevant documents based on frequency of selection (not examining importance weights), as reported in Table 6.18 there are notable differences. For example, subjects selected "recency" in the exploration stage with a high enough frequency to be approximately 5% of the selections in that stage, but fewer than 2% indicated "recency" was the most important criteria to them in that stage. The frequency count for the selections in this table were fairly consistent and did not show any statistically significant shifts across search stages.

Table 6.20: Criteria Selection by Search Stage - Relevant Documents - Most Important Criteria *

Criteria	initiation	differentiating	exploration	extracting	verifying
accuracy	13.89%(1)	10.71%(1)	13.79%(1)	13.84%(1)	5.88%(2)
advertisement	0.00%	0.00%	0.00%	0.63%	0.00%
affectiveness	2.78%	0.00%	1.72%	0.00%	0.00%
authority	5.56%(2)	1.79%	3.45%	3.77%	5.88%
bias	2.78%	1.79%	0.00%	0.00%	0.00%
breadth	0.00%	7.14%(3)	3.45%	2.52%	7.84%(1)
definitions	2.78%	1.79%	1.72%	3.14%	5.88%(2)
depth	2.78%	7.14%(3)	3.45%	3.77%	7.84%
descriptions	5.56%(2)	3.57%	3.45%	2.52%	0.00%
guidelines	0.00%	0.00%	0.00%	1.89%	1.96%
history	5.56%(2)	5.36%	5.17%(3)	3.14%	1.96%
novelty	2.78%	8.93%(2)	5.17%(3)	8.81%(2)	1.96%
recency	5.56%(2)	8.93%(2)	1.72%	6.29%(3)	3.92%(3)
source	5.56%(2)	3.57%	1.72%	5.03%	9.80%
structure	2.78%	5.36%	0.00%	4.40%	0.00%
time	2.78%	1.79%	0.00%	2.52%	0.00%
tips	0.00%	0.00%	0.00%	1.89%	1.96%
understandability	5.56%(2)	3.57%	6.90%(2)	1.89%	5.88%(2)

6.3.3 Not Relevant Judgments

Subjects reported few documents as "not relevant." These judgments represent just 10% of the relevance judgments made by subjects (see Table 6.12). Table 6.21 lists the frequency counts for not relevant documents across search stages. When taken as a whole, an analysis of variance indicates a statistically significant association between criteria choices and search stages (χ^2= 81.8313, df = 64, p = 0.0658). Because of the low frequency counts in the individual cells, additional statistical analysis was not performed.

Table 6.21: Not Relevant Judgments Across Search Stage

Criteria	initiation	differentiating	exploration	extracting	verifying
accuracy	4	3	3	1	0
advertisement	5	2	5	8	0
affectiveness	3	3	1	7	1
authority	1	0	2	0	2
bias	3	1	0	1	0
breadth	3	4	9	1	0
depth	6	5	12	3	0
guidelines	1	2	5	0	0
history	0	3	8	1	1
novelty	0	1	2	2	0
recency	7	6	5	3	2
source	1	0	4	1	1
structure	0	1	3	2	1
time	1	0	2	0	0
tips	2	1	4	3	1
understandability	1	5	5	3	1
Total	38	37	70	36	10

6.3.4 Deliverable Due - Stage in Task Completion

As with Study 1, subjects were assigned a specific deliverable to be completed approximately every week. These deliverables represent "stages in task completion" and by evaluating the data reported by subjects within the time frame in which the

deliverables were due, the relevance judgment process can be examined in relation to these stages.

The results in Table 6.22 show that subjects reported document selection consistently over the duration of the project, with the exception of the time frame for the Final Deliverable where they performed approximately twice as many evaluations as for any previous deliverable. Table 6.23 shows the criteria selected by subjects in relation to stage in task completion. These selections do show some shifts in preference for criteria across search stage. These results appear to show shifts in preference for a number of criteria: "accuracy," "authority," "guidelines," "recency." When each criteria is tested individually across stage in task completion, a number of statistically significant associations are found as shown in Table 6.24.

Table 6.22: Documents Selected by Deliverable Due

Deliverable	Count	Percent
Abstract	326	21.79%
Detailed Outline	286	19.12%
Rough Draft	316	21.12%
Final Presentation	568	37.97%
Total	1496	100.00%

Table 6.23: Criteria Selection by Deliverable Due - Relevant Documents

Criteria	Abstract	Detailed Outline	Rough Draft	Final Presentation
accuracy	17	30	31	45
advertisement	4	1	0	5
affectiveness	16	9	25	33
authority	8	9	12	26
bias	5	4	4	14
breadth	19	13	18	24
definitions	16	10	8	22
depth	20	21	21	41
descriptions	25	15	13	36
guidelines	9	3	8	18
history	11	13	7	13
novelty	25	24	24	38
recency	21	18	17	46
source	14	14	13	24
structure	19	19	13	31

Criteria	Abstract	Detailed Outline	Rough Draft	Final Presentation
time	6	10	5	16
tips	13	0	8	14
topic	45	49	64	87
understandability	21	24	15	40
Total	314	286	306	573

Table 6.24: Analysis of Variance Results for Deliverables*

Criteria	x2
accuracy	12.77
advertisement(1)	6.8
affectiveness(1)	8
authority	15.18
bias(1)	10.48
definitions(1)	8.57
depth(2)	12.07
descriptions	15.05
guidelines	12.32
recency	22.31
structure(1)	8.34
time(1)	8.08
tips(2)	14.03
understandability(2)	13.68

* df =3, $p < .001$, except: (1) $p < .10$, (2) $< .05$

Table 6.25 shows criteria weights assigned in relation to stage in task completion. This shows some shifts in the level of importance for criteria in relation to stage in task completion. Based on these results, it appears that some significant shifts occur between the Rough Draft and Final Presentation deliverable. Additional analysis indicated a statistical association at various levels of significance between several criteria and stage in task completion as shown in Table 6.26. This table shows that for a specific level of importance, in moving from the Rough Draft to the Final Draft stage, subjects were more

likely to select the criteria identified at that level of importance during the Final Draft deliverable than during the Rough Draft deliverable.

Table 6.25: Criteria Weights by Deliverable Due

Weight	Abstract	Detailed Outline	Rough Draft	Final Presentation
not very important	13	4	17	28
slightly important	24	15	17	46
somewhat important	40	42	39	81
important	134	133	110	248
very important	103	92	123	170
Total	314	286	306	573

Table 6.26: Importance Value Changes - Rough Draft to Final Draft

Criteria	Importance	x2	p
accuracy	important	7.76	< .05
authority	important	6.25	< .05
definitions	very important	7.36	< .05
depth	important	3.13	< .05
descriptions	very important	8.07	< .05
novelty	important	2.67	=.10
recency	important	4.55	< .05
structure	somewhat important	5.33	< .05

6.3.5 Analysis of Grouped Relevance Criteria

Examination of results indicated that subjects generally selected more than one criteria as important to their relevance judgment process. As a data filter, to determine which groups of criteria subjects used, a number of correlation matrices were generated (see Appendix V). In reviewing these matrices, a value of .15 was used as a threshold to indicate some degree of correlation between criteria choices. Those criteria pairs which met this threshold are discussed in the sections below.

6.3.5.1 Grouped Relevance Criteria - All Search Stages

Table 6.27 shows the criteria pairs with a correlation coefficient above .15 which were generated using regression analysis of the results. It appears that across all search stages, subjects had some preference for certain criteria pairs. The criteria of "structure"

generated some of the highest coefficients in this group when paired with the criteria of "recency," "guidelines," "definitions," and "accuracy." The criteria of "recency" and "understandability" also generated a high correlation coefficient, as did the criteria of "guidelines" and "tips." The criteria of "accuracy" also generated a number of pairwise groupings that met the .15 threshold. This is a reflection of the fact that the criteria of "accuracy" was a consistent selection with the subjects throughout the search process (see Table 6.10).

Table 6.27: Correlation Coefficients >= .15 for Relevant Documents - All Search Stages

Criteria	Coefficient
structure - understandability	0.196
structure - tips	0.218
structure - time	0.153
recency - understandability	0.263
recency - time	0.133
recency - structure	0.270
guidelines - tips	0.360
guidelines - time	0.151
guidelines - structure	0.216
depth - understandability	0.173
depth - recency	0.152
depth - description	0.187
definitions - time	0.152
definitions - structure	0.245
definitions - source	0.160
definitions - novel	0.177
definitions - guidelines	0.195
definitions - descriptions	0.212
breadth - understandability	0.113
breadth - depth	0.198
affectiveness - guidelines	0.232
affectiveness - breadth	0.221
accuracy - structure	0.178
accuracy - source	0.165

Criteria	Coefficient
accuracy - recency	0.169
accuracy - novel	0.158
accuracy - guidelines	0.174
accuracy - description	0.200
accuracy - breadth	0.163
accuracy - affectiveness	0.229

Since the focus of this research is on relationships to search stage, grouped relevance criteria were examined within each search stage. These results are presented in the following sections.

6.3.5.2 Grouped Relevance Criteria - Initiation Stage

Table 6.28 identifies the criteria pair groupings from the initiating stage which met the .15 correlation coefficient threshold. Results from this stage demonstrate some preference for the criteria of "novelty" (see Table 6.18). The criteria of "recency," "novelty," "affectiveness," "descriptions," and "depth" all have a high correlation with the criterion of "definitions" in this stage. The criteria of "history" also has a high correlation with the criteria of "depth" and "time," indicating subjects demonstrated a tendency to select these criteria in groups.

Table 6.28: Correlation Coefficients >= .15 for Relevant Documents - Initiating Stage

Criteria	Coefficient
definitions - understandability	0.330
definitions - time	0.329
recency - understandability	0.328
recency - structure	0.241
recency - definitions	0.457
novelty - understandability	0.253
novelty - tips	0.163
novelty - source	0.145
novelty - definitions	0.483
novelty - recency	0.173

Criteria	Coefficient
history - time	0.447
guidelines - understandability	0.150
guidelines - definitions	0.446
guidelines - recency	0.283
guidelines - history	0.421
descriptions - understandability	0.189
descriptions - structure	0.284
descriptions - definitions	0.562
descriptions - novelty	0.459
depth - understandability	0.189
depth - time	0.219
depth - definitions	0.562
depth - novelty	0.163
depth - history	0.447
depth - descriptions	0.414
breadth - structure	0.165
breadth - recency	0.190
breadth - guidelines	0.196
breadth - description	0.247
authority - understandability	0.495
authority - structure	0.165
authority - source	0.334
authority - definitions	0.253
authority - history	0.185
authority - guidelines	0.253
authority - depth	0.319
authority - breadth	0.150
affectiveness - definitions	0.421
affectiveness - recency	0.239
affectiveness - guidelines	0.421
advertisement - understandability	0.352
advertisement - time	0.227
advertisement - definitions	0.307
advertisement - recency	0.314
advertisement - guidelines	0.307
advertisement - affectiveness	0.153
accuracy - source	0.209
accuracy - definitions	0.262

Criteria	Coefficient
accuracy - novelty	0.353
accuracy - description	0.402
accuracy - breadth	0.158
accuracy - affectiveness	0.331

6.3.5.3 Grouped Relevance Criteria - Exploration Stage

Table 6.29 lists the criteria pairs that met the threshold for correlation in the exploration stage. A number of criteria grouped with the criteria of "accuracy" in this stage, again demonstrating the popularity of this criteria with this sample. The criteria of "guidelines" and "tips" also generated a strong correlation, as did the criteria of "descriptions" and "source."

Table 6.29: Correlation Coefficients >= .15 for Relevant Documents - Exploration Stage

Criteria	Coefficient
time - understandability	0.266
time - tips	0.329
structure - tips	0.250
source - time	0.233
recency - understandability	0.281
recency - structure	0.181
history - understandability	0.269
history - recency	0.181
guidelines - tips	0.409
guidelines - structure	0.274
description - understandability	0.207
description - time	0.161
description - source	0.514
description - recency	0.207
description - guidelines	0.169
depth - understandability	0.379
depth - recency	0.306
depth - novelty	0.160
depth - history	0.198
depth descriptions	0.306

Criteria	Coefficient
definitions - understandability	0.183
definitions - source	0.216
affectiveness - guidelines	0.266
accuracy - understandability	0.256
accuracy - tips	0.291
accuracy - time	0.162
accuracy - structure	0.185
accuracy - source	0.345
accuracy - novelty	0.320
accuracy - history	0.264
accuracy - guidelines	0.309
accuracy - description	0.221
accuracy - definitions	0.255
accuracy - breadth	0.157
accuracy - affectiveness	0.153

6.3.5.4 Grouped Relevance Criteria - Extracting Stage

Table 6.30 identifies the criteria pairs that met the correlation threshold of .15 for the "extracting" stage. Because of the high number of criteria evaluations in this stage, and the variety of criteria selections, a variety of pairwise correlations were found with a lower coefficient than other stages. The criteria of "guidelines" and "tips" generated a high correlation, as did "affectiveness" and "guidelines."

Table 6.30: Correlation Coefficients >= .15 for Relevant Documents - Extracting Stage

Criteria	Coefficient
structure - understandability	0.237
structure - tips	0.266
recency - tips	0.266
recency - structure	0.220
guidelines - tips	0.343
guidelines - time	0.228
guidelines - structure	0.168
descriptions - structure	0.222

Criteria	Coefficient
descriptions - novelty	0.242
descriptions - guidelines	0.153
depth - source	0.275
depth - understandability	0.166
depth - descriptions	0.154
definitions - time	0.220
definitions - novelty	0.188
definitions - guidelines	0.195
breadth - understandability	0.228
breadth - source	0.172
breadth - guidelines	0.220
breadth - description	0.255
breadth - depth	0.261
authority - source	0.168
authority - guidelines	0.150
affectiveness - time	0.149
affectiveness - guidelines	0.304
affectiveness - description	0.188
affectiveness - depth	0.192
affectiveness - breadth	0.231
accuracy - breadth	0.213
accuracy - affectiveness	0.239

6.3.5.5 Grouped Relevance Criteria - Verifying Stage

Table 6.31 identifies the criteria pairs that met the .15 threshold for the "verifying" stage. Strong correlations were found for several of the paired criteria for "tips" and "guidelines," an indication of a strong preference to select these criteria in pairs even though individually these were not commonly selected criteria during this stage (see Table 6.16 and 6.18). Strong correlations were also found for "source" quality, and "guidelines."

Table 6.31: Correlation Coefficients >= .15 for Relevant Documents - Verifying Stage

Criteria	Coefficient
definitions - accuracy	0.277
definitions - breadth	0.288
depth - accuracy	0.200
depth - affectiveness	0.232
descriptions - accuracy	0.281
descriptions - affectiveness	0.179
descriptions - definitions	0.489
descriptions - depth	0.152
guidelines - accuracy	0.493
guidelines - definitions	0.456
guidelines - descriptions	0.391
history - accuracy	0.226
history - guidelines	0.197
novelty - accuracy	0.579
novelty - advertisements	0.268
novelty - definitions	0.308
novelty - descriptions	0.295
novelty - guidelines	0.328
recency - accuracy	0.441
recency - descriptions	0.257
recency - guidelines	0.368
recency - novelty	0.475
source - accuracy	0.338
source - definitions	0.277
source - guidelines	0.493
source - novelty	0.422
source - recency	0.441
structure - accuracy	0.427
structure - definitions	0.246
structure - descriptions	0.453
structure - guidelines	0.562
structure - novelty	0.444
structure - recency	0.411
structure - source	0.337
time - advertisement	0.158
time - novelty	0.333

Criteria	Coefficient
time - structure	0.166
tips - accuracy	0.323
tips - definitions	0.406
tips - descriptions	0.456
tips - guidelines	0.759
tips - novelty	0.277
tips - recency	0.207
tips - source	0.323
tips - structure	0.504
understandability - accuracy	0.182
understandability - definitions	0.220
understandability - guidelines	0.220
understandability - history	0192
understandability - descriptions	0.211

6.3.5.6 Grouped Relevance Criteria in Relation to Search Stage

The analysis discussed previously established various pairwise groupings for relevant documents during certain search stages. Additional analysis was performed to determine whether these criteria groups showed a statistically significant increase (at $p < .10$) in importance across early (initiation/differentiation + exploration) and late (extracting) search stages. This combination of search stages provides a basis of a roughly equivalent number of document selections for the early search stages versus late search stages, and thus differences in frequency counts between the selection of criteria are more likely to be due to criteria preference than preference for a search stage. The results of this analysis are shown in Table 6.32.

Table 6.32: Grouped Relevance Criteria - Early versus Late Search Stage*

Criteria	x2
novelty - source	5.4
novelty - structure	2.79
description - structure	6.429
recency - structure	3.6
structure - guidelines	4
accuracy - guidelines	6.25

* $p < .10$

6.4 Discussion

Study 3 sought to examine the relationship of relevance criteria selection across search stage, stage in task completion, and with groups of relevance criteria. The methods from Study 2 were used in this study, with minor changes. The findings reported here provide further confirmation of the relationships detected in Study 1 and Study 2.

6.4.1 Major Findings

The major findings for this study are as follows.

1. Subjects showed a preference for the criteria of "accuracy," "depth," "novelty," and "recency" across all search stages (research question 1).
2. With partial relevance judgments, subjects showed a preference for the criteria of "accuracy," "affectiveness," "authority," "descriptions," "guidelines," "novelty," "recency," "source," "structure," "time," and "understandability" and analysis determined that a number of these criteria had a statistical correlation with search stage selection as shown in Table 6.15 (research question 1, research question 3).
3. With relevant document judgments, subjects showed a preference for "accuracy," "affectiveness," "authority," "bias," "breadth," "depth," "definitions," "descriptions," "guidelines," "history," "novelty," "recency," "source," "structure,"

"time," and "understandability" and analysis determined these criteria selections had a statistical correlation with search stage selection as shown in Table 6.17 (research question 3).

4. Subjects demonstrated a preference for "accuracy," "advertisement," "affectiveness," "authority," "bias," "definitions," "depth," "descriptions," "guidelines," "recency," "structure," "time," "tips," "understandability" and the selection of a number of these criteria demonstrated a statistical relationship across all search stages as shown in Table 6.19 (research question 3).
5. For relevant documents across stage in task completion, the criteria of "accuracy," "authority," "descriptions," "guidelines," and "recency" demonstrate a statistically significant relationship with stage in task completion as shown in Table 6.24 (research question 2).
6. Subjects demonstrated a preference for the criteria groups of "novelty-source," "novelty-structure," "description-structure," "recency-structure," "structure-guidelines," and "accuracy-guidelines" and the selection of these groups and analysis determined there was a statistical relationship at various levels of significance across *early* versus *late* search stages as shown in Table 6.32 (research question 4).

6.4.2 Detailed Discussion

A number of relationships were detected based on the analysis performed in this study and provide empirical evidence for all research questions. The following sections detail the relationships detected based on relevance criteria selections by subjects. Analysis is divided into sections based on the type of relevance judgment.

6.4.2.1 Partially Relevant Documents

Partially relevant documents represented 22 % of the selections made by subjects. Unlike Study 2, partially relevant document selection was not concentrated in the "initiation" stage. In this study, partial relevance judgments were made consistently throughout the search process.

A variety of relevance criteria were important to subjects during this stage in terms of frequency of selection: "recency," "ability to understand," and "novelty." The criteria of "source" quality was selected infrequently in early search stages, but subjects indicated a preference for this criteria in the extracting stage. Subjects appeared to be more discerning in their evaluation of partially relevant documents, selecting more criteria with a higher frequency in later search stages.

The large number of criteria in Table 6.15 represent a statistically significant finding that subjects are selecting partially relevant documents with a more varied set of criteria in later search stages. Results also indicated that for some criteria, selection decreases, but then subjects show an increased interest in these criteria during the "verifying" stage. Based on selection counts, in this sample, the criteria of "accuracy," "affectiveness," and "ability to understand" decrease in importance after the "initiation" stage, but then increase in importance in the "verifying" stage. This finding suggests a relationship between criteria selected for relevance judgment and the cognitive actions required for "verifying."

6.4.2.2 Relevant Documents

The majority of relevance judgments made by subjects in this study was for relevant documents. The criteria of "accuracy" "depth" "novelty" and "recency" were all consistently selected by subjects in judging documents relevant, indicating that subjects

in this study used a number of criteria to judge document relevance, and the "accuracy" and "depth" [of coverage] for the document were consistently important criteria used in making the relevance judgment. The criteria of "recency" and "novelty" are also important, but not just in later stages as some previous studies indicated. The criteria of "novelty" increases and then decreases in frequency of selection as the search process advances. This finding differs from previous studies which suggest that searchers look for new (novel) documents in later search stages (Vakkari, 2000). The selection of the criteria of "recency" first decreases in selection, then increases in the "verifying" search stage. This finding implies that subjects are searching for recent (current) documents as they prepare to conclude their research, and may suggest some effort on the part of the subject to "fill in" missing portions of their research with more current information. Given the technical nature of the research project assigned to the subjects (computer technology) , more current information is generally more important than less current information.

A number of criteria demonstrated a statistically significant increase in importance for relevant judgments as subjects progressed through the search process (see tables 6.17 and 6.19). This finding indicates that subjects are more discerning in later search stages, using a variety of criteria to make their relevance judgment. The identification of the various criteria which are more important to subjects in later search stages adds clarity to this finding.

6.4.2.3 Non-Relevant Documents

The subjects in this study reported not relevant documents as a small percentage of total documents evaluated (10%). The selection of the criteria used to evaluate these documents did demonstrate a statistically significant shift across all criteria across all search stages. This finding is significant (at the $p = .05$ level) and indicates that subjects

evaluate documents differently as they progress through the search process, and that those revised document evaluations also pertain to not relevant judgments.

6.4.2.4 Deliverable Due - Stage in Task Completion Analysis

Subjects in this sample evaluated documents consistently across the first three stages of task completion, indicating a smooth progression through the search process. They evaluated almost 40% of all documents in the preparation of the Final Presentation, suggesting that some subjects may have done some significant portion of the work in the final stage of task completion. Additionally, subjects appear to be repeating the information search process within the stage in task completion (see Table 6.2). This finding suggests that in practice, searchers may make multiple iterations through an ISP within stages in task completion.

Subjects demonstrated a preference for a number of criteria within stages in task completion, and they reported a statistically significant increase in preference for those criteria as they progressed through the search process. These statistically significant findings indicate that subjects preference for the criteria identified in Table 6.24 and 6.26 increased over the course of the work task. Many of these criteria were reported as increasing in importance in the search stage analysis, but this analysis also reveals that the criteria of "bias" as an important criteria for subjects as they work to complete their assigned task. This indicates that in this sample, subjects were more interested in the position of the author (the author's bias) in later stages of task completion than in early stages.

6.4.2.5 Grouped Relevance Criteria

Subjects in this sample demonstrated a preference for a number of groupings. The statistically significant findings in Table 6.32 indicate that subjects demonstrated a

change in preference for a number of these criteria groupings in relation to the progress through the search process. These groupings indicate that subjects are interested in the quality and structure of the source, and the novelty of the source. Subjects also have an interest in the recency of the document (the document is current), and that the document contains some guidelines. Given the technical nature of the report assigned to subjects, the use of guidelines appeared to be more important to subjects as they moved through the search process. So as subjects move into later stages of the search process, they appeared more likely to select these criteria in groups, an indication that they are looking for documents which fulfill not just one criteria, but groups of criteria.

Chapter 7 - Cross Study Analysis

The studies examined relevance criteria over the progression of time as a subject gathered information for a research project. Time was measured either as search stage (Study 1,2, and 3), or as stage in task completion (Study 2 and Study 3). The data collected and analyzed over the course of three studies involved 174 subjects who made 1,798 distinct document judgments and recorded 3,011 criteria selection records. Results analyzed from all studies demonstrated a number of statistically significant associations with relevance criteria selections and search stage, with relevance criteria selections and stage in task completion, and with groups of relevance criteria and information search process stage.

Study 1 examined the relevance criteria choices in relation to search stage in a controlled laboratory environment. The search session (episode) was short-term (1 to 2 hours). The results of this study provided data for the analysis of research questions 1 and 3.

Study 2 expanded and refined the methods of Study 1 and examined relevance criteria choices across both search stage and stage in task completion. The study was longitudinal, extending over a 5 week period. The 82 subjects were allowed to work in a naturalistic environment at their own pace, performing searches and submitting data over the Web. Results from this study provided data for research questions 1,2,3, and 4.

Study 3 further refined the methods of Study 2 and provided additional data for analysis. This study expanded on the methods of Study 2 by allowing the 53 subjects to indicate indicate the importance of a relevance criterion selection. Study 3 also allowed the subject to indicate which criteria was most important to the subject, and which criteria

was least important. Data collected was used for analysis of research questions 1,2,3, and 4.

7.1 Cross Study Analysis - Research Question 1

Research question 1 asked whether or not users have a preference for specific criteria choices in relation to certain relevance judgments. In order to examine statistical results across studies, the methodological differences between the studies must be understood. The methods for Study 2 and Study 3 involved three levels of relevance judgments: *relevant*, *not relevant*, and *partially relevant*. In Study 1, subjects indicated relevance on a scale of 1 to 10. Analysis of results from Study 1 involved the collection of data for distinct bands of relevance which were identified as *low relevance* (1 to 3), *partial relevance* (4 to 7), and *high relevance (8 to 10)*. For the purposes of this cross-study analysis, the results for high relevance are interpreted as "relevant" judgments which can be compared to judgments of relevant documents in Study 2 and Study 3.

The results in Table 7.1 compare the percentage of selection for criteria common to all three studies. These results provide some indication that subjects demonstrated a consistent preference for some criteria across different studies using different methods. The criteria of "recency," "depth," and "ability to understand" were consistently selected by subjects in each of the studies with a high level of frequency across all search stages. These results indicate that subjects in this study were interested in understandable, current (recent) documents, with sufficient depth of coverage. This finding may be an indication that these criteria represent basic prerequisites for document selection.

The criteria of "bias" may have had more importance to subjects in Study 1 where the assigned question was to "compare and contrast" two technical concepts. The criteria of "amount of information" was eliminated from Study 3 where subjects were expected to

choose either the criteria of "depth" (deep coverage of a topic), or "breadth" (broad coverage of the topic and various related topics); what had previously been selected as "amount of information" was therefore distributed over the criteria of "depth" and "breadth," a change which is not reflected in this table.

The criteria of "accuracy" and "structure" were important to subjects in Study 2 and Study 3 where research was performed on the Web. This finding suggests that subjects were concerned with the structure of documents on the Web, where documents are interactive pages which can be difficult to navigate, or documents may contain links which are not useful to the subject. The selection of the criteria of "structure" may also reference a number of deficiencies of information retrieval using hypertext Web navigation. Subjects may also have had some awareness the information retrieved from the Web comes from a variety of sources with varying degrees of quality, as reflected in the consistent selection of the criteria of "accuracy" for Study 2 and Study 3.

Table 7.1: Cross-Study Comparison of Frequency of Selection for Common Relevance Criteria*

Criteria	Study 1	Study 2	Study 3
recency	8.00%	8.00%	7.00%
authority	8.00%	3.00%	3.00%
bias	7.00%	2.00%	4.00%
depth/scope	10.00%	9.00%	7.00%
ability to understand	10.00%	14.00%	8.00%
amount of information	9.00%	9.50%	n/a
accuracy	n/a	11.00%	9.00%
structure	n/a	7.25%	6.00%

* percent of criteria selected across all search stages

7.2 Cross Study Analysis - Research Question 2

Research question 2 concerned whether or not there was a relationship between criteria selection and *stage in task completion*. Study 1 did not examine stage in task

completion. For Study 2 and Study 3, stage in task completion was considered to be the interim deliverables subjects had to submit on a consistent basis throughout the duration of their assigned research project. These deliverables were due at a specific point in time. By examining the relevance judgment information submitted during the time period when subjects were required to submit the deliverable, an association can be made between the criteria used for relevance judgments and stage in task completion (deliverable due).

Table 7.2 identifies the criteria selections which demonstrated a statistically significant association with stage in task completion in Study 2 and Study 3. The criteria of "recency," "structure," and "ability to understand" were common to both studies, strengthening the statistically significant finding that these criteria are important to subjects, and that the level of importance for the criteria increases as the subject progresses through stages in task completion. The criteria of "depth" in Study 3 can be considered equivalent to the criteria of "amount of information" in Study 2, and when considered as such strengthens the statistically significant finding that "depth" (amount of information) increases in importance to subjects as they progress through stages in task completion. The criteria of "authority" was important to subjects, a finding which suggests that subjects were aware that not all documents found on the Web are reliable, and research performed on the Web should involve an evaluation of whether or not the document being retrieved is authoritative. The more specific criteria of "tips," "guidelines," and "descriptions" was not provided in Study 2. The statistically significant finding that these criteria are important to subjects suggests that availability of more specific criteria for searching may be useful to individuals seeking to acquire information.

Table 7.2: Cross-Study Comparison of Statistically Significant Criteria Choice Associations to Stage in Task Completion

Criteria	Study 2	Study 3
affectiveness	x	
accuracy		x
advertisement	n/a	x
amount of information	x	
authority	x	x
breadth	x	
bias		x
definitions	n/a	
depth		x
descriptions	n/a	x
guidelines	n/a	x
tips	n/a	x
novelty	x	
recency	x	x
structure	x	x
ability to understand	x	x

7.3 Cross Study Analysis - Research Question 3

Research question 3 involved examining whether or not there is a relationship between criteria selections and stage in search process. Table 7.3 contains a cross study comparison of the statistically significant findings of associations between criteria selections and stage in search process. This adds strength to the findings in the individual studies that criteria choices change due to cognitive state changes across search stages. A statistically significant association for the criteria of "ability to understand" was found across all three studies, further strengthening the finding of association. This may be another indication that the "ability to understand" along with the 'epistemic value' of a document (Wang and Soergel, 1998) is a basic prerequisite for positive relevance judgments.

The statistically significant finding that the criteria of "accuracy," "novelty," "source," "time," and "structure" become more important to subjects as they progress through the search process may be an indication that as subjects become more discerning in making relevance judgments later in the search process, their evaluation focuses on these criteria. This may once again be a reflection of issues involved with using the unorganized, unevaluated, raw information available on the Web to perform research. As subjects progressed through the search process, they became more discerning in their evaluations, and they sought new documents with a navigable document structure from accurate sources. Also, later in the search process, time is limited and the time constraints of the research project assignment loom, as reflected in the selection of the "time" criterion.

Table 7.3: Cross-study Comparison of Statistically Significant Associations between Criteria Selections and Search Stage Progress

Criteria	Study 1	Study 2	Study 3
clarity	x	n/a	n/a
precision	x	n/a	n/a
depth/scope	x	x	
bias		x	
descriptions	n/a	n/a	x
guidelines	n/a	n/a	x
authority	x		x
instructional	x	n/a	n/a
ability to understand	x	x	x
accuracy	n/a	x	x
affectiveness	n/a		x
novelty	n/a	x	x
recency	x		x
source quality	n/a	x	x
structure	n/a	x	x
time	n/a	x	x

7.4 Cross Study Analysis - Research Question 4

Research question 4 involved the relationship of sets or groups of criteria to search stage progress. Study 1 did not examine criteria groups. Analysis of the results of Study 2 found a number of groups of criteria which exhibited a relationship with search stage progress, as shown in Table 5.32. Study 3 also found a number of statistically significant relationships as shown in Table 6.32. These statistically significant findings demonstrate the relationship between groups of criteria selections by subjects and the information search process, but find no common statistically significant groups between the two studies. This may be due to the smaller number of subjects in the second study (53 for Study 3 versus 82 for Study 2), and the changes in the relevance criteria selections available to subjects for Study 3.

Chapter 8 - Discussion and Conclusion

These findings strengthen the cognitive concept of relevance as complex and dynamic, and provide indications that changes in the user's cognitive state are partially manifested as changes in criteria selections over the course of the search process as described in chapter 3. These changes in the user's cognitive state, are reflected in changes in the use of criteria during the relevance judgment process, which in turn affect their relevance assessments. These criteria provide the focus with which the searcher determines relevance, and this focus changes as the user's cognitive state changes.

As detailed in the chapters on the individual studies, analysis of results from each study separately demonstrated associations between relevance criteria selections and search stage, and for Study 2 and Study 3, for relevance criteria selections and stage in task completion. Study 2 and Study 3 results also demonstrated statistically significant associations for groups of criteria and search stage. The previous chapter further strengthened these findings through a cross-study analysis, demonstrating consistent user behavior across the three studies.

8.1 Summary of Major Findings

The exploratory research conducted in these studies sought to examine the relationship of relevance criteria selections to search stage progress. Specifically, the research questions for these studies are as follows.

1. Does the user's choice of some relevance criterion change in relation to relevance judgments?
2. Does the importance of some *relevance criterion* change in relation to *stage in task completion* as indicated by the frequency of criterion selection, and/or a weight indicating importance as assigned by the user?

3. Does the importance of some *relevance criterion* change in relation to a user-identified *stage in the search process* as indicated by the frequency of criterion selection, and/or a weight indicating importance as assigned by the user?
4. Are there sets of *relevance criteria* choices which change in importance in relation to a *user-identified stage* in the search process as indicated by the frequency of criterion selection, and/or a weight indicating importance as assigned by the user?

A summary of the major findings of these studies is as follows.

1. **Research Question 1:** Subjects demonstrated a preference for specific criteria across all three studies. Though there were variations in the selection, the criteria of "recency," "authority," "bias," "depth/scope," "ability to understand," "amount of information," "accuracy," and "structure" were consistently selected across studies (see Table 7.1).
2. **Research Question 2:** Study 2 and 3 examined relevance criteria choices in relation to stage in task completion. The criteria of "authority," "structure," "recency," and "ability to understand" were identified in both Study 2 and Study 3 and demonstrated a statistically significant relationship between criteria selection and stage in task completion. There were other criteria which demonstrated a relationship between stage in task completion and relevance criteria as shown in Table 7.2.
3. **Research Question 3:** The data collected in Study 1,2, and 3 all demonstrated a relationship between criteria selection and search stage. The criteria of "ability to understand" demonstrated a statistically significant association in all three studies. Other criteria were common to two of the studies as shown in Table 7.3.

4. **Research Question 4:** Study 2 and 3 examined the selection of groups of criteria in relation to search stage. A number of criteria groups were found and demonstrated a statistically significant association with stage in task completion as shown in Table 5.32 and Table 6.32.

8.2 Discussion

These statistically significant findings indicate that as users progressed through the search process their relevance judgments became more discerning, results consistent with previous studies (Tang and Solomon, 1998; Spink et al, 1998). Consistent findings across the three studies presented here strengthen the findings of the individual studies that demonstrated a statistically significant association between use of criteria to make relevance judgments and their choice of stage in the search process, and separately their choice of stage in task completion. These results are consistent with previous non-statistical studies which reported some evidence of changes in criteria choices across various search stage models (Vakkari, 2001; Vakkari & Hakala, 2001; Tang & Solomon, 2001; Wang & White, 1999; Hirsh, 1999). The results from these three studies add statistical strength to those findings, and provide additional findings through the use of a more detailed search stage model and findings of user preference for specific criteria across the stages of this search model. Additional findings of user preference for criteria across stage in task completion extend the findings of these previous studies and provide some confirmation of the model proposed in chapter 3.

Previous studies indicated that subjects used groups of criteria to make relevance judgments (Wang and White, 1999), but did not identify those groupings, nor has any previous research attempted to associate groups of criteria with search stage progress. The findings in Study 2 and Study 3 provide indications of the use of several groups of

criteria by subjects, and found statistically significant associations between the selections of those groups of criteria and search stage.

8.3 Limitations of Studies

The research methods employed allowed subjects to select topicality as a criteria, but did not specifically analyze topicality in relation to the selection of other criteria. This approach was based on the assumption that the document first needed to be 'on topic' before other criteria would be considered (Wang and Sorgel, 1998; Crystal and Greenberg, 2006). This research focused on the use of more specific relevance criteria.

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 was drawn from a pool of undergraduate students who are business majors in a business school at an American university. These students are part of the Millennial generation, the generation born between 1980-1995, and there is research to suggest their search behavior may be different from that of other demographic groups. Research has indicated that students in this age group may be more motivated by convenience and inclined to accept any source as valid and may be disinclined to pursue other sources of information to verify what they have found (Harley, Dreger, Knobloch, 2001; Weiler, 2004; Vondracek, 2007).

All students were taking the same course and were given the same assignment, thus all had the same *work task*. 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.

Study 3 had a large percentage of students for whom English was a secondary language and may have impacted the results from that sample. Study 2 had a very small percentage of students who reported English as a second language, and the Study 1 sample was not surveyed for their primary language.

Though subjects were 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 have also limited the amount of research the subjects could perform.

Task is also known to influence relevance judgments and in this research was treated as a constant since all subject's were 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 search stage selection data that has been collected and examined in this study indicates a reasonable distribution of selections by subjects and thus provides some indication that the model was appropriate and provides a reasonable foundation for research question 3 and 4. The examination of relevance criteria selections by stage in task completion in study 2 and study 3 do not depend on the search stage model and demonstrated a number of statistically significant correlations which add strength to the general finding that criteria selections are related to progress over a search process (as

measured by time). Due to the time constraints of this research, a detailed analysis of the search stage data selected by subjects has not been done and is to be investigated in future studies.

The Yahoo! search engine was used to generate the search results for the modified search engine used for Study 2 and 3. The reason for this choice was the technical ease in working with Yahoo! search results output. It is a limitation these studies 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.

8.4 Suggestions for Future Research

These findings provide a foundation for the detailed analysis of the use of criteria during the information search process. While there was some attempt to identify a combination of general and more specific criteria in Study 3, future studies should continue to add more detailed criteria specific to a subject area.

This research attempted to isolate task influences on the relevance judgment process, and focused on a specific research task. Subjects were assigned a search task indirectly as part of a research project assignment. The topic was not necessarily a topic of interest for the subject. Future research should examine other types of work tasks to determine the influence of work task on criteria selection.

The subjects for this study were undergraduate students between 19 and 25 years of age. Subjects in this age group represent a relatively small portion of the total population. Future studies should therefor examine the behaviors of subjects from different age groups. Though the sample size used was adequate for the statistical analysis performed, a larger sample size selected from a more diverse sample population would improve the significance and generalizability of the results.

This research provided some isolation of the influence of domain knowledge on the selection of criteria. It was assumed that all subjects had a lack of knowledge in the subject area. Subjects were undergraduate students with a variety of research skills. Future research should attempt to distinguish between research skills of subjects, and the domain knowledge of subjects. This would allow an association between the user's knowledge and background, and selection of criteria for relevance judgment.

The results of these studies provide direct suggestions for improvements in information retrieval (IR) systems. By acknowledging the role of the searcher's use of criteria beyond topicality, search engines should employ more robust use of additional criteria in the search process. Though some search engines currently provide "advanced" features which allow additional criteria to be added to the query (for example, dates, author), the use of these advanced features is limited and often counter-intuitive. These IR systems also do not provide a recognition of how the importance of these criteria change over the course of the search process. An IR system that recognizes a "search episode" session, and search progress over the course of that session, or potentially multiple sessions, is needed. An adaptive IR system should also employ some mechanism to alter criteria weights as the user progresses through the search process. Some of the findings presented here provide suggestions for what those weights could be. For example, "ability to understand" would be a criteria weighted more heavily in early search stages, and "accuracy" and "depth" would be weighted more heavily in later search stages. Such as system could potentially improve the search process by retrieving documents which are more likely to judged relevant by the searcher.

Chapter 9 - Conclusion

The criteria used to judge relevance effectively provide a lens into the cognitive changes occurring during the information search process. Though prior research has identified criteria used to judge relevance and has provided some hints as to how these criteria are selected and when, our knowledge of the selection of these criteria and their relationship to the information search process is incomplete.The purpose of these studies was to increase our understanding of these criteria choices by examining how they change over the course of an information search process.

The studies detailed here examined the relevance criteria selections of subjects as they progressed through an information search process. In three distinct studies, subjects conducted a search to either answer an assigned question or to prepare an assigned research report. During their search, subjects examined documents and recorded information on their stage in the information search process, their relevance judgment, the criteria used to make that judgment, and in studies 2 and 3, the importance of the criteria used in making that relevance judgment. This data was then analyzed using descriptive statistics, regression analysis and analysis of variance. Methods for data collection varied across each study as detailed in chapter 3.

9.1 Major Findings

As identified in chapter 8, analysis of the results of these studies resulted in a number of major findings. Analysis found that subjects used a consistent set of relevance criteria across the three studies, and these choices varied in relation to progress through the information search process model described in chapter 3. These similarities across distinct studies strengthen the findings of the relationship between criteria selection and information search process. The repeatability of these results for relevance criteria

choices and stage in task completion across the three studies further strengthens these findings.

In each study, the relationship between relevance criteria judgment and search stage was examined. The most commonly selected criteria were identified across the three studies. Findings indicated that users are consistently interested in the same criteria in relation to a relevance judgment and provide some indication that more criteria are found for relevant documents than for non-relevant documents.

The methods for Study 2 and Study 3 had subjects prepare a research project. The research project had a number of interim deliverables which represent stages in task completion. Analysis of the results examined the relevance criteria choices of subjects in relation to these stages in task completion. Findings indicated that users are interested in different criteria later in the search process. Specifically "authority," "recency," and "ability to understand" are more important to subjects as they progress through stages in task completion. These findings suggest that subjects seek more authoritative, recent sources that can be understood, and they seek these with increased importance in later stages of task completion.

The relationship of relevance criteria choices and stage in the information search process was examined in all studies. The information search process model used was the synthesis of the information search process models described in chapter 3. Across all studies, findings indicated that the criteria of "ability to understand" demonstrated a consistent increase in importance as subjects progressed through the information search process. This finding suggests that understanding of content is more important as subjects near completion of their search process and they have a stronger need to understand the material being reviewed. Across Study 2 and Study 3, the criteria of "accuracy,"

"novelty," "source quality," "structure," and "time" also increase in importance in later stages of the search process. These findings suggest that users seek more accurate, quality sources which are new and have a useful structure in later stages of the search process.

Study 2 and Study 3 examined sets or groups of relevance criteria and found several consistent sets of selections across both studies. Analysis determined that the selection of several of these groups was impacted by progress through the information search process. These findings suggest that users do not commonly use one criteria to make a relevance judgment, but instead use groups of criteria, and the selection of some of these criteria groups increase in importance in later stages of the information search process.

9.2 Significance of Findings

These statistically significant findings add to previous studies which examined the use of criteria for relevance judgments. These studies report a number of interactions of specific criteria for both search stage progress and stage in task completion, criteria which have not been previously identified. These findings are strengthened by replication of results across the three studies. New findings also include interactions with sets of criteria and search stage progress, a line of research which has not been previously examined.

These studies used both a search stage model, and a stage in task completion model which related to interim deliverables required of the subjects. In comparing the subject's progress through the information search process to the subject's stage in task completion some surprising results were found. Subjects appeared to be iterating through the information search process for each deliverable. Though these results were not completely unexpected, the degree to which subjects appeared to be backtracking and

repeating search stages was unexpected. Further exploration of these results is beyond the scope of this research, but future studies will seek to examine the relationship between these models.

These findings indicated that subjects seek more authoritative, recent sources that can be understood, and they seek these with increased importance in later stages of task completion. These findings also indicate that users seek more accurate, quality sources which are new and have a useful structure in later stages of the search process. Subjects searching are also inclined to use multiple criteria to judge document relevance, and these criteria are commonly selected in identifiable groups which also vary by stage in the information search process.

These results provide direct suggestions for IR system design. An adaptive IR system could capitalize on these findings. A system which recognizes criteria beyond topicality and relates these criteria to progress through an information search process would be more aligned with the cognitive changes of the user than current static IR systems. Such as system could potentially be more effective at selecting documents which fulfilled the criteria which were predicted to be more important to subjects at that point in their information search process.

9.3 Closing Remarks

Relevance is a foundational concept for information science and a concept that is key to IR system evaluation. A deeper understanding of relevance, and specifically the dynamic nature of relevance, is imperative to increasing the satisfaction of IR system users. Without a more complete understanding of how IR users evaluate documents and the criteria used to evaluate the document, improving IR document selection may remain an elusive goal.

These studies provide important results which enhance our knowledge of the rich, dynamic relevance judgment process. Across three studies amidst the highly variable selection of documents with distinct topics, consistency could be found in the selection of a number of relevance criteria, and groups of criteria in relation to information search process progress. These studies help to clarify our understanding of relevance and provide a richer picture of the complex, dynamic process of relevance judgments. They provide direction for future research to further examine criteria selection in the dynamic relevance judgment process and ultimately exploit these lessons learned in improved IR system design and increased user satisfaction.

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.

Bruce, H. (1994). A cognitive view of the situational dynamism of user-centered relevance estimation. *Journal of the American Society for Information Science, 45* (3), 142-148.

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

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.

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

Harley, B., Dreger, M., Knobloch, P.. 2001. The postmodern condition: Students, the Web, and academic library services . *Reference Services Review, 29* (1): 23-32.

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. (1996). Cognitive perspectives of information retrieval interaction: Elements of cognitive IR theory. *Journal of Documentation, 52* (1), 3-50.

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.

Kuhlthau, C. (2004). Seeking Meaning: A process approach to library and information services, 2nd Ed.. Libraries Unlimited**:** Norwood, NJ.

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.

Lin, S., & Belkin, N. (2005). Validation of a model of information seeking over multiple search sessions. *Journal of the American Society for Information Science and Technology, 56* (4), 393-415.

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. (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.

Saracevic, T. (2007a), Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology, 58* (13), 1915-1933.

Saracevic, T. (2007b), Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology, 58* (13), 2126-2144.

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.

Spink, A., Wilson, T., Ford, N., Foster, A., & Ellis, D. (2002). Information seeking and mediated searching study. Part 3. Successive Searching. *Journal of the American Society for Information Science and Technology, 53* (9), 716-727.

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

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.

Vickery, B.C.. (1959). Subject analysis for information retrieval. In (ed.), *Proceedings of the International Conference on Scientific Information*, National Science Foundation.

Vondracek, R.. 2007. Comfort and convenience? Why students choose alternatives to the library. *Libraries and the Academy, 7* (3): 277-293.

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.

Weiler, A.. 2004. Information-seeking behavior in generation Y students: Motivation, critical thinking, and learning theory . *Journal of Academic Librarianship, 31* (1): 46-53.

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: Research Project Assignment for Subjects - Study 2 and Study 3

General Description

This research project requires you to prepare a presentation for an approved Information Systems topic. As you prepare this presentation, a series of interim *deliverables* will be required. These deliverables include a project description which explains what you will write about the topic, an outline for the presentation, a rough draft of the presentation (PowerPoint or similar format), and a final draft of the presentation (PowerPoint or similar format) and notes. The topic will be either one of those listed at the end of this document, or an alternative topic. In either case, your topic **must be approved** by the instructor.

You do not have to deliver the presentation, but the research you do will be graded and must demonstrate knowledge of the topic. Your slides should be informative and should include notes (as part of the slide) which describe what is on the slide.

Deliverables

The deliverables for this project and their due dates are as follows.

Deliverable	Description	Due Date	Where Delivered
Research Project Description	Several paragraphs which explain what you will write about the the topic This should identify the content your project will contain and how that content will flow	6/6	Class Website (Wiki)
Detailed Outline	One or more pages of outline text which explains the flow and subtopics which you will present Note: you are not committed to using this outline in the final presentation	6/15	Class Website (Wiki)
Rough draft of presentation slides	PowerPoint (or similar, eg. OpenOffice) slides - uploaded to the Digital Dropbox on Blackboard	6/17	Blackboard Digital Dropbox

Deliverable	Description	Due Date	Where Delivered
Final presentation slides	PowerPoint (or similar, eg. OpenOffice) slides - uploaded to the Digital Dropbox on Blackboard	6/24	Blackboard Digital Dropbox

Grading

The following rubric will be used for grading

Element	Percentage
Clarity • clear and formatted so that it can be read easily • free from grammar and spelling mistakes	20
Completeness of coverage • covers all pertinent points for the topic • covers the topic in sufficient depth	30
Understanding of material • should demonstrate an understanding of the topic/material based on your research	10
Quality and use of sources • a variety of sources/sites should be used • quality, reputable sources/sites should be used • should use between 10 and 15 sources	20
Correctness • any statements made should be factually correct according to your sources • sources and information used should be current	20
Total	100

Searching for Information

This project requires you to use the Web to search for information for your research, and that as part of this effort, you use a specific search engine. For each source or [Web] site you find and examine using the search engine, you will be expected to record information about the web site/page. This information will indicate where you are in the search process, whether or not you think the document is relevant to your topic (whether you plan to use the document in your research) and what criteria you used to make that decision. The search engine you will use for this assignment will simplify the recording of this information. You will be shown how to use the search engine in class.3

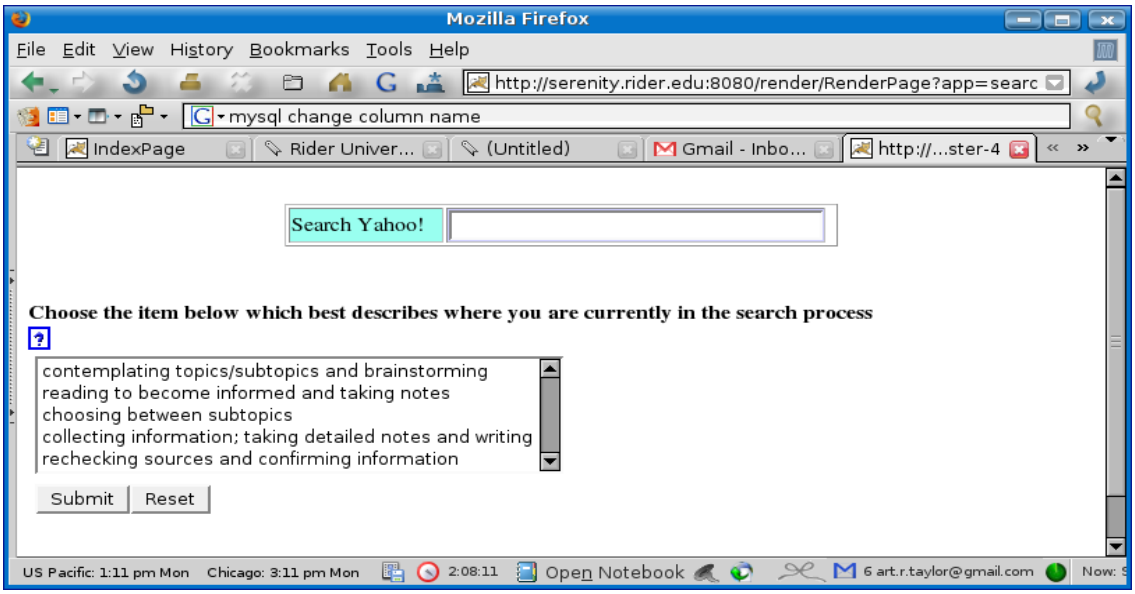
Originality of Work

You are expected to submit original material. Any detection of plagiarism or copying could result in a score of zero for the presentation in additional to other disciplinary action from the Rider CBA.

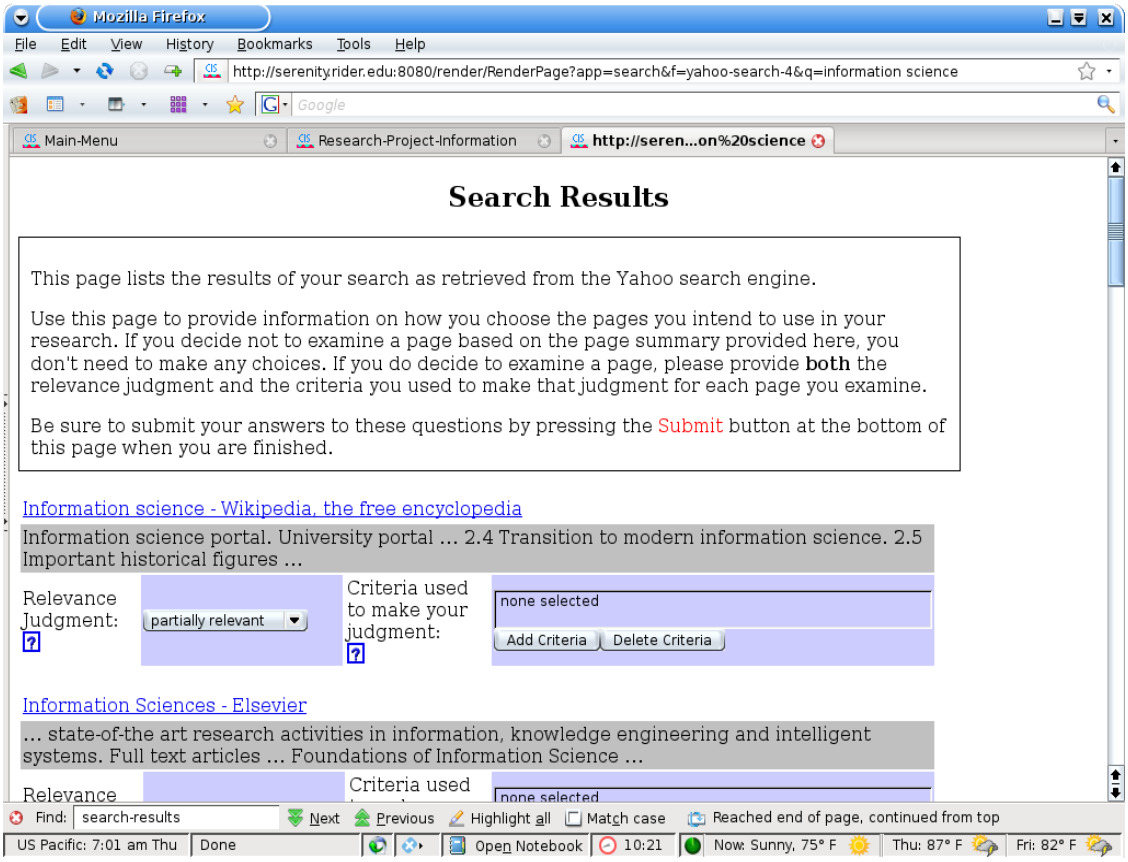
Current Topics

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 as a Monopoly and It's Influence on Information Technology
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

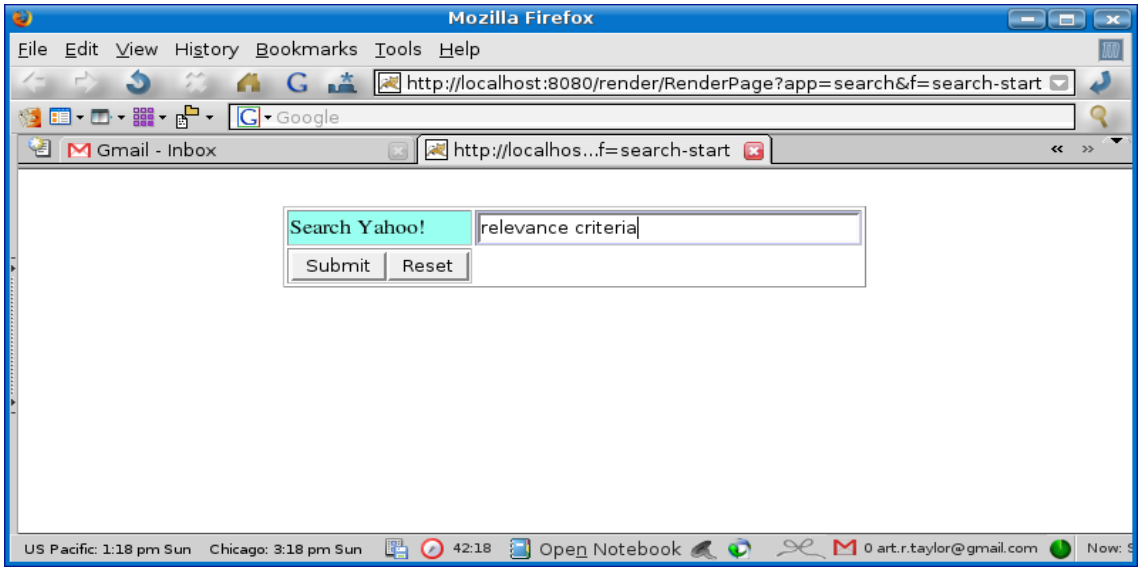
Appendix B: Search Engine Interface - Study 2 Pilot



Appendix C: Search Results - Study 2 Pilot

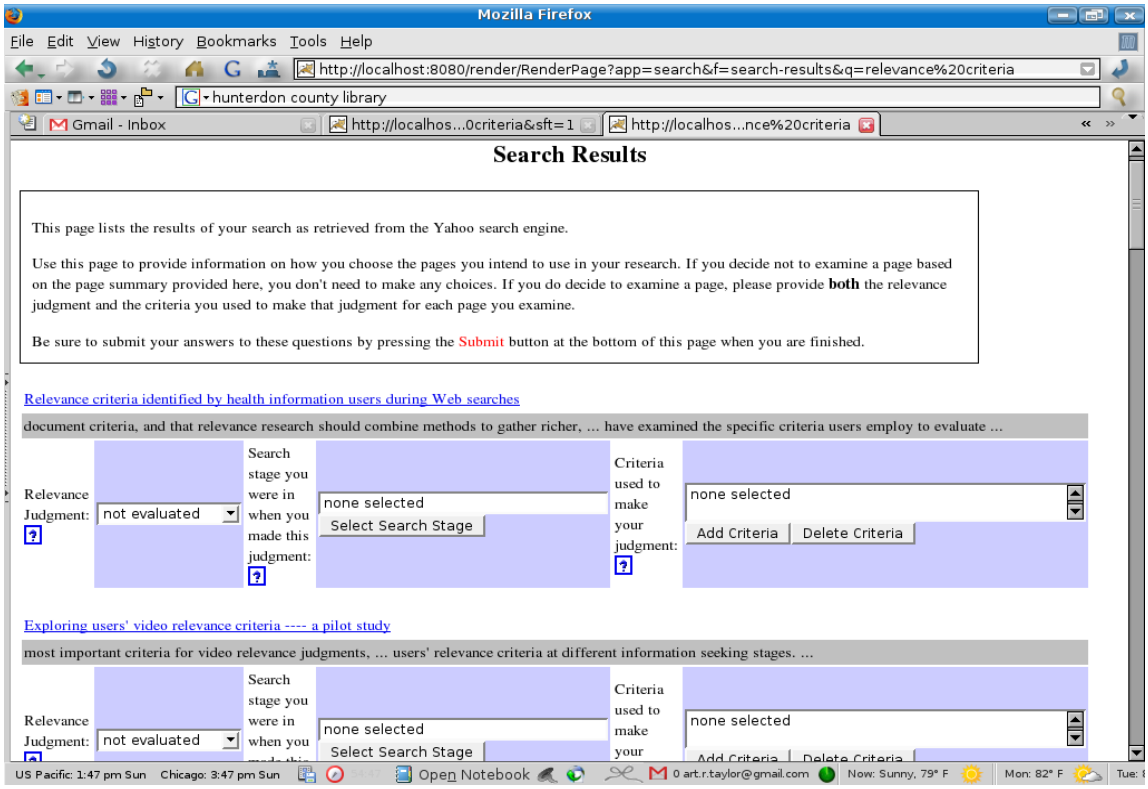


Appendix D: Search Engine Interface - Study 2 and Study 3

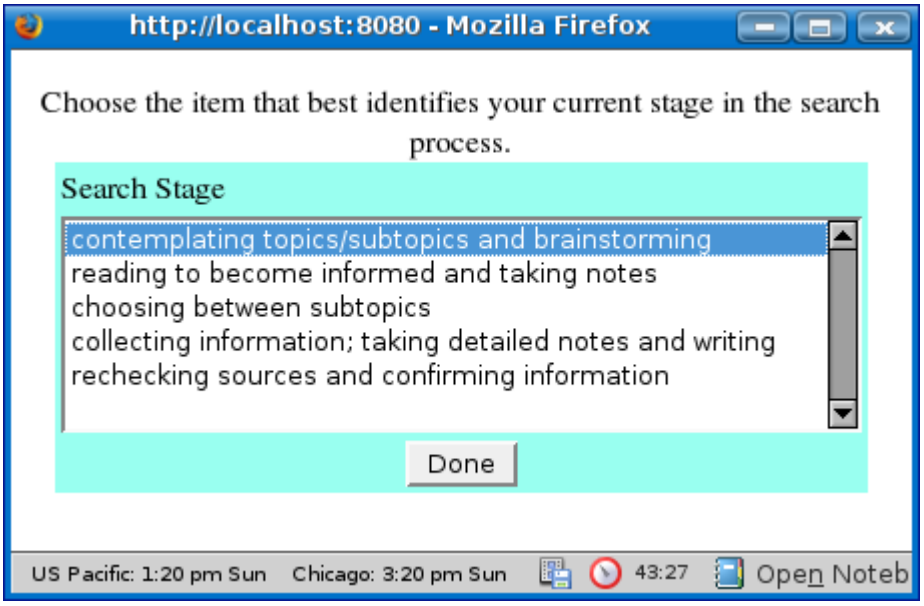


Appendix E: Search Results Page - Study 2 and Study 3

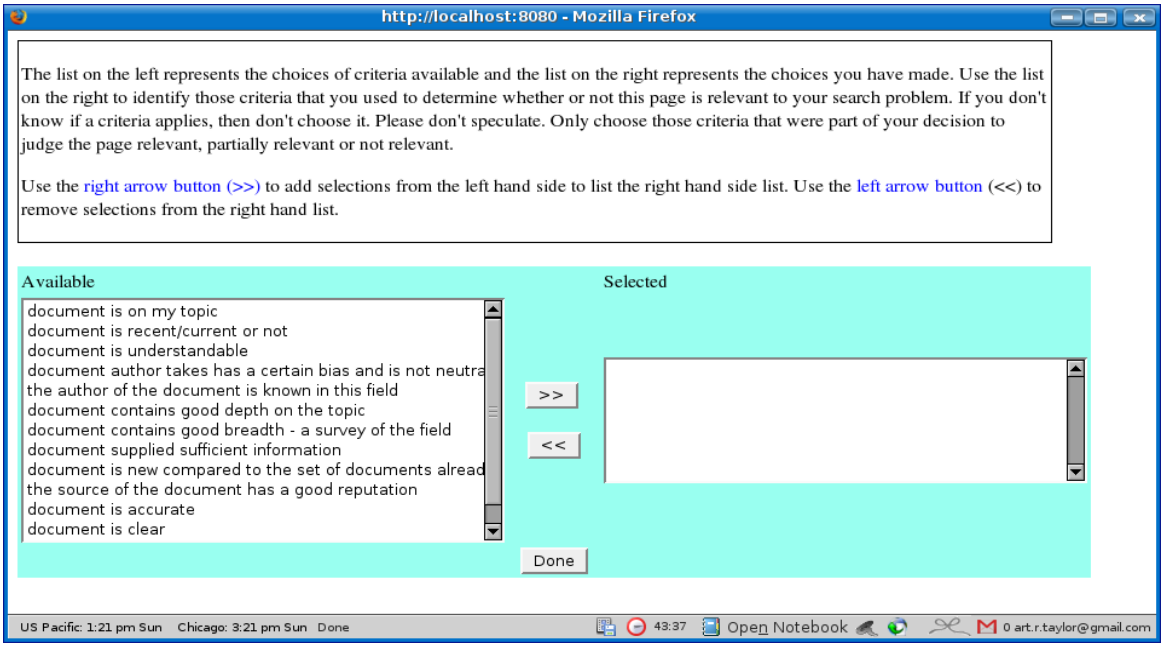
* Note: relevance judgment choices are presented in a drop-down list which presents a mutually exclusive choice of *relevant*, *not relevant*, and *partially relevant/unsure about relevance*.



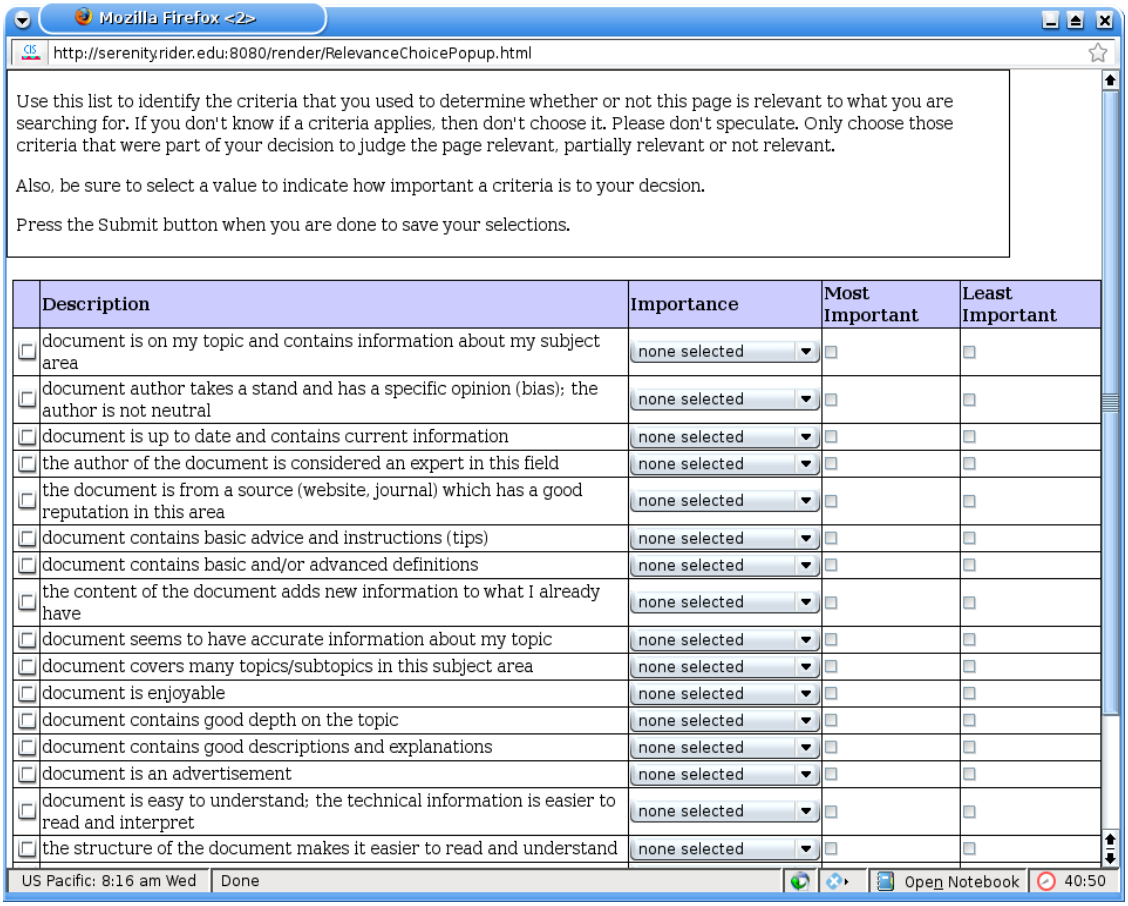
Appendix F: Search Stage Selection - Study 2 and Study 3



Appendix G: Criteria Selection - Study 2 Pilot and Study 2

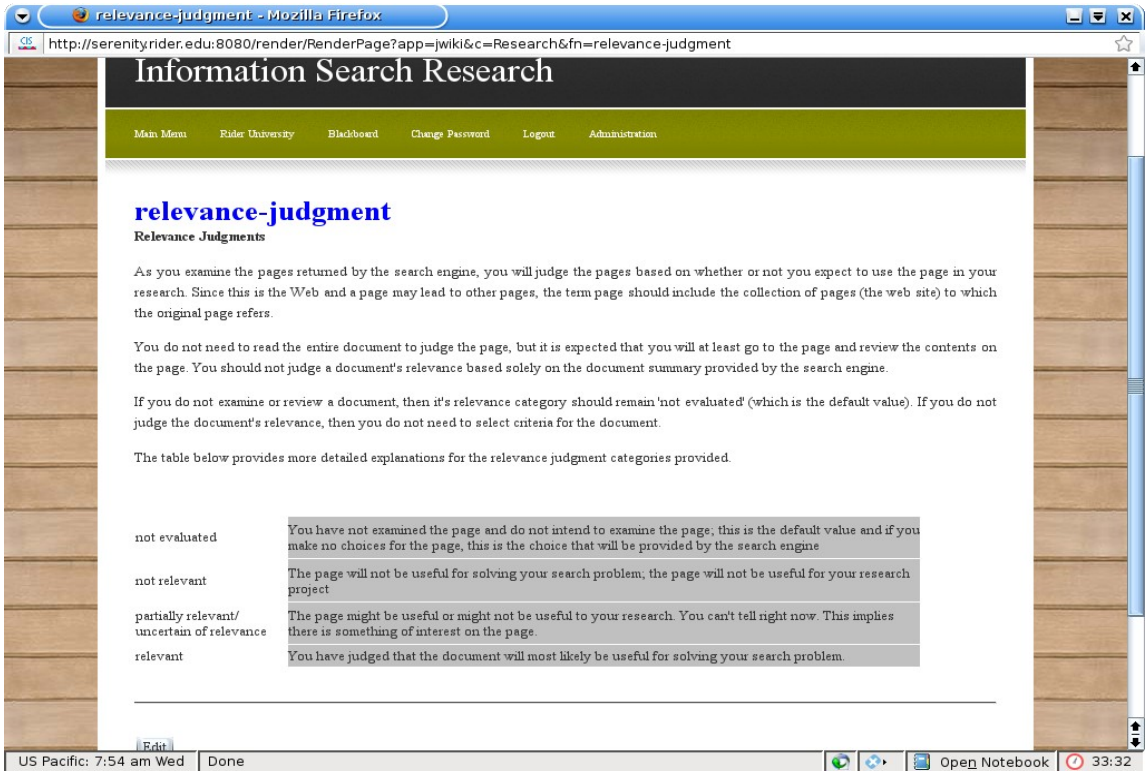


Appendix H: Criteria Selection and Assignment of Criteria Importance - Study 3

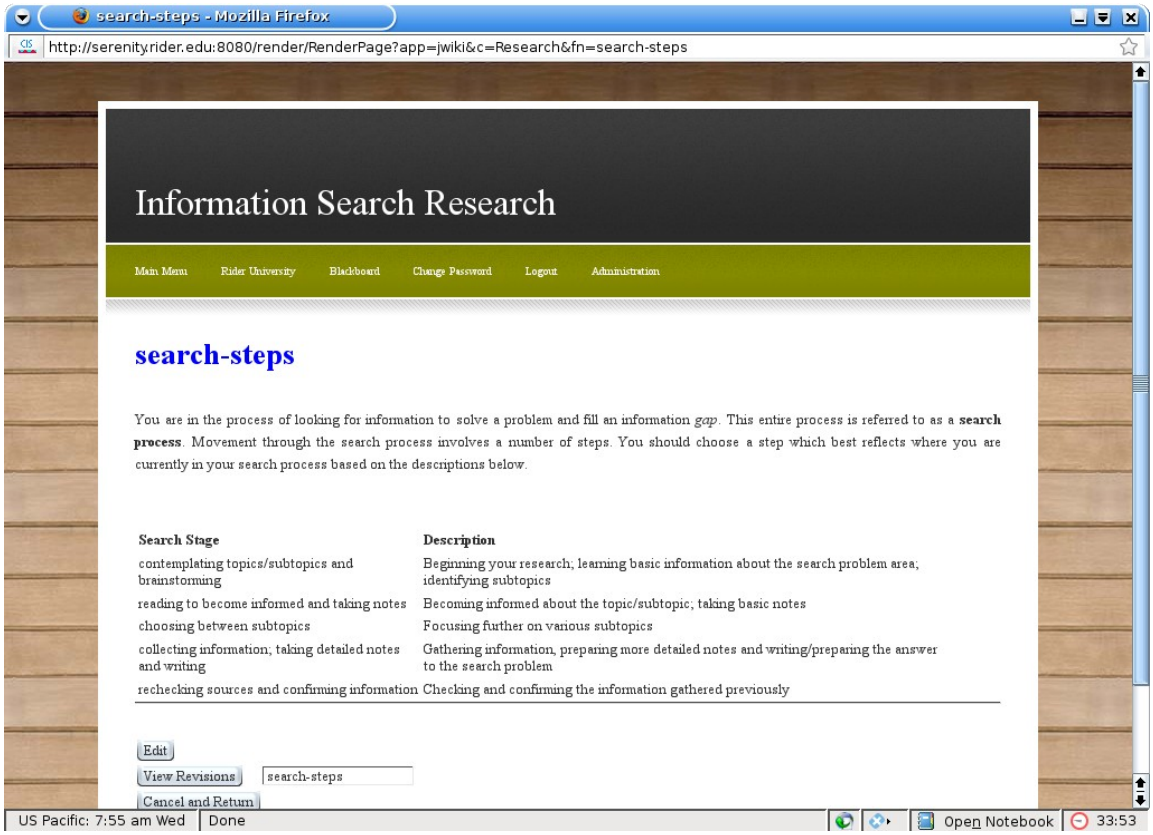


Importance drop-down list selections:	Not very important, Slightly important, Somewhat important, Important, Very Important

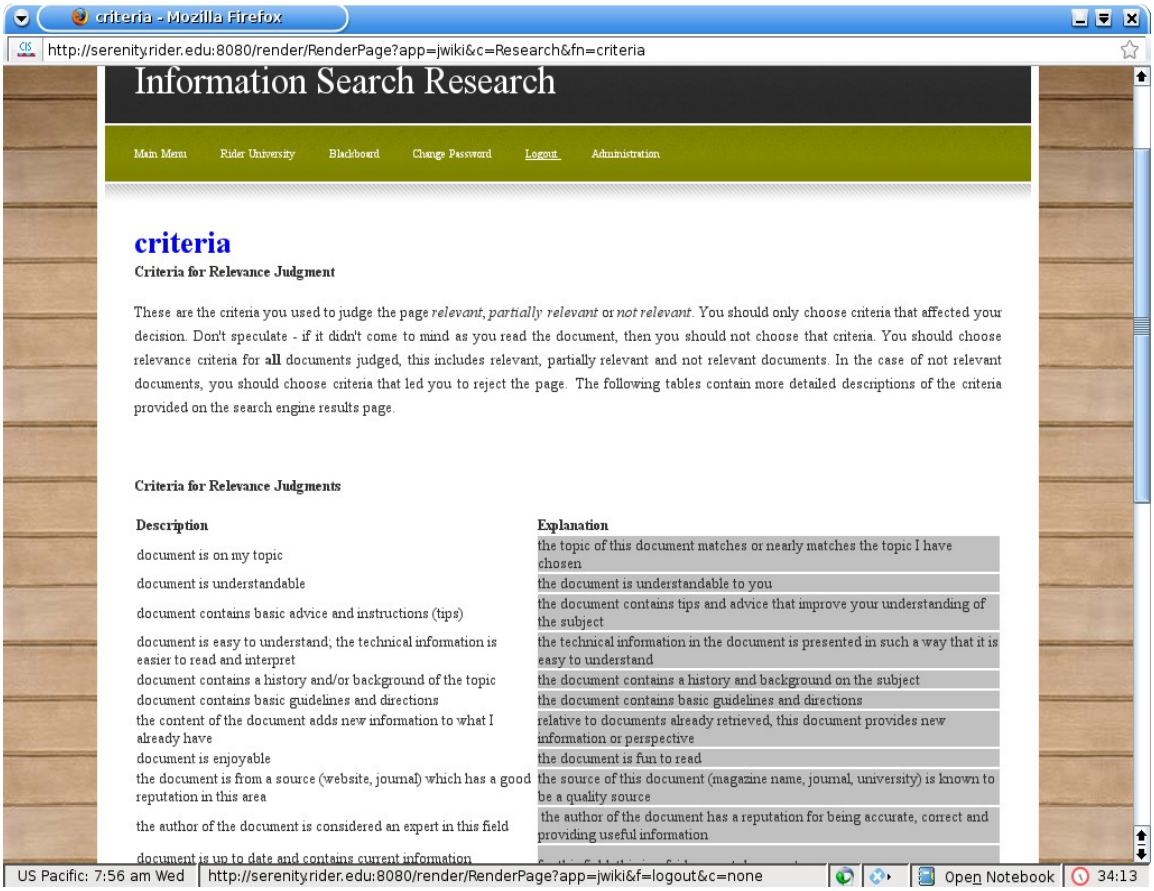
Appendix I: Relevance Judgment Help



Appendix J: Search Stage Help



Appendix K: Criteria for Relevance Judgment Help



Appendix L: Research Topics Assigned to Subjects - Study 2 and 3

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

Note: Subjects could add a topic if approved by the instructor; a very small percentage of subjects chose their own topic.

Appendix M: Collection of Demographic Information - Study 1

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 ___

Appendix N: Relevance Questions - Study 1

Indicate whether or not this document was relevant on a scale of 1 to 10, with 10 being most relevant and 1 being least relevant.

Question	Text
0	What was your reason for selecting/rejecting this document?
1	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Amount of information'
2	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Specificity'
3	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Clarity of presentation'
4	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Ability to Understand'
5	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Depth/Scope'
6	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Precision of Document'
7	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Recency of Document Publication'
9	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Interest in Topic'
10	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Instructional'
11	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Authority of Author'
12	On a scale of 1 to 10 indicate why you considered this document relevant/irrelevant based on the criteria of 'Bias of Author'

Appendix O: Search Stage Questions - Study 1

2. Which of the following applies to where you were in your search process when you accepted or rejected this document.	
☐a.) Becoming informed on the topic.	
☐b.) Learning about the topic.	
☐c.) Trying to focus on the topic/subtopic.	
☐d.) Defining and extending focus.	
☐e.) Browsing for information on the focus I've identified.	
☐f.) Extracting useful information.	
☐g.) Verifying information retrieved.	
☐h.) Completion and presentation of information.	

Appendix P: Search Question - Study 1

Search Test 5
Consider that you have been assigned the following question as part of an open book, open Internet exam. Conduct a search for documents which you would find useful in answering these questions. Attempt to find at least ten documents which you would find useful.

> *Compare and contrast the benefits of using a fixed exchange rate versus a flexible exchange rate for international transactions.*

Appendix Q: Post Research Survey - Study 2 and Study 3

1. How frequently do you perform online searches?	
a.) Once a month	
b.) Once a week	
c.) Once a day	
d.) Several times a day	

2. What is your current year in college?	
a.) Freshman	
b.) Sophomore	
c.) Junior	
d.) Senior	
e.) Graduate	

3. What is your age?

4. What is your gender?	
a.) Male	
b.) Female	

5. Is English your primary language?	
a.) Yes	
b.) No	

Appendix R: Criteria Selection Correlation Matrix - Study 2

Note: Correlation coefficient values in bold are those over the .15 threshold.

Table T-1: Partial Relevance Correlation Matrix - All Stages

Criteria	understand	topic	structure	source	recency	novelty	depth	breadth	auth	amt	accuracy
understand	1.000	-0.140	0.014	-0.013	-0.022	-0.020	**0.196**	0.105	0.076	-0.032	0.010
topic	-0.142	1.000	0.010	-0.088	0.139	0.028	0.020	-0.121	0.121	-0.122	-0.166
structure	0.014	0.010	1.000	-0.107	-0.075	-0.064	-0.013	0.040	**0.226**	0.042	0.149
source	-0.013	-0.090	-0.107	1.000	0.109	**0.174**	0.094	**0.183**	-0.040	0.052	0.211
recency	-0.022	0.140	-0.075	0.109	1.000	**0.282**	-0.085	-0.048	-0.065	-0.092	-0.092
novelty	-0.020	0.028	-0.064	**0.174**	**0.282**	1.000	-0.049	-0.054	-0.024	0.063	0.024
depth	**0.196**	0.020	-0.013	0.094	-0.085	-0.049	1.000	**0.129**	-0.049	**0.345**	0.098
breadth	0.105	-0.121	0.040	**0.183**	-0.048	-0.054	0.129	1.000	-0.054	0.071	0.089
authority	0.076	0.121	**0.226**	-0.040	-0.065	-0.024	-0.049	-0.054	1.000	0.063	0.024
amount	-0.032	-0.122	0.042	0.052	-0.092	0.063	**0.345**	0.071	0.063	1.000	0.098
accuracy	0.010	-0.166	**0.149**	**0.211**	-0.092	0.024	0.098	0.089	0.024	0.098	1.000

n=758

Table T-2:Relevant Documents Correlation Matrix - All Stages

Criteria	understand	topic	structure	source	recency	novelty	depth	breadth	authority	amount	accuracy
underst	1.000	0.123	**0.148**	**0.155**	0.131	0.067	**0.155**	**0.152**	0.108	**0.170**	**0.242**
topic	0.123	1.000	0.059	0.008	**0.161**	0.030	0.077	-0.016	-0.014	-0.048	0.013
structure	**0.148**	0.059	1.000	0.023	0.136	0.076	**0.174**	**0.158**	0.043	**0.164**	**0.186**
source	**0.155**	0.008	0.023	1.000	0.125	0.083	0.110	**0.209**	**0.157**	**0.17**	**0.187**
recency	0.131	**0.161**	0.136	0.125	1.000	0.089	0.053	0.139	**0.153**	0.038	0.100
novelty	0.067	0.030	0.076	0.083	0.089	1.000	0.056	0.042	**0.145**	0.078	0.088
depth	**0.155**	0.077	**0.174**	0.110	0.053	0.056	1.000	**0.179**	0.095	**0.289**	**0.197**
breadth	**0.152**	-0.016	**0.158**	0.209	0.139	0.042	**0.179**	1.000	0.076	**0.207**	**0.201**
authority	0.108	-0.014	0.043	**0.157**	**0.153**	**0.145**	0.095	0.076	1.000	0.073	0.070
amount	**0.170**	-0.048	**0.164**	**0.165**	0.038	0.078	**0.289**	**0.207**	0.073	1.000	0.108
accuracy	**0.242**	0.013	**0.186**	**0.187**	0.100	0.088	**0.197**	**0.201**	0.070	0.108	1.000

n=758

Table T-3: Relevant - Initiation Stage Correlation Matrix

Coefficient	underst	topic	structure	source	recency	novelty	depth	breadth	authority	amt	accur
underst	1.000	-0.004	0.208	**0.368**	**0.364**	-0.040	**0.153**	**0.377**	0.003	**0.363**	**0.402**
topic	-0.004	1.000	0.011	-0.111	**0.239**	-0.015	-0.107	-0.221	-0.153	-0.139	-0.069
structure	**0.208**	0.011	1.000	0.038	0.048	-0.118	**0.270**	0.078	-0.070	**0.322**	**0.223**
source	**0.368**	-0.111	0.038	1.000	**0.155**	**0.176**	**0.149**	**0.471**	0.168	**0.277**	**0.385**
recency	**0.364**	0.239	0.048	**0.155**	1.000	-0.045	**0.179**	0.107	-0.026	0.113	**0.295**
novelty	-0.040	-0.015	-0.118	**0.176**	-0.045	1.000	**0.216**	0.087	0.103	0.111	0.104
depth	**0.153**	-0.107	**0.270**	**0.149**	**0.179**	**0.216**	1.000	0.018	**0.174**	**0.352**	**0.294**
breadth	**0.377**	-0.221	0.078	**0.471**	0.107	0.087	0.018	1.000	0.017	**0.483**	**0.273**
authority	0.003	-0.153	-0.070	**0.168**	-0.026	0.103	**0.174**	0.017	1.000	-0.049	0.076
amount	**0.363**	-0.139	**0.322**	**0.277**	0.113	0.111	**0.352**	**0.483**	-0.049	1.000	**0.294**
affectiveness	0.132	0.065	-0.074	**0.223**	**0.220**	-0.032	**0.192**	**0.237**	**0.485**	**0.175**	**0.177**
accuracy	**0.402**	-0.069	**0.223**	**0.385**	**0.295**	0.104	**0.294**	**0.273**	0.076	**0.294**	1.000

n=104

Table T-4: Exploration Stage Correlation Matrix

Criteria	underst	topic	struct	source	recency	novelty	breadth	authority	amt	accur	depth
understand	1.000	0.175	**0.227**	0.062	0.073	-0.095	0.100	**0.166**	**0.245**	**0.170**	0.088
topic	**0.175**	1.000	0.132	0.075	0.071	0.104	-0.117	-0.165	-0.077	-0.015	-0.028
structure	**0.227**	0.132	1.000	0.057	0.028	-0.115	**0.245**	0.110	0.159	**0.152**	0.142
source	0.062	0.075	0.057	1.000	-0.045	-0.056	0.135	**0.210**	-0.081	**0.159**	0.104
recency	0.073	0.071	0.028	-0.045	1.000	0.128	0.040	0.006	0.040	0.012	0.020
novelty	-0.095	0.104	-0.115	-0.056	0.128	1.000	-0.113	-0.068	-0.113	-0.077	-0.034
breadth	0.100	-0.117	**0.245**	0.135	0.040	-0.113	1.000	-0.065	-0.047	**0.207**	0.113
authority	**0.166**	-0.165	0.110	**0.210**	0.006	-0.068	-0.065	1.000	**0.179**	-0.001	0.082
amount	**0.245**	-0.077	**0.159**	-0.081	0.040	-0.113	-0.047	**0.179**	1.000	-0.060	**0.404**
accuracy	**0.170**	-0.015	**0.152**	**0.159**	0.012	-0.077	**0.207**	-0.001	-0.060	1.000	0.127
depth	0.088	-0.028	0.142	0.104	0.020	-0.034	0.113	0.082	**0.404**	0.127	1.000

n=128

Table T-5: Extracting Stage Correlation Matrix

Criteria	underst	topic	structure	source	recency	novelty	depth	breadth	authority	amount	accuracy
understand	1.000	0.082	0.125	0.108	0.136	0.130	0.142	0.136	**0.201**	0.015	**0.224**
topic	0.082	1.000	0.038	0.009	**0.169**	0.025	0.126	0.051	0.101	-0.088	0.015
structure	0.125	0.038	1.000	0.013	**0.223**	0.130	0.133	0.120	0.063	0.088	**0.168**
source	0.108	0.009	0.013	1.000	**0.167**	0.056	0.141	**0.148**	0.101	0.140	**0.158**
recency	0.136	**0.169**	**0.223**	0.167	1.000	0.119	0.063	**0.176**	**0.266**	0.046	0.107
novelty	0.130	0.025	0.130	0.056	0.119	1.000	0.063	0.085	**0.221**	0.102	0.104
depth	0.142	0.126	0.133	0.141	0.063	0.063	1.000	**0.207**	**0.168**	**0.263**	0.131
breadth	0.136	0.051	0.120	**0.148**	**0.176**	0.085	**0.207**	1.000	**0.189**	**0.201**	**0.197**
authority	**0.201**	0.101	0.063	0.101	**0.266**	**0.221**	**0.168**	**0.189**	1.000	0.072	**0.151**
amount	0.015	-0.088	0.088	0.140	0.046	0.102	**0.263**	**0.201**	0.072	1.000	0.066
accuracy	**0.224**	0.015	**0.168**	**0.158**	0.107	0.104	0.131	**0.197**	**0.151**	0.066	1.000

n=355

Table T-6: Verifying Stage Correlation Matrix

Criteria	underst	topic	structure	source	recency	novelty	depth	breadth	authority	amount	accuracy
underst	1.000	**0.242**	-0.118	0.000	-0.041	-0.053	0.000	-0.100	-0.127	**0.309**	-0.106
topic	**0.242**	1.000	**0.149**	-0.125	0.207	0.085	**0.204**	-0.125	-0.181	**0.168**	0.128
structure	-0.118	**0.149**	1.000	**-0.236**	**0.155**	0.144	**0.285**	**0.220**	-0.032	**0.164**	**0.246**
source	0.000	-0.125	**-0.236**	1.000	**0.145**	**0.159**	-0.268	-0.160	**0.339**	**0.201**	-0.120
recency	-0.041	**0.207**	**0.155**	**0.145**	1.000	**0.182**	-0.003	**0.504**	0.100	0.131	**0.156**
novelty	-0.053	0.085	0.144	**0.159**	**0.182**	1.000	-0.020	0.005	-0.011	0.127	**0.241**
depth	0.000	**0.204**	**0.285**	-0.268	-0.003	-0.020	1.000	0.164	-0.164	-0.164	**0.370**
breadth	-0.100	-0.125	**0.220**	-0.160	**0.504**	0.005	**0.164**	1.000	-0.153	-0.023	0.085
authority	-0.127	-0.181	-0.032	**0.339**	0.100	-0.011	-0.164	-0.153	1.000	0.144	-0.136
amount	**0.309**	**0.168**	**0.164**	**0.201**	0.131	0.127	-0.164	-0.023	0.144	1.000	-0.027
accuracy	-0.106	0.128	**0.246**	-0.120	**0.156**	**0.241**	**0.370**	0.085	-0.136	-0.027	1.000

n=59

Appendix S: Frequency Counts for Groups of Relevance Criteria by Search Stage - Study 3

Frequency Count for Groups of Relevance Criteria

Group	initiation	exploration	differentiating	extracting	verifying
1	5	5	2	23	1
2	21	18	15	71	3
3	19	6	9	46	2
4	18	6	5	36	2
5	29	27	16	105	10
6	27	20	14	80	12
7	15	1	5	33	4
8	4	3	3	11	4
9	14	16	7	59	9
10	20	16	7	53	3
12	21	5	5	40	4
13	21	5	5	40	4
14	10	2	3	27	3
15	8	10	3	47	5
16	8	9	3	34	6
17	2	4	3	26	3
18	21	18	15	71	3
19	5	5	2	23	1
20	8	11	11	41	4
21	19	6	9	46	2
22	14	17	7	48	4
23	4	5	2	26	2

Appendix T: Relevance Criteria List - Study 3

Criteria	Description
accuracy	document seems to have accurate information about my topic
advertisement	document is an advertisement
affectiveness	document is enjoyable
authority	the author of the document is considered an expert in this field
bias	document author takes a stand and has a specific opinion (bias); the author is not neutral
breadth	document covers many topics/subtopics in this subject area
definitions	document contains basic and/or advanced definitions
depth	document contains good depth on the topic
descriptions	document contains good descriptions and explanations
guidelines	document contains basic guidelines and directions
history	document contains a history and/or background of the topic
novelty	the content of the document adds new information to what I already have
recency	document is up to date and contains current information
source	the document is from a source (website, journal) which has a good reputation in this area
structure	the structure of the document makes it easier to read and understand
time	document is useful because of time constraints
tips	document contains basic advice and instructions (tips)
topic	document is on my topic and contains information about my subject area
understandability	document is easy to understand; the technical information is easier to read and interpret

Appendix U: Selection of Criteria by Criteria Code Weight - Study 3

Criteria	code_weight	initiation	differentiating	exploration	extracting	verifying
accuracy	2	1	1		1	
	3			1	2	1
	4	5	9	14	27	11
	5	9	17	15	38	1
advertise	1	1		1	3	
	2				1	1
	3		1		1	2
	4	1	1		1	
	5		1			
affective	1	3	13	4	7	2
	2		2	1	3	
	3		4	5	19	
	4	4	3	5	13	1
	5	2		1	1	1
authority	1					1
	2	1		1		1
	3	1	3	1	6	1
	4	3	3	2	12	4
	5	3	1	2	9	5
bias	1	1			2	
	2	1			1	2
	3		2	2	4	1
	4	4	1	3	5	
	5				2	1
breadth	1	1			1	
	2	3	1		2	2
	3		2	1	8	1
	4	4	8	12	12	2
	5	2	6	5	11	1
definition	1				1	
	2	1	2	1	5	
	3	1	4	3	9	
	4	1	4	5	13	8
	5	1	3	8	4	1
depth	1			1	1	
	2			3	3	1
	3	1	3	1	7	10
	4	4	12	10	31	6
	5	1	13	3	13	3
descript	1					1
	2				3	2
	3			1	3	2
	4	3	7	8	41	4

Criteria	code_weight	initiation	differentiating	exploration	extracting	verifying
	5	3	7	7	8	3
guideline	1	1				
	2	1	1	1	2	
	3	1	2	3	5	1
	4		1	5	13	3
	5	1	1	1	4	
history	2	2	1		1	
	3	4	4	2		1
	4	2	1	5	7	2
	5	1	4	4	6	
novelty	1	2			1	1
	2		1	1	4	4
	3	2		2	9	4
	4	6	8	8	20	4
	5	3	12	8	22	3
recency	1				1	
	2	3	2	5	7	2
	3	2	2		6	6
	4	6	3	8	23	10
	5	3	9	6	22	4
source	1	3			2	
	2				2	1
	3	1	2	2	4	
	4	4	5	6	16	5
	5	1	3	3	13	7
structure	1	1			3	1
	2	4	1	1	6	3
	3		5	5	14	3
	4	7	3	4	26	2
	5	2	7	1	3	1
time	1	1		1	2	
	2	2	2		1	1
	3	1	1		6	2
	4	1	2	2	5	1
	5	1	3	1	5	1
tips	1	1	1	1	2	
	2				5	
	3			1	7	1
	4	5	1	3	7	4
	5			2	3	
topic	1		1		1	
	2	2	1		3	1
	3	1		3	4	2
	4	10	14	23	57	25
	5	13	31	29	67	11
understand	1	2		1		

Criteria	code_weight	initiation	differentiating	exploration	extracting	verifying
	2	1		2	7	4
	3	2	3	1	16	5
	4	5	5	9	31	3
	5	2	5	6	6	5

Appendix V: Criteria Selection Correlation Matrix - Study 3

Note: Correlation coefficient values in bold are those over the .15 threshold.

Correlation Matrix: Relevant - All Search Stages

	understand	tips	time	structure	source	recency	novel	history	guides	descr	depth	def	breadth	auth	affective	accuracy
understand	1.000	0.115	0.133	**0.196**	0.007	**0.263**	0.071	0.086	0.129	0.207	**0.173**	0.094	0.113	0.006	0.018	0.130
tips	0.115	1.000	0.113	**0.218**	0.065	0.114	0.092	-0.046	**0.360**	**0.151**	-0.009	0.097	-0.002	0.014	0.037	0.119
time	0.133	0.113	1.000	**0.153**	0.116	0.133	0.109	0.132	0.151	0.029	0.087	**0.152**	0.065	-0.024	0.105	0.107
structure	**0.196**	**0.218**	0.153	1.000	0.055	0.270	0.115	0.091	**0.216**	**0.245**	0.096	0.143	0.076	-0.013	0.022	**0.178**
source	0.007	0.065	0.116	0.055	1.000	**0.150**	**0.189**	0.055	0.106	0.135	0.081	**0.160**	0.065	0.109	0.041	**0.165**
recency	**0.263**	0.114	0.133	**0.270**	0.150	1.000	0.108	0.085	0.097	0.088	**0.152**	0.111	0.039	0.119	0.032	**0.169**
novelty	0.071	0.092	0.109	0.115	**0.189**	0.108	1.000	-0.002	0.122	**0.241**	0.105	**0.177**	0.103	0.043	0.107	**0.158**
history	0.086	-0.046	0.132	0.091	0.055	0.085	-0.002	1.000	0.039	0.074	0.090	0.107	0.118	0.048	0.061	0.073
guidelines	0.129	**0.360**	**0.151**	0.216	0.106	0.097	0.122	0.039	1.000	**0.200**	-0.023	**0.195**	0.113	0.067	**0.232**	**0.174**
description	**0.207**	0.151	0.029	**0.245**	0.135	0.088	**0.241**	0.074	**0.200**	1.000	**0.187**	**0.212**	**0.208**	0.017	0.125	**0.200**
depth	0.173	-0.009	0.087	0.096	0.081	0.152	0.105	0.090	-0.023	0.187	1.000	0.052	**0.198**	0.111	0.136	**0.158**
definitions	0.094	0.097	**0.152**	0.143	**0.160**	0.111	0.177	0.107	**0.195**	**0.212**	0.052	1.000	0.129	-0.014	0.045	0.144
breadth	0.113	-0.002	0.065	0.076	0.065	0.039	0.103	0.118	0.113	**0.208**	**0.198**	0.129	1.000	0.061	0.221	**0.163**
authority	0.006	0.014	-0.02	-0.01	0.109	0.119	0.043	0.048	0.067	0.017	0.111	-0.014	0.061	1.000	-0.009	0.041
affective	0.018	0.037	0.105	0.022	0.041	0.032	0.107	0.061	0.232	0.125	0.136	0.045	**0.221**	-0.009	1.000	**0.229**
accuracy	0.130	0.119	0.107	**0.178**	**0.165**	**0.169**	**0.158**	0.073	0.174	**0.200**	**0.158**	0.144	**0.163**	0.041	**0.229**	1.000

n=657

Correlation Matrix: Initiating Stage

	understand	tips	time	structure	source	def	recent	novel	history	guide	descr	depth	breadth	auth	affect	advert	accur
understand	1.000	0.037	0.037	0.102	0.047	**0.330**	**0.328**	**0.253**	0.047	0.150	**0.189**	**0.189**	0.009	**0.495**	0.047	**0.352**	-0.043
tips	0.037	1.000	0.024	-0.153	-0.053	0.096	-0.153	**0.163**	-0.053	0.096	0.024	0.024	-0.074	-0.030	0.114	-0.094	-0.028
time	0.037	0.024	1.000	0.138	-0.053	**0.329**	-0.007	0.014	0.447	**0.562**	0.024	**0.219**	0.086	0.144	0.114	**0.227**	-0.028
structure	0.102	-0.153	0.138	1.000	0.115	0.110	**0.241**	-0.049	-0.009	0.110	**0.284**	0.138	**0.310**	**0.165**	0.115	-0.163	0.094
source	0.047	-0.053	-0.053	0.115	1.000	0.024	0.115	**0.145**	0.003	0.024	-0.053	-0.053	-0.027	0.334	-0.139	-0.120	**0.209**
def	**0.330**	0.096	**0.329**	0.110	0.024	1.000	**0.457**	**0.483**	**0.223**	**0.446**	**0.562**	**0.562**	**0.196**	**0.253**	**0.421**	**0.307**	**0.262**
recency	**0.328**	-0.153	-0.007	**0.241**	0.115	**0.457**	1.000	**0.173**	0.115	**0.283**	0.138	0.138	**0.190**	0.035	**0.239**	**0.314**	-0.013
novelty	**0.253**	**0.163**	0.014	-0.049	0.145	**0.483**	**0.173**	1.000	-0.108	0.129	**0.459**	**0.163**	-0.143	0.061	**0.145**	**0.332**	**0.353**
history	0.047	-0.053	**0.447**	-0.009	0.003	**0.223**	0.115	-0.108	1.000	**0.421**	0.114	**0.447**	0.110	**0.185**	0.003	-0.120	-0.158
guide	**0.150**	0.096	**0.562**	0.110	0.024	**0.446**	**0.283**	0.129	**0.421**	1.000	0.096	0.096	**0.196**	**0.253**	**0.421**	**0.307**	-0.079

descr	**0.189**	0.024	0.024	**0.284**	-0.053	**0.562**	**0.138**	**0.459**	0.114	0.096	1.000	**0.414**	**0.247**	-0.030	**0.281**	-0.094	**0.402**
depth	**0.189**	0.024	**0.219**	0.138	-0.053	**0.562**	**0.138**	**0.163**	**0.447**	0.096	**0.414**	1.000	**0.247**	**0.319**	0.114	-0.094	0.115
breadth	0.009	-0.074	0.086	**0.310**	-0.027	**0.196**	**0.190**	-0.143	0.110	**0.196**	**0.247**	**0.247**	1.000	**0.150**	0.110	-0.129	**0.158**
auth	**0.495**	-0.030	0.144	0.165	0.334	**0.253**	0.035	0.061	**0.185**	**0.253**	-0.030	**0.319**	**0.150**	1.000	0.036	0.174	0.009
affect	0.047	0.114	0.114	0.115	-0.139	**0.421**	**0.239**	**0.145**	0.003	**0.421**	**0.281**	0.114	0.110	0.036	1.000	0.153	0.331
advert	**0.352**	-0.094	**0.227**	-0.163	-0.120	**0.307**	**0.314**	**0.332**	-0.120	**0.307**	-0.094	-0.094	-0.129	**0.174**	**0.153**	1.000	0.063
accur	-0.043	-0.028	-0.028	0.094	0.209	**0.262**	-0.013	**0.353**	-0.158	-0.079	**0.402**	0.115	**0.158**	0.009	**0.331**	0.063	1.000

n=42

Correlation Matrix: Exploration Stage

	understand	tips	time	struct	source	recent	novel	hist	guide	descr	depth	def	breadth	auth	affect	advert	accur
understand	1.000	0.120	**0.266**	-0.083	**0.181**	**0.281**	-0.006	**0.269**	0.121	0.207	0.379	**0.183**	0.087	-0.184	-0.021	-0.072	**0.256**
tips	0.120	1.000	**0.329**	**0.250**	0.120	-0.200	0.120	-0.011	**0.409**	0.048	0.133	-0.075	-0.084	-0.100	0.048	-0.040	**0.291**
time	**0.266**	**0.329**	1.000	-0.105	**0.233**	-0.010	-0.010	0.064	0.077	**0.161**	0.138	0.006	0.138	-0.074	0.014	-0.029	**0.162**
struct	-0.083	**0.250**	-0.105	1.000	-0.076	**0.181**	0.093	-0.076	0.274	0.049	0.108	0.033	0.019	-0.130	0.049	-0.051	**0.185**
source	**0.181**	0.120	**0.233**	-0.076	1.000	**0.181**	**0.181**	0.032	-0.061	**0.514**	0.019	**0.216**	0.019	0.010	-0.138	-0.051	**0.343**
recency	**0.281**	-0.200	-0.010	**0.181**	**0.181**	1.000	0.138	**0.181**	-0.062	**0.207**	**0.306**	**0.183**	0.087	0.045	-0.098	-0.072	0.127
novel	-0.006	0.120	-0.010	0.093	**0.181**	**0.138**	1.000	0.005	0.121	**0.207**	**0.160**	0.034	**0.233**	-0.069	0.131	-0.072	**0.320**
history	**0.269**	-0.011	0.064	-0.076	0.032	**0.181**	0.005	1.000	-0.173	0.049	**0.198**	0.033	0.019	-0.130	0.049	-0.051	**0.264**
guide	0.121	**0.409**	0.077	**0.274**	-0.061	-0.062	0.121	-0.173	1.000	0.169	0.043	-0.037	-0.050	0.023	**0.266**	-0.048	**0.309**
descr	**0.207**	0.048	**0.161**	0.049	**0.514**	0.207	**0.207**	0.049	**0.169**	1.000	**0.306**	**0.250**	**0.151**	0.079	-0.130	-0.064	**0.221**
depth	**0.379**	0.133	0.138	0.108	0.019	**0.306**	**0.160**	**0.198**	0.043	0.306	1.000	0.128	0.107	0.056	-0.082	-0.070	-0.040
def	**0.183**	-0.075	0.006	0.033	0.216	**0.183**	0.034	0.033	-0.037	**0.250**	0.128	1.000	**0.204**	-0.052	-0.066	-0.067	**0.255**
breadth	0.087	-0.084	0.138	0.019	0.019	0.087	**0.233**	0.019	-0.050	**0.151**	0.107	**0.204**	1.000	0.056	**0.151**	-0.070	**0.157**
auth	-0.184	-0.100	-0.074	-0.130	0.010	0.045	-0.069	-0.130	0.023	0.079	0.056	-0.052	0.056	1.000	-0.043	-0.036	-0.157
affect	-0.021	0.048	0.014	0.049	-0.138	-0.098	0.131	0.049	**0.266**	-0.130	-0.082	-0.066	0.151	-0.043	1.000	-0.064	**0.153**
advert	-0.072	-0.040	-0.029	-0.051	-0.051	-0.072	-0.072	-0.051	-0.048	-0.064	-0.070	-0.067	-0.070	-0.036	-0.064	1.000	-0.102
accur	**0.256**	**0.291**	0.162	**0.185**	**0.343**	**0.127**	**0.320**	**0.264**	**0.309**	0.221	-0.040	0.255	**0.157**	-0.157	0.153	-0.102	1.000

n=72

Correlation Matrix: Extracting Stage

	understand	tips	time	struct	source	recent	novel	hist	guide	descr	depth	def	breadth	auth	affect	advert	accur
understand	1.000	0.132	0.027	**0.237**	-0.060	**0.203**	0.085	-0.026	0.132	0.210	**0.166**	-0.016	**0.228**	-0.029	0.103	-0.062	0.061
tips	**0.132**	1.000	**0.177**	**0.266**	0.005	**0.266**	0.053	-0.049	**0.343**	0.120	-0.008	0.113	0.100	0.063	0.049	0.105	0.065
time	0.027	**0.177**	1.000	0.097	**0.222**	**0.181**	0.082	0.036	**0.228**	0.047	0.048	**0.220**	0.021	-0.092	**0.149**	0.138	0.108

	understr	tips	time	struct	source	recent	novel	hist	guide	descr	depth	def	breadth	auth	affect	advert	accur
struct	**0.237**	**0.266**	0.097	1.000	0.012	**0.220**	0.073	0.087	**0.168**	0.222	0.081	0.117	0.038	-0.023	-0.008	-0.046	0.109
source	-0.060	0.005	**0.222**	0.012	1.000	0.110	**0.157**	0.107	0.079	0.077	**0.275**	0.087	**0.172**	**0.168**	0.101	-0.090	0.114
recent	**0.203**	**0.266**	**0.181**	**0.220**	0.110	1.000	-0.004	-0.023	0.104	0.028	0.126	0.049	0.088	0.137	0.083	-0.061	0.117
novel	0.085	0.053	0.082	0.073	**0.157**	-0.004	1.000	0.074	0.086	0.242	0.051	**0.188**	0.046	0.088	0.048	-0.055	-0.046
hist	-0.026	-0.049	0.036	0.087	0.107	-0.023	0.074	1.000	0.009	0.161	-0.010	0.029	0.126	0.112	-0.015	-0.053	-0.095
guide	0.132	**0.343**	**0.228**	**0.168**	0.079	0.104	0.086	0.009	1.000	0.153	0.025	**0.195**	**0.220**	**0.150**	**0.304**	0.105	0.129
descr	**0.210**	0.120	0.047	0.222	0.077	0.028	**0.242**	**0.161**	**0.153**	1.000	**0.154**	0.100	**0.255**	0.059	**0.188**	0.012	0.128
depth	**0.166**	-0.008	0.048	0.081	**0.275**	0.126	0.051	-0.010	0.025	0.154	1.000	-0.051	**0.261**	0.093	**0.192**	-0.053	**0.204**
def	-0.016	0.113	**0.220**	**0.117**	0.087	0.049	**0.188**	0.029	**0.195**	0.100	-0.051	1.000	-0.036	-0.030	0.083	-0.085	0.061
breadth	**0.228**	0.100	0.021	0.038	0.172	0.088	0.046	0.126	**0.220**	0.255	**0.261**	-0.036	1.000	0.117	**0.231**	-0.010	**0.213**
auth	-0.029	0.063	-0.092	-0.023	**0.168**	0.137	0.088	0.112	**0.150**	0.059	0.093	-0.030	0.117	1.000	0.095	-0.077	0.125
affect	0.103	0.049	**0.149**	-0.008	0.101	0.083	0.048	-0.015	**0.304**	0.188	**0.192**	0.083	**0.231**	0.095	1.000	0.115	**0.239**
advert	-0.062	0.105	0.138	-0.046	-0.090	-0.061	-0.055	-0.053	0.105	0.012	-0.053	-0.085	-0.010	-0.077	0.115	1.000	0.048
accur	0.061	0.065	0.108	0.109	0.114	0.117	-0.046	-0.095	0.129	0.128	**0.204**	0.061	**0.213**	0.125	**0.239**	0.048	1.000

n=182

Correlation Matrix: Verifying Stage

	accur	advert	affect	auth	breadth	def	depth	descr	guide	hist	novel	recent	source	struct	time	tips	understr
accur	1.000	-0.088	0.015	0.029	0.081	**0.277**	**0.200**	**0.281**	**0.493**	**0.226**	**0.579**	**0.441**	0.338	**0.427**	-0.016	**0.323**	**0.182**
advert	-0.088	1.000	-0.049	-0.091	0.135	0.078	-0.130	-0.083	-0.039	-0.042	**0.268**	-0.131	-0.088	-0.073	**0.158**	-0.046	0.016
affect	0.015	-0.049	1.000	-0.139	0.131	0.063	**0.232**	**0.179**	-0.060	-0.064	0.126	0.063	**0.164**	0.050	-0.084	-0.070	-0.175
auth	0.029	-0.091	-0.139	1.000	-0.173	-0.090	0.077	-0.142	-0.112	0.073	0.061	0.119	0.029	-0.108	-0.157	-0.131	-0.142
breadth	0.081	0.135	0.131	-0.173	1.000	**0.288**	0.111	0.096	-0.075	-0.079	**0.157**	-0.030	-0.043	-0.005	**0.299**	-0.087	0.030
def	**0.277**	0.078	0.063	-0.090	**0.288**	1.000	0.080	**0.489**	**0.456**	-0.089	**0.308**	0.129	**0.277**	**0.246**	0.034	**0.406**	**0.220**
depth	**0.200**	-0.130	**0.232**	0.077	0.111	0.080	1.000	**0.152**	-0.159	-0.006	0.111	0.099	0.121	0.134	-0.094	-0.186	-0.148
descr	**0.281**	-0.083	**0.179**	-0.142	0.096	**0.489**	**0.152**	1.000	**0.391**	-0.108	**0.295**	**0.257**	**0.281**	**0.453**	-0.005	**0.456**	**0.211**
guide	**0.493**	-0.039	-0.060	-0.112	-0.075	**0.456**	-0.159	**0.391**	1.000	**0.197**	0.328	**0.368**	**0.493**	**0.562**	-0.067	**0.759**	**0.220**
hist	**0.226**	-0.042	-0.064	0.073	-0.079	-0.089	-0.006	-0.108	**0.197**	1.000	0.028	0.129	0.056	0.089	-0.072	-0.060	**0.192**
novel	**0.579**	**0.268**	0.126	0.061	**0.157**	**0.308**	0.111	**0.295**	**0.328**	0.028	1.000	**0.475**	**0.422**	**0.444**	**0.333**	**0.277**	0.024
recent	**0.441**	-0.131	0.063	0.119	-0.030	0.129	0.099	**0.257**	**0.368**	0.129	**0.475**	1.000	**0.441**	**0.411**	-0.106	**0.207**	0.113
source	**0.338**	-0.088	**0.164**	0.029	-0.043	**0.277**	0.121	**0.281**	**0.493**	0.056	**0.422**	**0.441**	1.000	**0.337**	-0.016	**0.323**	0.099
struct	**0.427**	-0.073	0.050	-0.108	-0.005	**0.246**	0.134	**0.453**	**0.562**	0.089	**0.444**	**0.411**	**0.337**	1.000	**0.166**	**0.504**	0.096
time	-0.016	**0.158**	-0.084	-0.157	**0.299**	**0.034**	-0.094	-0.005	-0.067	-0.072	**0.333**	-0.106	-0.016	0.166	1.000	-0.079	0.072

	accur	advert	affect	auth	breadth	def	depth	descr	guide	hist	novel	recent	source	struct	time	tips	underst
tips	**0.323**	-0.046	-0.070	-0.131	-0.087	**0.406**	-0.186	**0.456**	**0.759**	-0.060	**0.277**	**0.207**	**0.323**	**0.504**	-0.079	1.000	0.060
underst	**0.182**	0.016	-0.175	-0.142	0.030	**0.220**	-0.148	**0.211**	**0.220**	0.192	0.024	0.113	0.099	0.096	0.072	0.060	1.000

n=59

Appendix W: Criteria and Level of Importance by Stage in Task Completion - Study 3

Criteria	Level Impt	Abstract	Detailed Outline	Rough Draft	Final Presentation
accuracy	2				3
	3	1			3
	4	3	19	7	22
	5	13	11	24	17
advertisement	1	1			2
	2	1			
	3	1			1
	4	1	1		1
	5				1
affectiveness	1	2	3	13	4
	2	1	2	2	1
	3	4	2	7	14
	4	8	2	2	11
	5	1		1	3
authority	1	1			
	2				2
	3	2	2	2	5
	4	4	2	3	13
	5	1	5	7	6
bias	1			1	1
	2	1			
	3	1	2	2	3
	4	2	2	1	8
	5	1			2
breadth	1	1			1
	2	4	1		2
	3	1	3	2	2
	4	8	5	8	13
	5	5	4	8	6
definitions	1				1
	2	4	1	2	
	3	5	3	2	4
	4	5	3	3	7
	5	2	3	1	10
depth	1	1			1
	2	2	1	1	2
	3	3	4	1	3
	4	9	12	11	21
	5	5	4	8	14
descriptions	1				1
	2		1	1	1
	3	1	1		3

Criteria	Level Impt	Abstract	Detailed Outline	Rough Draft	Final Presentation
	4	16	10	10	18
	5	8	3	2	13
guidelines	1				1
	2			2	1
	3	1	3	3	3
	4	6		3	9
	5	2			4
history	2	1	2		1
	3	3	2	2	3
	4	3	4	4	5
	5	4	5	1	4
novelty	1	3			1
	2	1			3
	3	2	3	2	6
	4	9	10	8	16
	5	10	11	14	12
recency	1				1
	2	3	1	1	12
	3	3		2	7
	4	11	7	6	16
	5	4	10	8	10
source	1				5
	2	1		1	1
	3	4	1	1	
	4	2	8	8	10
	5	7	5	3	8
structure	1	1		1	2
	2	1	1	4	3
	3	2	6	2	10
	4	10	9	6	10
	5	5	3		6
time	1		1	1	2
	2	2	2		2
	3		3	3	3
	4	1	1	1	6
	5	3	3		3
tips	1	3			2
	2			1	3
	3	3		1	2
	4	7		4	5
	5			2	2
topic	1				1
	2		1		4
	3	1	1	2	4
	4	20	27	19	37

Criteria	Level Impt	Abstract	Detailed Outline	Rough Draft	Final Presentation
	5	24	20	43	41
understandability	1			1	2
	2	2	2	2	5
	3	2	6	5	5
	4	9	11	6	20
	5	8	5	1	8

VITA

Arthur R. Taylor

Education

B. A. Finance - University of South Florida, 1981

Masters of Information Systems, George Mason University, 1990

Ph.D. in Information Science, Rutgers University, January 2009 (expected)

Summary of Experience

Detailed Work Experience

Rider University, Visiting Professor/Lecturer, September 2002 – Present

Sun Microsystems, October 2000 – November 2001 - Senior Course Instructor

Xtivia Solutions, February 2000 - October 2000 - Senior Consultant

Spectrum Technology Group (ClBER), February 1998 - October 2000 - Project Lead - Senior Consultant

Informix Software, October 1991 - January 1998 - Senior Consultant

NVR, 1989 - October 1991 - Project Managemer

The Orkand Corporation, 1985 - 1989 - Consultant

Prior Experience, 1983 - 1985

Series of jobs in the role of programmer analyst, systems support and systems operations.

Awards

Rider Summer Fellowship - 2005, 2006

Publishing

Books (Peer Reviewed)

Taylor, A., Buege, B., & Layman, R. (2003). Hacking Exposed: J2EE and Java – Developing Secure Applications with Java Technology. New York, NY: Prentice-Hall.

Taylor, A. (2003). J2EE and Beyond: Designing and Developing with Java. New York, NY: Prentice-Hall.

Taylor, A. (2003). JDBC: Database Programming with J2EE. New York, NY: Prentice-Hall.

Taylor, A. (2001). JSP and Java, the Complete Guide to Web Development. New York, NY: Prentice-Hall.

Taylor, A. (1999). Informix Power Reference. New York, NY: Prentice-Hall.

Taylor, A. (1997). JDBC Developers Resource. New York, NY: Prentice-Hall.

Taylor, A. & Lacy-Thompson, T. (1996). Informix NewEra: A Developer's Resource. New York, NY: Prentice-Hall.

Taylor, A. (1995). Advanced Informix-4GL Programming. New York, NY: Prentice-Hall.

Edited Book Chapters

Taylor, A. (2009). "Gen M and the Information Search Process" In The Myths, Realities, and Practicalities of Working with Gen M: A Handbook for Educators and Librarians. Eds. Vibiana Bowman and Robert J. Lackie. NY: Neal-Schuman, 2009.

Refereed Book Chapters

Taylor, A. & Eder, L. (2005). A comparison of authorization, authentication and auditing in Windows and Linux. In Enterprise Information Systems Assurance, Handbook of Research on Information Security, Ed. Merrill Warkentin. Idea Group Publishing, Hershey, USA.

Taylor, A. R.(in press). The Last Line of Defense: A Comparison of Windows and Linux Authentication and Authorization Features . In Jatinder Gupta and Sushil Sharma (Ed.) Enterprise Information & Security Authentication. Hershey, USA: Idea Group Publishing.

Refereed Journal Articles

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* (4)*,* 1071-1084.

Taylor, A. R., Zhang, X., Amadio, W. J. (revised and resubmitted). "Examination of relevance criteria choices and the information search process," *Journal of Documentation*.

Journal Articles Under Review or Resubmitted

Anderson, D. Christie, E. Kirkyla, A. Mardenfeld, S. & Taylor, A. (under review). "Choosing media based on political stance," *Rocky Mountain Communication Review*.

Taylor, A. R. (under review). The information search behavior of the Millennial generation. *Journal of the American Society for Information Science and Technology*.

Taylor, A. R. (under review). User relevance choices and the information search process. *Information Processing and Management*.

Published Trade Journal Articles

Taylor, A. (2002). United Front: Solving application integration problems with J2EE. *Intelligent Enterprise, 5* (6).

Taylor, A. (1999). Java's middle kingdom: Simplifying multi-tier application integration. *Intelligent Enterprise, 2* (12).

Taylor, A. (1999). Beyond JDBC: the New JDBC 2.0 API. *Intelligent Enterprise, 2* (6).

Taylor, A. (1998). Three-tiers for web-OLAP: Using middleware with online analytical processing. *Database Programming and Design, 11* (17).

Taylor, A. (1997). JDBC Sampler: the new JDBC API. *Database Programming and Design, 10* (12).

Taylor, A. (1997). Java-Charging the Web. *Database Programming and Design, 10* (10).

Taylor, A. (1996). Application Partitioning with Informix-NewEra. *Informix Tech Notes, 6,* (2).

Published Non-refereed Conference Papers

N.J. Belkin, M. Cole, J. Gwizdka, Y.-L. Li, J.-J. Liu, G. Muresan, D. Roussinov, C. L. Smith, A. Taylor, X.-J. Yuan (2005) *Rutgers Information Interaction Lab at TREC 2005: Trying HARD*, in Proceedings of TREC 2005, Gaithersburg, MD, November 2005.

Published Refereed Conference Papers

Anderson, D. Christie, E. Kirkyla, A. Mardenfeld, S. & Taylor, A. (2006). "Choosing Media Based on Political Stance," NJCA Spring 2006, Montclair, NJ.

Published Poster Presentations

Taylor, A. R. (2007). "Improving Searches of Digital Collections: Relevance Criteria Choices and the Information Search Process, Libraries in the Digital Age", Dubrovnik, Croatia, May 28- June 2, 2008

Reviewer

Enterprise Information Systems Assurance (2005), Hershey, PA, USA.

Professional Memberships

Association for Computing Machinery