Chapter 6

Study One: Exploration and IR

6.1 Abstract

Law firms are under constant pressure to reduce the billing charged to clients. Due to the fact that all information retrievals in eDiscovery must be reviewed by humans, there is a significant interest in reducing the number of non-relevant documents to be reviewed. Reducing this number will result in significant cost savings to the firm and their clients. This study addresses the objective of cost savings in human review by developing and evaluating a process for user exploration to reduce the search space. The goal is to improve the retrieval result developed from user exploration of a small representative sample of the full corpus. The desired effect is to reduce the search space such that fewer documents are needed for human review.

This study uses four behavioral measures to predict an individual's performance and three exploration measures to predict IR results. The data collected are recorded using an IT artifact developed for this study and evaluated for its utility based on the common performance measures of recall and precision.

The research conducted by this study provides insight into the relationship between recall and precision previously validated in the literature but never explained; how behavioral preferences impact IR performance; and how exploration variables measuring documents and time can be used to predict productivity in IR results.

6.2 Introduction

eDiscovery is an instance of document retrieval of a bounded collection/corpus. In such an instance the collection is domain specific, the search is ad hoc, and the typical user is highly educated in the domain – either through direct prior experience or emersion during an investigative process. (Bill Hamilton, November 2011). Study One: *Exploration* focuses on these two conditions. These conditions of eDiscovery IR go largely unexploited by users.

As previously stated, eDiscovery IR is highly context dependent, meaning that relevant search terms are often linked to subject matter knowledge or a controlled vocabulary of the corpus items. For example, if a user enters the word "the" as a search term, he/she will no doubt return 100% recall of documents but there will be almost zero (0) precision, resulting in the return of the entire corpus — not very useful.

A more useful result would be if a user could surgically identify terms with the goal of reducing the search space that produces a retrieval set containing a high percentage of relevant documents and a low percentage of non-relevant documents. In this study we investigate whether exploration is an effective method to learn context in order to achieve this goal.

This study addresses the gap between brute force, trial and error techniques, and test collection reviews presently employed by eDiscovery practitioners, by designing an artifact to support user exploration of a bounded corpus. The study seeks to explain how users determine their IR strategies and how exploration can be used most effectively to improve IR performance.

In 1985, Blair and Maron conducted a series of experiments designed to address the problem of finding relevant documents from a collection. The collection contained 40,000 objects. They found that on average, recall was in the 20% range – quite unacceptable. Twenty-

five years later, instances of low recall are still common in large corpus IR (*TREC Proceedings 2010*). One explanation for this phenomenon lies in the limitation of a user to predict the matching of terms to relevant documents – recall terms producing *relevant* documents, and also not producing *non-relevant* documents (Blair and Maron, 1985). This study addresses this limitation by developing an artifact to support exploration as a method to better predict matching of terms to relevant documents. One of the objectives of this study is to provide insight into how users develop their search strategies (Bates, 1979). We begin our research in this domain by researching the phenomenon of exploration.

Exploration is a natural and intuitive method to use when probing a large collection of documents in an attempt to reduce the scope of the search space. We see common and frequent examples every day when a person searches the web for information on a subject matter or topic. In such instances the user chooses terms, and sometimes operators, as an initial predictive approximation for the information being requested, and then adjusts the query criteria as results appear. This approach makes conventional sense when conducting a search of scale free collections with no preconceived definition of document(s) satisfying the information need, and where the information need is the presence or the absence of a document containing the information requested rather than a specific answer to a specific factual question.

6.2.1 Exploration

The concept of exploration has been associated with learning (Berlyne, 1963; March, 1991); familiarization (Barnett, 1963), and information search (Debowski et al., 2001). In fact work done by Berlyne in the 1960s classifies exploration as a “fundamental human activity” (Demangeot and Broderick, 2010).

Exploration is seen as a behavior motivated by curiosity. Exploration that is goal directed is classified as extrinsic (Berlyne, 1960). Extrinsic exploration typically has a specific task purpose, whereas intrinsic exploration is motivated by learning (Berlyne, 1960; Demangeot and Broderick, 2010). (Kaplan and Kaplan, 1982) argue that exploration arises from our need to make sense of our environment. (March, 1991) writes about exploration and exploitation. He views exploration and exploitation as competing tensions in organizational learning.

(Berlyne, 1963) suggests that specific exploration is a means of satisfying curiosity. The goals of exploration as a means for making sense of our environment and satisfying curiosity are represented in the problem domain of information retrieval and eDiscovery. (Debowski et al., 2001) view exploratory search as a "screening process," and state that exploration identifies items "to become the focus of attention." They suggest that exploration leads to learning through the examining and scrutinizing of items.

The first artifact developed for this study is designed to support the user in exploring a corpus of items and facilitating examining and scrutinizing, so that the user may obtain contextual knowledge about the collection.

6.2.2 Prior Exploration/Search Research

The study of exploration and search behavior and the mental models users create to execute them is not entirely new. In the early 2000s much research focused upon strategies that users formulate for web searches. (Holschler and Strube, 2000) examined the types of knowledge and strategies involved in web-based information seeking. They found that users with higher levels of knowledge were more flexible in their approach and were better able to tackle

search problems than those who were less knowledgeable. They characterize the information space as "diverse and often poorly organized content." This contrasts with the bounded space of eDiscovery which is typically organized around the subject matter in question. Their finding that experts can outperform less experienced users will be extended by this study by evaluating whether knowledge acquired by exploration can improve a user's ability to tackle the search problem of eDiscovery.

(Muramatsu and Pratt, 2001) evaluated a system designed to provide users with "light weight feedback" about their queries. They found that transparency is "helpful and valuable." Their conclusion was that interfaces "allowing direct control of query transformation may be helpful to users." The exploration study and learning study in this dissertation extend their work by designing two separate artifact tools to provide just such control. We evaluate the efficacy of both artifacts in chapter 6 and chapter 8 of this dissertation.

Other research has focused upon browsing behavior and categorizing search behaviors into types. Bates (1989) coined the phrase "berry-picking" to refer to individuals' search strategy being in constant evolution. A study done by Catledge and Pitkow at Georgia Institute of Technology captured client-side user events to study browsing and search behavior (Catledge and Pitkow, 1995). Their study evaluated frequency and depth and found support for three different types of searcher characterizations based on Cove and Walsh's original work in 1988: *Serendipitous browser*, *General purpose browser*, and *Searcher* (Cove and Walsh, 1988).

Broder (2002) proposed a taxonomy of web search to include *transactional*—a web mediated activity, *navigational*—seeking a specific site, and *informational*—a page containing a

particular need. This dissertation studies the user's informational need, and also seeks to explain his/her navigational behavior that may affect the IR result produced.

(Muylle et al., 1999) undertook a study to better understand web search behaviors and motivations of consumers and business people. The study found three constructs describing search behavior: (1) *exploratory* – title scanning, (2) *window* – document scanning, and (3) *evolved* – document scrutinizing. The research conducted in this dissertation adapts these constructs to measure scanning, skimming and scrutinizing behavior in the users conducting exploration of the corpus.

(Navarro-Prieto et al., 1999) studied how people search for information and focused on the "cognitive strategies" followed by the user. Not surprisingly, they found three prevailing strategies: (1) *Top-down*—broad based followed by narrowing down, (2) *Bottom-up*—specific terms for specific fact finding, and (3) *Mixed*—employing both strategies in parallel. Also not surprising, they found that experience mattered. The users who were more experienced developed more complex rules for their searches and followed a top-down approach.

The research and findings mentioned above are consistent with the other research found in this domain: (1) that experience matters, and (2) that experience affects the complexity of search strategy and choice of search terms.

This dissertation seeks to extend the research done on search by evaluating how users can improve their knowledge through exploration, and leverage that knowledge through an automated tool to improve IR results.

6.3 Motivation

eDiscovery extraction is modeled differently than an open ended IR search such as "scanning the web for general information on a topic," or a prior art search for say, a patent. What makes eDiscovery unique is the manner in which the user frames the universe to be searched; the corpus/collection is bounded — it is defined in a way that those who understand the context of the documents to be sought tend to produce better IR results. The reason for this is that the user in eDiscovery IR is a high compensated individual, an educated professional or team, highly focused on the topic of interest. The topic arises out of a specific series of transactions or related events that are defined by time, population, location, and other ad hoc circumstances making the IR corpus bounded in a particular way that the IR result (relevance) is highly dependent on content and context. This unique set of IR circumstances leads to our main research question in this chapter of *whether a user can acquire knowledge of context and content through exploration of the bounded corpus*. The assumption is that more knowledge on a topic will produce better IR results than less knowledge on a topic.

Why is this question important? The elements that make IR of a bounded corpus unique also make it an important phenomenon to study. Consider the fact that bounded collections represent the recorded actions of parties to everyday transactions. As our society becomes more and more dependent on digital storage of recorded transactions the ability to effectively extract relevant documents from large collections of similar items will continue to be a value proposition in terms of time and cost. Whether it is a consumer and merchant, a dispute between commercial actors, or in the case of Ms Zubalake, an employee against her employer, bounded collections are becoming increasingly more frequent in information retrieval.

Learning and Relevance Feedback in Information Retrieval: A Study in the Application of User Knowledge to Enhance Performance

By: Harvey Hyman, PhD

Most bounded searches in eDiscovery follow a standard pattern of: interviewing personnel expert in the domain, ascertaining the standards of storage and organization, developing potential search terms, and producing test reports indicating the frequency of terms within documents (occurrences). "Hit reports" are reviewed and terms are refined based on the frequencies observed. Quite often, this is done prior to reviewing any document extractions. Patterns are determined and more test runs are commenced; only then are documents reviewed from the extracted sets. As one interviewee describes, IR users engage in a "brute force" trial and error approach to arrive at a focal point of key terms to use for extraction (Aaron Laliberte, November 2011).

Search terms are proxies for relevance descriptors of a document (Salton and Buckley, 1988). Weighting of terms enhances the effectiveness of document description by using statistical occurrences and term frequencies. This technique is fundamental to indexing methods (Luhn, 1957; Spark-Jones, 1971). A major limitation associated with term frequency is the difficulty of distinguishing between the frequency of occurrence in the relevant documents and the frequency of occurrence in the entire collection. There is an assumption here that search terms may be known *a priori* or may surface as a result of patterns discovered during exploration of the collection. One hypothesis in this study is that *exploration of the collection will provide the eDiscovery user with the ability to describe the document he/she is seeking, and therefore select better search terms*. A second hypothesis here is that *exploration of the corpus will provide a greater understanding about the nature of the documents (relevant and not), and lead to better decisions for selection of search terms, resulting in improved recall and precision*.

6.3.1 Background

eDiscovery is a domain where the nature of the IR is highly user dependent and highly context oriented (Oard et al., 2010; Baron, 2005; Grossman and Cormack, 2011). This nature exploits the weaknesses of term based search alone. Term based search is well suited when a query is narrow in focus and particularity. eDiscovery users have found that term search alone is inadequate when context is important, resulting in (over-inclusion) precision loss, or (under-inclusion) recall loss, (Paul and Baron 2007; *Sendona Conference,* 2007; Oard et al., 2010). This study evaluates whether a user can learn the context of a collection and make better decisions about the selection of search terms to improve the IR result. The prototype developed for this study allows the user to select the "level of importance" of a search term for the method of weighting.

6.4 Methods

The method used in this study is a controlled experiment. The purpose of the experiment is to measure the affect upon IR performance of user exploration of a small sample of a large corpus. Performance is measured by the dependent variables *Recall* and *Precision* as previously defined. There are two sets of explanatory variables used. The first set is comprised of behavioral scales known to be associated with preferences that are predictive in the use of technology and innovativeness. The second set is comprised of operational measures to represent the constructs of scanning and scrutinizing behaviors associated with exploration of digital collections.

The document sample consists of 300 randomly selected documents from the overall collection of 680,000 objects. The task, treatment and data collection are conducted via the prototype developed for this study. The prototype application built for this experiment is housed

on a server and accessed by the participants using a URL link from their self provided laptop computers. The computer screens from the application are displayed in Appendix - E.

Participants are assigned an eDiscovery task. Informed consent, task instruction and data collection instrument are displayed as computer screens – graphically depicted in Appendices – B, C and D.

All participants are given the same task. The task is to provide recall (search) terms and elimination terms (filters) in response to an eDiscovery request. The task has been adapted from the TREC Legal Track 2011 Conference Problem Set #401. A list of the exploration independent and dependent variables and covariates are displayed in Table 3.

Table 3: List of Variables Tracked in Exploration Study

Independent Variables	Dependent Variables
Exploration/Artifact Variables:	**Performance Measures:**
Number of documents viewed	Recall
Total Viewing Time	Precision
Viewing time per document	
Order of documents viewed*	
Covariates:	
Age, Gender, Education Level	
Legal, Discovery, eDiscovery experience	
Experience with ENRON collection	
Experience with financial terms, concepts and transactions.	

***-** Collected for Future Study

The independent exploration variables tracked in this study are: Total Number of Documents Viewed, Total Amount of Viewing Time, and Time Spent per Document. Linear regression analysis is used to measure the following relationships: (1) Correlations of independent variables with dependent variable *Recall*, (2) Correlations of independent variables

with dependent variable *Precision*, (3) Interactive effect of independent variables upon the dependent variables, and (4) Significance of covariates.

Demographic information has been collected to track co-variables. Specific *a priori* co-variables are legal experience, discovery experience, eDiscovery experience, familiarity with the Enron data set, experience in financial transactions, financial concepts and financial terminology. Age, gender, and educational level have also been collected. The above variables have been identified by our panel of experts and surfaced during our pilots for this study. Not all co-variables appeared during our data collection. The co-variables that have been identified as most relevant in this study are: Litigation experience scaled as 0 – 3 to represent no experience, less than 1 year, 1 to 2 years, and greater than 2 years; Knowledge of subject matter (Financial Terminology) scaled as 0 – 3 to represent no knowledge, some familiarity with terms, amateur investing, and professional training or experience.

6.4.1 Dataset

The dataset in this study is a corpus of electronic documents known as the Enron Collection, Version 2. The full corpus of this version contains approximately 650,000 to 680,000 email objects depending on the counting of attachments. This data set has been previously validated in the literature (TREC Legal Track Proceedings 2010, 2011).

The subset of data we use for the exploration artifact is a collection of 10,000 randomly selected documents from the full corpus. 1,000 of the documents have been selected from the validated set marked relevant, and 9,000 documents have been selected from the validated set marked not relevant. This allows us to make certain assumptions. The first assumption is that a random extraction from the subset should yield a recall of .10. Any level of recall above this

number indicates an improvement over chance – a result better than no human input at all. The second assumption is that the set of documents retrieved by the user selections can be measured for precision based on a common base line for relevance from the validated relevant documents.

6.4.1.1 Participants

The participants in this study are 60 third year law students from three different United States, East Coast law schools. Law students have been chosen for this study because they are familiar with fundamental principles of legal procedure and discovery process and techniques. **Participants have been given an economic incentive** to motivate them to try their best. The participants were told that the best performer in each group would be awarded a cash prize.

6.4.2 Process

An IT artifact developed for this dissertation is a job run procedure used to pre-process the data from the Enron collection. The objects in the collection which consist of emails and attachments need to be prepared in such a way that the text can be read by the engines of the system. Considerable time went into this phase. Some problems encountered with files included: password protected files, power point files that did not translate well with the chosen OCR tool, emails with URL links no longer valid, emails with attachments only and no text in the body, and emails with files pasted into the body instead of native text. The job run process is depicted in Figure 7.

The participants interact with the user-interface screens made available through an URL link to the server from their personal laptops. They are instructed by random assignment to select Group 1, Group 2, Group 3, from a list of radio buttons on the computer screen (depicted in the Appendix - E). The radio button chosen corresponds to amount of exploration time allowed.

Group 1 receives up to 15 minutes, Group 2 receives up to 30 minutes, and Group 3 receives 45 minutes to explore the 300 document sample. **The time allotments are maximums**, meaning each participant may terminate their individual session at any point during the study. For example, a participant in Group 3 may choose to terminate his/her exploration after 10 minutes. The actual application has 4 group buttons. The additional button is used for testing of the system. This allows us to segregate our system tests from the participant data.

The participants may conclude their exploration at any time by selecting the next button on the screen. The exploration behaviors are logged by the system as sessions, and tracked as the independent variables (IVs) Total Time, Time per Document, Number of Documents.

All three groups receive the same task. The purpose of using three groups is to spread out the time line. We found during the pilots that if participants are not given an anchor time they all cluster too close together and create a narrow variance in time measure. Therefore, the study uses artificial groups to spread out the time line thereby avoiding a tight cluster and increase the explanatory power of the variables.

The participants supply their recall terms and elimination terms through the interface screen. The user selections are logged per user and submitted to the job queue for processing as depicted in Figure 7.

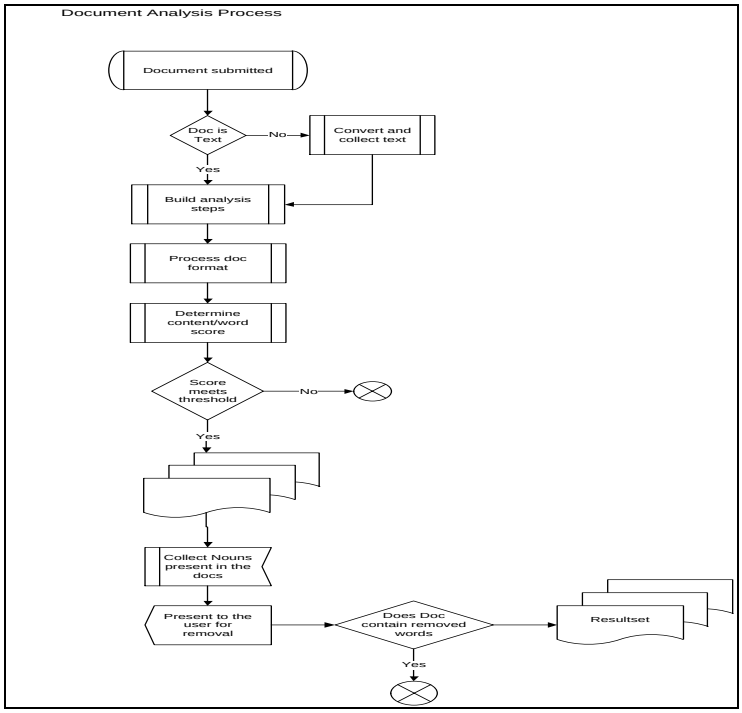

Figure 7: Job Run Process

6.5 Initial Pilot Studies

The first pilot conducted in this study consisted of 7 lay personnel. The purpose of the pilot was to receive feedback regarding presentation, clarity and ease of use. The current version of the system reflects feedback received from the pilot.

The second pilot was conducted with 10, second year law students. Five (5) students were placed in a control group which was given no time to explore, and 5 were given up to 60 minutes to explore the sample collection. The average time exploring the collection clustered around 43 minutes, with a single high of 60 minutes and a single low of 23 minutes. The average number of documents reviewed was 70, with a single high of 90 and a single low of 15. The average time per document was 45 seconds, with a single high of 2 minutes and a single low of 10 seconds.

Data was inconclusive on the issue of significant difference in recall or precision between the two groups. The reason for this probably has to do with the small number in the groups in order to detect a difference from zero. The small number of participants also makes it difficult to draw conclusions about how: *total time viewed, number of documents viewed* and *per document time,* may be predictive of *recall* and *precision* due to the concentrated clustering of the *total time explored*.

The most useful information provided from the second pilot was in the form of user feedback. The participants provided feedback consistent with the previous pilot. This increased our confidence in the design of presentation and environment for the full experiment. The main limitation in the second pilot indicted above is the small sample group making it difficult to draw conclusions about relationships of the variables. However, the pilot has confirmed the design, ease of use and quality issues of the system being tested.

6.6 Design of Full Study

Sixty (60) third year law students are the subjects of this study. They have been randomly assigned to three groups to spread out time performance. The individuals within each of the groups are allotted maximum time allowances to complete their exploration. The participants may terminate their exploration at any time. For example, an individual who is assigned to Group 4 is given up to 45 minutes to explore the corpus however the participant may choose to terminate the exploration at the 10 minute mark; there is no forced time range or minimum amount for the participants, just a maximum allowance depending on the Group assigned.

The participants are administered the behavioral questionnaires at the beginning of the study so that their responses are not affected by the task. The behavioral questionnaires are

designed to collect data on the four scales measuring user IR behavioral attitudes: Tolerance for Ambiguity (TOA), Locus of Control (LOC), Disposition Toward Innovation (DISPO), and Personal Innovation Toward Information Technology (PIIT). Three subjects from each group have been selected for verbal protocols and are encouraged to "think out loud." Post-task interviews are conducted with three (3) additional subjects from each group. The purpose for choosing subjects from each group is to select users from different levels of time exploration.

All subjects are administered the post-task questionnaire and usability study at the end, followed by a hearty thank you for their time and good-bye. The total time for participation ranged from 45 minutes to 110 minutes. The participants' sessions have been recorded by a server hosting the artifact/application.

Independent variables (IVs) representing *Total Time Exploring*, *Total Documents Viewed*, and *Time Spent Per Document* have been assigned to track user interaction with the artifact. A model that depicts the artifact IVs and their relationship to dependent variables (DVs) *Recall* and *Precision* is illustrated in Figure 8. The model for exploration artifact assumes an input, an output, and a process in the middle. The three IVs represent *input*, the two DVs represent *output*, and the exploration construct is in the middle representing the human cognitive *process*.

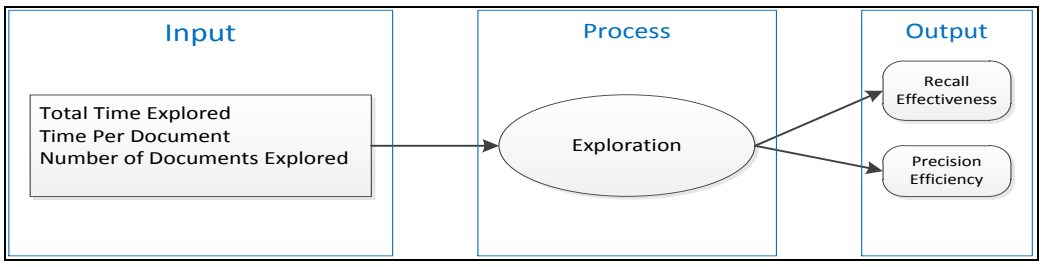

Figure 8: Model of Artifact IVs

Independent variables representing *tolerance for ambiguity (TOA), locus of control (LOC), dispositional innovativeness (DISPO),* and *personal innovativeness toward information technology* (PIIT) have been assigned to track user behavioral factors associated with information retrieval technology and innovation. This study focuses on the portion of the Information Retrieval Behavior Model from Vandenbosch and Huff in Figure 3.2, representing the impact of behavioral measures upon exploration and their relationship to the dependent variables (DVs) *Recall* and *Precision*. The adapted model is depicted below in Figure 9.

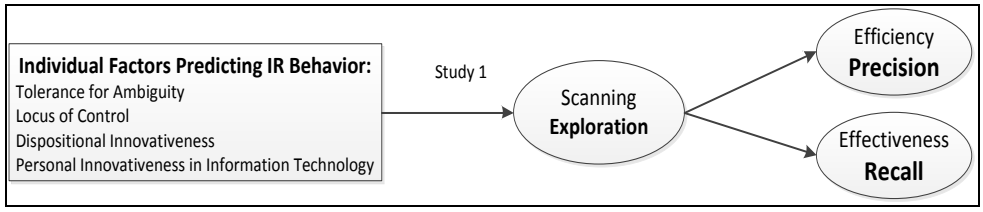

Figure 9: Information Retrieval Behavior Model for Study One

Covariates assigned in this study control for *litigation experience, familiarity with the corpus* and *knowledge of financial terms* (subject matter experience) associated with the task. Levels have been assigned to correspond to years of litigation experience, depth of financial knowledge, and exposure to the corpus. A table listing the co-variates, assigned levels, and descriptions is displayed in the Appendix - J.

6.6.1 Exploration Artifact Variables

Information seeking can be divided into broad exploration and precise specificity (Heinstrom, 2005). Broad exploration is a possible indicator of a wide overview strategy and knowledge building, whereas precise information seeking may be an indicator of a focused,

pinpointed search (Heinstrom, 2005). In the case of precision search, the user has a specific frame of reference from which to investigate and probe a collection.

Research has found that a "common approach" to large collection search is for the user to begin with "an already known term" (Lehman et al., 2010). The use of the known term typically leads to an item that informs the user with additional terms to improve the search for the next iteration. When more than one item is returned the user has the option of reviewing each item one at a time. But when a large volume of items is contained in the retrieval set, the user must apply some method to select items for further inspection from among the set. (Lehman et al., 2010) developed a visualization method for user exploration of large document collections. The visualization approach was employed by them to study user information seeking in Wikipedia. The results of their study found that, "visual navigation can be easily used and understood" (Lehman et al., 2010).

Browsing as an information seeking process has been established as a method when the information need is ill-defined (Kuhlthau, 1991; McKay et al., 2004). Browsing has been described as a fundamental information seeking function (Bates, 1979, 1989; Kuhlthau, 1991; McKay et al., 2004). Exploration is an underlying construct representing the human search behavior (Holschler and Strube, 2000; Muylle et al., 1999); it is operationalized in electronic search as browsing. This study operationalizes exploration by use of an artifact built as an interactive tool to support user exploration of a corpus by exploitation of selected items in order to learn context and content.

When a user finds multiple documents they will tend to switch back and forth between items; this activity describes the iterative approach to information seeking (McKay et al., 2004).

(Meuess et al., 2005) developed an XML retrieval system combining structure with text references. (McKay et al., 2004) developed three approaches to browsing using the Greenstone digital library database. Ignat et al., 2006 developed an automated tool designed to support exploration of large document collections by use of clustering; it is implemented using a standard web browser. (Chowdhury et al., 2011) focused on uncertainty as an underlying construct in Human Information Behavior (HIB).

The above referenced research has focused on investigating and describing users' information seeking behavior through exploration and browsing activities. The research in this dissertation is focused upon benchmarking user productivity in the search process.

In this study we have selected variables to measure user productivity in the exploration information seeking process. We have chosen *Total Time Explored* and *Time Spent Per Document* to measure the effort expended by exploration and to account for the exploration/exploitation trade-off (Holschler and Strube, 2000; Muylle et al., 1999; Hills, 2010; March, 1991). We have chosen *Total Number of Documents* explored to account for the iterative nature of information seeking (Bates, 1989; McKay, 2004).

The general proposition for this study is that an exploration method will outperform both random extraction and verbatim/non-function word extraction. The hypotheses representing this proposition are as follows:

H1a Random: Exploration outperforms random extraction measured in units of recall.

H1b Random: Exploration outperforms random extraction measured in units of precision.

H2a Verbatim: Exploration outperforms verbatim extraction measured in units of recall.

H2b Verbatim: Exploration outperforms verbatim extraction measured in units of precision.

The hypotheses for the exploration variables are as follows:

H1a: *Recall* is directly and positively correlated with *Total time exploring a corpus*.

H1b: *Precision* is directly and positively correlated with *Total time exploring a corpus*.

H2a: *Recall* is directly and positively correlated with the *Number of documents viewed in a corpus*.

H2b: *Precision* is directly and positively correlated with the *Number of documents viewed in a corpus*.

H3a: *Recall* is directly and positively correlated with *Time spent per document*.

.

H3b: *Precision* is directly and positively correlated with *Time spent per document*.

We did not have any prior theory about whether some of the variables might interact to produce effects upon recall and precision. Therefore, we used a null and alternative hypothesis for each:

H_0: *Total time exploring a corpus* affects *Recall* and *Precision* independent of *Number of documents viewed* and *Time per document*.

H_a: *Total time exploring a corpus* affects *Recall* and *Precision* depending on *Number of documents viewed* or *Time per document*.

H_0: *Number of documents viewed* affects *Recall* and *Precision* independent of *Total time exploring a corpus* and *Time per document*.

H_a: *Number of documents viewed* affects *Recall* and *Precision* depending on *Total time exploring a corpus* or *Time per document*.

H_0: *Time per document* affects *Recall* and *Precision* independent of *Total time exploring a corpus* and *Number of documents viewed*.

H_a: *Time per document* affects *Recall* and *Precision* depending on *Total time exploring a corpus* or *Number of documents viewed*.

6.6.2 Behavior Scales

The experiment in this chapter seeks to explain factors impacting IR results and uses an innovative tool to do so. Four behavioral scales have been chosen to measure preferences known to be associated with information retrieval and innovation. The goal is to determine which scales are significant in ability to predict IR performance of individuals, measured by the variables *Recall* and *Precision*. Personality traits have been associated with information seeking patterns and differences in search approaches and strategies (Heinstrom, 2005). The four behavioral scales used in this study are listed in Table 4. They are further explained in the next sections.

Table 4: List of Behavior Scales

Variable	Name	Description	Number of Items	Cronbach's Alpha
TOA	Tolerance for Ambiguity	The degree to which an individual is willing to accept ambiguity is "related to an individual's desire to create uncertainty and tend toward scanning behavior because they are not fearful of the ambiguity that often results." (Vandenbosch and Huff, 1997)	8	.80
LOC	Locus of Control	A person who has a higher LOC believes he/she has greater control over what happens to them rather than external factors. This individual is more likely to explore broadly due to greater confidence to produce results.	5	.85
DISPO	Dispositional Innovativeness	The measure of an individual's likeliness to try a new product, or think tangentially when solving a problem.	8	.85
PIIT	Personal Innovativeness in the Domain of Information Technology	The degree to which an individual has a preference for technology use.	4	.97

Tolerance for Ambiguity

Tolerance for Ambiguity (TOA) has been found to be associated with uncertainty in tasks intended to replace ambiguity with order (Vandenbosch and Huff, 1997; Rydell and Rosen, 1966; McCasky, 1976). The hypotheses are illustrated in Figure 10, below and in written form as follows;

H4a: *TOA is positively related to Recall.*

H4b: *TOA is negatively related to Precision*.

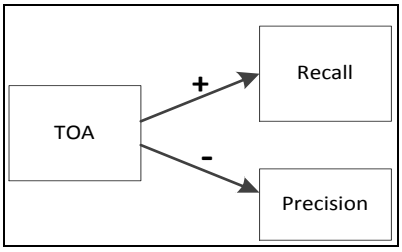

Figure 10: TOA effect upon Recall and Precision

Given that we know from previous studies that recall and precision are inversely related (Oard et al., 2010; Grossman and Cormack, 2011), we believe in this study that individuals seeking less ambiguity will prefer greater precision, whereas individuals willing to accept more ambiguity will prefer greater recall. The person more comfortable with ambiguity is more likely to seek broader exploration because he/she is not concerned with the additional non-relevant documents that may result. This is especially applicable to eDiscovey where lawyers often go on "fishing expeditions" as mentioned by Oard et al., 2010. The pre-task questionnaire designed to measure this construct has been adapted from the Rydell-Rosen Scale (1966). The original form contained 20 items which proved too unwieldy for our subjects. A confirmatory factor analysis was used to reduce the number of items. The final form contains 8 items and produced a Cronbach alpha of .80.

Locus of Control (LOC) is a measure of the degree to which individuals believe they control their own fate (Levenson, 1974). The LOC inventory developed by Levenson measures three factors: (1) Internal, the extent to which the person believes he or she is in control; (2) External, the extent to which a person believes his or her fate is controlled by others; (3) Chance, the extent to which the person believes their fate is determined by chance events.

Prior MIS research has found that individuals who believe they control their own fate are more likely to engage in scanning techniques for their IR (Vandenbosch and Huff, 1997; Levenson, 1974). Prior analysis of the Levenson three factor scale has shown it to be more reliable than similar scales measuring only two factors (Presson et al., 1997). For these reasons the Levenson three factor scale has been adapted for use in this study. The original form had 24 items. A confirmatory factor analysis was used to reduce the number of items to 5 with a Cronbach alpha of .85.

The proposition in this study is that scanning should be expected to be associated with broader search exploration and therefore, would favor recall over precision. The rationale is that individuals who believe they are in control of their performance results, rather than chance or others being in control, are more likely to conduct broader searches, leading to greater relevant documents returned. Broader searches are associated with return of greater non-relevant documents. We therefore believe that individuals with a higher preference on the LOC scale will explore with greater confidence, search broader, and produce higher recall, but lower precision. The hypotheses are illustrated in Figure 11, and presented in written form as follows;

H5a: *LOC is positively related to Recall*.

H5b: *LOC is negatively related to Precision*.

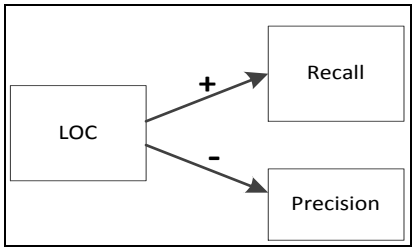

Figure 11: LOC effect upon Recall and Precision

Innovativeness can be described in several ways. It has been used in consumer research to predict an individual's predisposition to purchase new products (Roehrich, 2004; Steenkamp and Gielens, 2003). It has been shown to predict an individual's willingness to try a new technology (Agarwal and Prasad, 1998). It has been used to explain an individual's tendency to engage in thinking exercises such as puzzle solving and pondering (Pearson, 1970). When describing "cognitive innovation" Pearson describes the concept as "thinking for its own sake" (Venkatraman and Price, 1990, citing Pearson, 1970).

In this dissertation we are interested in how an individual's exploration attitudes and techniques can be explained through known and validated measures. In this case we have settled on two scales for measuring innovativeness. The first scale is designed to measure a user's disposition toward innovativeness. The second scale is designed to measure a user's personal innovativeness.

"Dispositional Innovativeness" (DISPO) has been shown to be significant in predicting consumers who are more likely to try a new product (Steenkamp and Gielens, 2003). In this dissertation participants are being asked to use a new method for eDiscovery IR. One of the hypotheses of this dissertation is that participants measuring higher on the scale of dispositional innovativeness will produce a higher IR result. The administered questionnaire contains eight (8)

items measured on a 1 to 5 scored scale, ranging from completely disagree = 1 to completely agree = 5. Cronbach alpha for this inventory is .85.

The proposition here is that individuals with a higher level of dispositional innovativeness are more likely to embrace a new system resulting in greater IR results. It is likely that such individuals are broader thinking and are willing to randomly jump around in their exploration due to their preference for the new and novel. These types of individuals are more tangential in their thinking and approach problem solving from unconventional points of view (Kirton, 1976; Vandenbosch and Huff, 1997). The hypotheses derived from the proposition are depicted in Figure 12 and in written form as follows:

H6a: *DISPO is positively related to Recall.*

H6b: *DISPO is negatively related to Precision.*

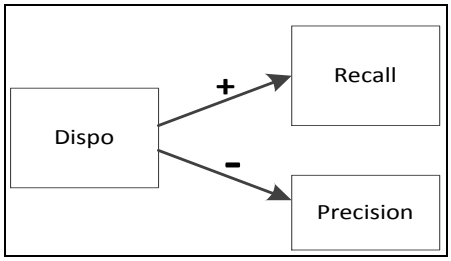

Figure 12: DISPO effect upon Recall and Precision

These hypotheses are measured using two different methods. The first method analyzes whether DISPO is significant and if the relationship is in fact positively correlated with *Recall* and negatively correlated with *Precision*. The second method utilizes a post-task usability study.

This study asks the users to rate the system on how well it helped them perform the task and how likely they are to use this system in a real life eDiscovery scenario. The results from the usability study are discussed in Chapter 9.

"Personal innovativeness in the domain of information technology" (PIIT) is associated with early adopters and individuals who are more comfortable with uncertainty (Agarwal and Prasad, 1998 citing Rodgers, 1995). Given that the eDiscovery user specifically operates in the domain of uncertainty, a measure of a user's PIIT may be helpful in predicting the same user's exploration preferences and resulting IR performance. The questionnaire contains 4 items and produced a Cronbach alpha of .97.

Agarwal and Prasad argue that individuals with higher PIIT levels are more likely to have positive attitudes toward an innovative technology. These attitudes translate to our experiment in terms of higher values in *Precision*. We believe that individuals with a preference toward technology will be more surgical in their exploratory behavior and produce higher precision. Given the documented inverse relationship between recall and precision, we believe the higher performance in *Precision* will result in a lower performance in *Recall*. The hypotheses are depicted in Figure 13 and in written form below:

H7a: *PIIT is negatively related to Recall.*

H7b: *PIIT is positively related to Precision.*

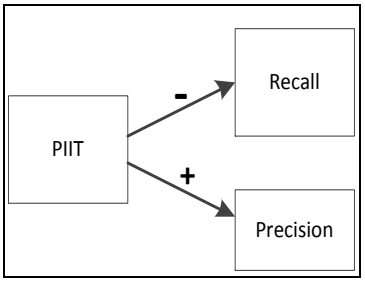

Figure 13: PIIT effect upon Recall and Precision

6.7 Data Analysis

SAS 9.2 was the statistical package chosen to support the analysis in this study. Collected data has been analyzed in several steps. The method of analysis in this case is a multiple linear regression. We are analyzing whether the independent (explanatory) variables are significant and whether interactive effects are present. We are also concerned with controlling for the listed covariates. A global F-test was used to evaluate the overall model and partial F-tests were used for testing interactive effects.

The behavioral scales have been analyzed using Cronbach's alpha. Two of the behavioral scales were extremely long (TOA and LOC); the original version of TOA had 20 items and the original version of LOC had 24 items. In order to reduce these scales to a manageable number of items for participants, a factor analysis was conducted for each scale. The scales were reduced to 8 items and 5 items respectively. Confirmatory Factor Analysis was used with varimax rotation. Cronbach alphas were calculated for the scales and are listed in Table - 4.

The first step was to transfer the pen and paper questionnaires to a spreadsheet for input into SAS. These questionnaires covered the four scales of TOA, LOC, DISPO, and PIIT. These behavioral scales were then analyzed to determine significance in a main effects and full model.

The models reflect the underlying theories represented by the hypotheses being tested. The initial theory of the behavioral scales is that individuals' IR performance can be predicted from their scores on the behavioral scales. The theory is represented by the hypotheses in the previous section and reduced to equations forming the behavioral models indicated below.

Main Effects Model: DV_{Recall}, $DV_{Precision} = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + e$

Full Model: DV_{Recall}, $DV_{Precision} = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + B_5X_1X_2 + B_6X_1X_3 + B_7X_1X_4 + B_8X_2X_3 + B_9X_2X_4 + B_{10}X_3X_4 + CV_1 + CV_2 + CV_3 + e$

Where:

X_1 = TOA,

X_2 = LOC,

X_3 = DISPO,

X_4 = PIIT,

CV_1 = Litigation Experience,

CV_2 = Enron Set Familiarity,

CV_3 = Subject Matter Familiarity (Financial Knowledge).

After the behavioral models were analyzed, we then conducted analysis upon the exploration models comprised of the independent variables *Total Time Explored* (TTE), *Time per Document* (PER), and *Number Of Documents Explored* (NUM). The initial theory is that individuals' IR performance can be predicted based on their exploration behavior measured by

the independent variables. The theory is represented by the hypotheses in the previous section and reduced to the equations for the exploration models indicated below:

Main Effects Model: DV_{Recall}, $DV_{Precision}$ = $B_0 + B_0 + B_1X_1 + B_2X_2 + B_3X_3 + e$

Full Model: DV_{Recall}, $DV_{Precision}$ = $B_0 + B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_1X_2 + B_5X_1X_3 + B_6X_2X_3 + B_7CV_1 + B_8CV_2 + B_9CV_3 + e$

Where:

X_1 = Total Time Explored,

X_2 = Time per Document,

X_3 = Number of Documents Explored,

CV_1 = Litigation Experience,

CV_2 = Enron Set Familiarity,

CV_3 = Subject Matter Familiarity (Financial Knowledge).

In addition to analyzing the behavioral models and the exploration models we also investigated relationships between the behavioral variables and the exploration variables. To test for this we set up an equation with the behavioral scales as independent variables and the exploration variables as dependent.

We had no prior theory about whether the behavioral variables would affect the exploration variables. Therefore, we used null and alternative hypotheses to test whether a significant relationship exists along with the linear equations for this model indicated below. In this case the null hypotheses represent the results that there is no significant difference from zero

between the independent variables and the dependent variables. The alternative hypotheses represent the results that at least one of the independent variables is significantly different from zero.

Main Effects Model: $DV_1, DV_2, DV_3 = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + e$

H_0: $B_1 = B_2 = B_3 = B_4 = 0$

Ha: At least one Beta $\neq 0$

Full Model: $DV_1, DV_2, DV_3 = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + B_5X_1X_2 + B_6X_1X_3 + B_7X_1X_4 + B_8X_2X_3 + B_9X_2X_4 + B_{10}X_3X_4 + e$

H_0: $B_1 = B_2 = B_3 = B_4 = B_5 = B_6 = B_7 = B_8 = B_9 = B_{10} = 0$

Ha: At least one Beta $\neq 0$

Where:

DV_1 = Total Time Explored,

DV_2 = Time per Document,

DV_3 = Number of Documents Explored,

X_1 = TOA,

X_2 = LOC,

X_3 = DISPO,

X_4 = PIIT.

The covariates were coded on a continuous scale based on levels of experience as described previously. The plan was to analyze the covariates for significance. When we collected the data it turned out that none of the 60 students had experience with the data set nor did any of them have financial experience, so those two co-variables dropped out of the equation. Also, only three students answered that they had litigation experience and none had eDiscovery experience. The analysis of CV for litigation experience indicated that there was no significant difference in the result. This could be due to the fact that there were only three out of 60 students that answered in the affirmative on this question.

6.8 Results

The average number of documents reviewed was 43, with an average of 27.5 minutes total time and 58 seconds – just under one minute – spent per document. The average number of documents produced was 503 with an average recall of .50 and an average precision of .61.

IR performance results have been compared across three alternative methods for IR extraction: (1) The exploration approach measured by average *Recall* and *Precision* based upon incremental units of the *Total Time Explored* variable, (2) A random extraction of the 503 documents, representing the average number of documents produced by the participants' selections, (3) An extraction of documents based on the eDiscovery task verbatim after removing non-function words.

The graph of *Recall* against time appears in Figure 14. *Recall* performance for participants exploring the corpus for less than 15 minutes produced results in the .3 to .5 range, but outliers at the 14 and 15 minute data points make conclusions about a general trend within this time frame difficult to draw.

Participants exploring the corpus in the 23 - 30 minute time frame produced results in the .4 to .5 range and follow a mostly flat trend line. There is a gap up from .2 at the 15 minute mark to .5 at the 23 minute mark; but with no data points between 16 and 22 minutes, it is difficult to draw a conclusion about why this trend occurs.

There is also a significant trend upward between 30 minutes and 42 minutes, with a gap up between the 40 and 42 minute marks and no data point at 41 minutes. The recall performance results are fairly flat after an initial jump in performance between 30 and 42 minutes. This warrants further study to determine if there is a diminishing return beyond 42 minutes. A future experiment is planned to include a time frame up to 60 minutes to investigate this relationship.

Exploration outperformed random extraction at every data point. However, the verbatim non-function method was a better choice for users who spent 15 minutes or less and was competitive in several data points in the 15 – 30 minute range. Exploration outperformed verbatim non-function at all data points over 30 minutes; the effect also seems to flatten after an initial jump in this time range. This too will need to be studied further to determine if the effect remains flat, meaning that no further exploration yields improvement or if a time range beyond 45 minutes may continue to improve *Recall*. The study will be performed over a larger sample of users to make the results more generalizeable.

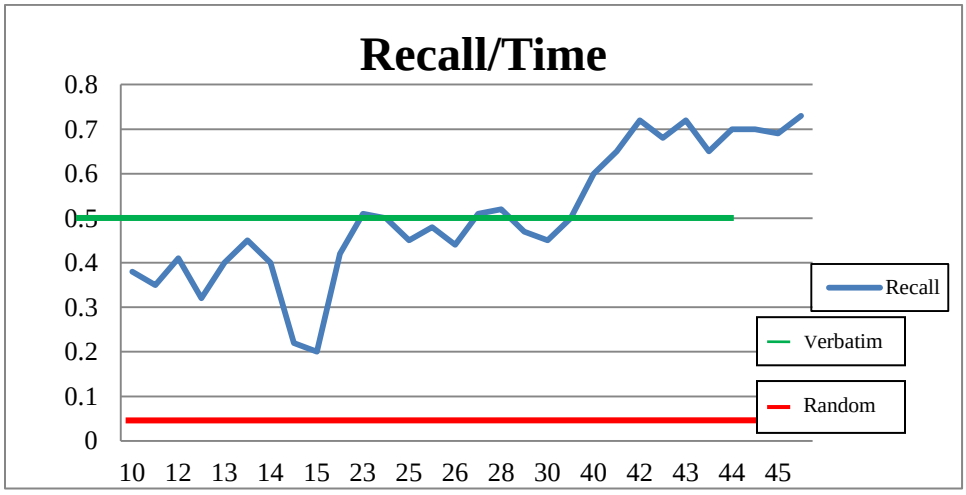

Figure 14: Recall over Time

The graph of precision against time appears in Figure 15. Precision performance follows a different trend than recall.

Participants exploring the corpus on the shorter end of the timeline, 23 minutes or less, produced results in a range of .6 to .8, with an outlier at the 14 minute mark. Participants exploring the corpus on the higher end of the timeline, greater than 40 minutes, produced results that were consistently above .6.

Participants exploring the corpus in the middle of the timeline, greater than 23 minutes but less than 40 minutes produced the worst results; they were in the .4 to .6 range. This is a strange result given that we expected participants to improve generally as time increased, and in this time range performance dropped. We have no explanation for why this may be so. We plan to further study this effect to investigate whether in fact a middle time frame exists that should be avoided by eDiscovery users.

The main results indicate that exploration can be an effective method for producing better precision than random extraction or verbatim /non-function word, and perhaps most effective (results above .60) when a user spends over 40 minutes or less than 13. However, it is difficult to justify such a conclusion with a small sample and with data point gaps within the ranges analyzed. Therefore, a future study has been planned to investigate this effect with a larger sample of users.

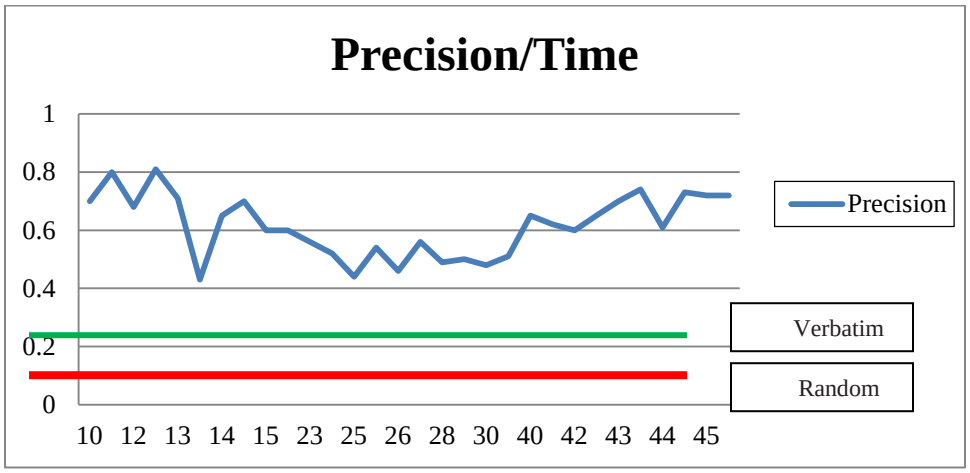

Figure 15: Precision over Time

A random extraction of documents was produced equal to the average number of documents produced by the participants in the study to determine if the exploration method would outperform chance. Given that the subset of documents contained 1,000 relevant out of 10,000 documents, a random extraction should produce 10% relevant documents. The average number of documents extracted based on participant performance was 503. If a random selection of 500 documents from the corpus was performed, the expectation would be approximately 50 documents out of 500 should be relevant. This would yield a precision of .10 and a recall of .05.

Given that there were 60 participants, we performed 60 random extractions and averaged the results.

When we performed the random extractions our average result was actually in line with expected chance performance. The average number of relevant documents extracted was 51, with a high of 68 and a low of 38.

Given that the worst performance using the exploration method was .20 for recall and .43 for precision, exploration outperformed random extraction.

In situations when the eDiscovery user has no *a priori* guidance for what search structure or terms that might produce relevant documents, sometimes the specific words from the request itself can be used as a good starting point to probe for initial trial and error results. The theory is that the terms in the request may in fact be significant indicators of relevant context.

When we performed this type of extraction we produced 2120 documents from the 10,000 item corpus, with 455 relevant. This extraction represents a recall of 455/1000 (.455) and a precision of 455/2120 (.215) – a pretty good starting point if the user has no prior knowledge. The exploration approach produced an average recall of .50 and an average precision of .61, outperforming Method 3 in both measures.

The results show that hypotheses H1aRandom and H1bRandom are both supported. The exploration participants outperformed random extraction at all data points in the study.

Hypothesis H2aVerbatim is partially supported. The exploration participants outperformed verbatim extraction in all data points greater than 30 minutes. Exploration did not outperform verbatim extraction in data points under 15 minutes and produced mixed results in the 15 to 30 minute range.

Hypothesis H2bVerbatim is supported. The exploration participants outperformed verbatim extraction for precision for all data points in the study.

There are several possibilities that may explain the results reported above. The most obvious explanation could be that a certain minimum amount of time must be given to a user to produce any improvement over verbatim extraction. This study suggests that, unless a user is prepared to spend more than 23 minutes on exploration, don't bother, simply use an automated approach such as verbatim.

Another explanation may be that, after a certain amount of time is spent exploring, there is a significant leap in knowledge acquired about the corpus. This study suggests that the number may be as little as 40 minutes or more to achieve this leap.

The flatter results produced in the 23 to 40 minute range are a mystery. There are several speculative explanations we could suggest. One possibility is; there may be a range of time spent in exploration that produces no increased effect, meaning if a user is going to spend less than 42 minutes, then the user might as well reduce that time to 23, because the additional 19 minutes will not produce any more productivity in recall.

6.8.1 Statistical Analysis of Three Models

Global F-tests were performed on the three models: Exploration, Behavioral, and Behavioral-Exploration. Each model was analyzed separately for *Recall* and for *Precision* using null and alternative hypotheses. A table summarizing the results for the behavioral and exploration hypotheses is contained in the discussion section.

A global F-test has been performed for Recall and for Precision. A summary of results appear in Table 5 and Table 6 on the next page.

The null and alternative hypotheses are as follows:

Recall	**Precision**
H_0: $B_1 = B_2 = B_3 = 0$	H_0: $B_1 = B_2 = B_3 = 0$
Ha: At least one Beta $\neq 0$	Ha: At least one Beta $\neq 0$

Where:

B_1 = Slope for Total time explored,

B_2 = Slope for Number of documents viewed,

B_3 = Slope for Time per document.

Table 5: SAS 9.2 Printout for Recall Variables

The REG Procedure
Model: MODEL1
Dependent Variable: **RECALL**

Number of Observations Read 60
Number of Observations Used 60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.08166	0.36055	105.21	<.0001
Error	56	0.19191	0.00343		
Corrected Total	59	1.27357			

Root MSE	0.05854	R-Square	0.8493
Dependent Mean	0.50733	Adj R-Sq	0.8412
Coeff Var	11.53878		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.23835	0.03069	7.77	<.0001
TOTALTIM	TOTAL TIME	1	0.00947	0.00142	6.69	<.0001
PERDOCTI	PER DOC TIME	1	-0.02839	0.03653	-0.78	0.4404
TOTALDOC	TOTAL DOCS VIEWED	1	0.00055612	0.00071032	0.78	0.4370

Table 6: SAS 9.2 Printout for Precision Variables

The REG Procedure
Model: MODEL1
Dependent Variable: **PRECISION**

Number of Observations Read 60
Number of Observations Used 60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.26712	0.08904	12.68	<.0001
Error	56	0.39312	0.00702		
Corrected Total	59	0.66024			

Root MSE	0.08379	R-Square	0.4046
Dependent Mean	0.61600	Adj R-Sq	0.3727
Coeff Var	13.60160		

Parameter Estimates

Variable	Label	Parameter DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.66597	0.04392	15.16	<.0001
TOTALTIM	TOTAL TIME	1	-0.00975	0.00203	-4.81	<.0001
PERDOCTI	PER DOC TIME	1	-0.00128	0.05229	-0.02	0.9805
TOTALDOC	TOTAL DOCS VIEWED	1	0.00502	0.00102	4.94	<.0001

The global F-test for the Recall exploration model and the Precision exploration model are both significant at alpha .01. However, *Recall* and *Precision* differ in which IVs are significant predictors. *Total Time Explored* is significant at alpha .01 for recall and precision. However, *Number of Documents Viewed* is significant at alpha .01 for *Precision*, but not for *Recall*; and *Time per document* was not supported for either *Recall* or *Precision*.

Table 7: Summary of Exploration Model Results

Independent Variables	Alpha	Dependent Variable Effected
Total Time Exploring*	.01	Recall & Precision
Number Documents*	.01	Precision
Time per Document	Not Significant	

***- Interactive Effect upon Precision**

The exploration independent variables have been analyzed for interactive effects. Total Time Explored and Total Number of Documents Viewed were found to have an interactive effect upon *Precision* and the relationship was significant at alpha .01. This suggests that the impact upon *Precision* by the *total time* spent in exploration depends on the *total number of documents* viewed, and the impact upon *Precision* by the *total number of documents* viewed depends on the *total time* explored. No other interactive effect was found to be supported. SAS 9.2 printout results from interactive tests appear on the next two pages in Table 8 and Table 9.

Table 8: Results from SAS 9.2 printout for interactive effect upon Recall

The REG Procedure
Model: MODEL1
Dependent Variable: RECALL RECALL

Number of Observations Read	60
Number of Observations Used	60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1.09217	0.18203	53.18	<.0001
Error	53	0.18140	0.00342		
Corrected Total	59	1.27357			

Root MSE	0.05850	R-Square	0.8576
Dependent Mean	0.50733	Adj R-Sq	0.8414
Coeff Var	11.53156		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.33478	0.12329	2.72	0.0089
TOTALTIM	TOTAL TIME	1	0.00912	0.00404	2.26	0.0280
PERDOCTI	PER DOC TIME	1	-0.04297	0.15645	-0.27	0.7847
TOTALDOC	TOTAL DOCS VIEWED	1	-0.00474	0.00328	-1.45	0.1540
TTPD		1	-0.00175	0.00644	-0.27	0.7864
TTTD		1	0.00009655	0.00005987	1.61	0.1128
TDPD		1	0.00160	0.00300	0.53	0.5965

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable RECALL

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	0.00350	1.02	0.3897
Denominator	53	0.00342		

Table 9: Results from SAS 9.2 printout for interactive effect upon Precision

The REG Procedure
Model: MODEL1
Dependent Variable: **PRECISION**

Number of Observations Read	60
Number of Observations Used	60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	0.38057	0.06343	12.02	<.0001
Error	53	0.27967	0.00528		
Corrected Total	59	0.66024			

Root MSE	0.07264	R-Square	0.5764
Dependent Mean	0.61600	Adj R-Sq	0.5285
Coeff Var	11.79238		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.15779	0.15308	7.56	<.0001
TOTALTIM	TOTAL TIME	1	-0.01673	0.00501	-3.34	0.0015
PERDOCTI	PER DOC TIME	1	-0.33524	0.19426	-1.73	0.0902
TOTALDOC	TOTAL DOCS VIEWED	1	-0.01290	0.00407	-3.17	0.0026
TTPD		1	0.00460	0.00799	0.58	0.5675
TTTD		1	0.00033831	0.00007434	4.55	<.0001
TDPD		1	0.00467	0.00373	1.25	0.2157

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable PRECISIO

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	0.03782	7.17	0.0004
Denominator	53	0.00528		

A global F-test has been performed upon the behavioral model. A summary of results appear in Table 10 on the next page.

The null and alternative hypotheses are:

Recall

H_0: $B_1 = B_2 = B_3 = B_4 = 0$

Ha: At least one Beta $\neq 0$

Precision

H_0: $B_1 = B_2 = B_3 = B_4 = 0$

Ha: At least one Beta $\neq 0$

Where:

X_1 = Tolerance for ambiguity (TOA),

X_2 = Locus of control (LOC),

X_3 = Disposition toward innovativeness (DISPO),

X_4 = Personal innovativeness in information technology (PIIT).

Table 10: Summary of Behavioral Model Results

Independent Variables	Alpha	Dependent Variable Effected	Beta Estimate
TOA	.01	Precision	.005
LOC	.01	Recall	-.013
DISPO	.05	Precision	.008
PIIT	Not Significant		

The global F-test for the *Recall* behavioral model and the *Precision* behavioral model are both significant at alpha .01. However, just like the exploration model, the behavioral model

differed in which variables were significant for *Recall* and which were significant for *Precision*. LOC was significant for *Recall* at alpha .01.

TOA was significant for *Precision* at alpha .01 and DISPO was significant for *Precision* at alpha .05. PIIT was not supported for *Recall* or *Precision*. The printouts for these results appear on the next pages in Table 11 and Table 12.

Table 11: SAS 9.2 Printout for Recall Variables

The REG Procedure
Model: MODEL1
Dependent Variable: **RECALL**

Number of Observations Read	60
Number of Observations Used	60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1.16472	0.29118	147.12	<.0001
Error	55	0.10885	0.00198		
Corrected Total	59	1.27357			

Root MSE	0.04449	R-Square	0.9145
Dependent Mean	0.50733	Adj R-Sq	0.9083
Coeff Var	8.76897		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.52230	0.04589	11.38	<.0001
LOC	LOC	1	-0.01291	0.00194	-6.64	<.0001
TOA	TOA	1	0.00043654	0.00149	0.29	0.7702
DISPO	DISPO	1	-0.00091858	0.00293	-0.31	0.7547
PIITSUM	PIIT SUM	1	0.00320	0.00124	2.59	0.0124

Table 12: SAS 9.2 Printout for Precision Variables

The REG Procedure
Model: MODEL1
Dependent Variable: **PRECISION**

Number of Observations Read 60
Number of Observations Used 60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	0.60044	0.15011	138.06	<.0001
Error	55	0.05980	0.00109		
Corrected Total	59	0.66024			

Root MSE	0.03297	R-Square	0.9094
Dependent Mean	0.61600	Adj R-Sq	0.9028
Coeff Var	5.35284		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.22744	0.03401	6.69	<.0001
LOC	LOC	1	-0.00012484	0.00144	-0.09	0.9312
TOA	TOA	1	0.00542	0.00110	4.91	<.0001
DISPO	DISPO	1	0.00833	0.00217	3.84	0.0003
PIITSUM	PIIT SUM	1	0.00003059	0.00091712	0.03	0.9735

The behavioral variables have been analyzed for interactive effects. Interaction between the independent variables was not found to be supported in the individual p-values but was support at alpha .01 in the partial F test. This conflicting result suggests there may be multi-collinearity among two or more of the variables. To account for this possibility we have tested whether any of the IVs correlate.

The Pearson Coefficient results indicate that DISPO and TOA are highly correlated. We plan to study this effect in future experiments to determine if one of the variables should be removed from the equation for parsimony. We also found that LOC and PIIT are highly negatively correlated. PIIT was not found to be significant as a main effect; however, this relationship suggests that we need to be careful drawing conclusions about the IVs' effects on Recall and Precision and we will need to further investigate this effect in our future work with larger populations. The SAS 9.2 results reports for interactive effects and multi-collinearity have been reproduced on the next pages in Table 13, Table 13.1 and Table 14.

Table 13: SAS 9.2 Printout for Recall Variables

The REG Procedure
Model: MODEL1
Dependent Variable: RECALL RECALL

Number of Observations Read 60
Number of Observations Used 60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	1.21496	0.13500	115.16	<.0001
Error	50	0.05861	0.00117		
Corrected Total	59	1.27357			

Root MSE	0.03424	R-Square	0.9540
Dependent Mean	0.50733	Adj R-Sq	0.9457
Coeff Var	6.74866		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.38595	0.13234	2.92	0.0053
LOC	LOC	1	0.01269	0.01158	1.10	0.2782
TOA	TOA	1	0.00620	0.00397	1.56	0.1244
DISPO	DISPO	1	-0.00244	0.00566	-0.43	0.6687
PIITSUM	PIIT SUM	1	0.00908	0.00787	1.15	0.2541
PIITSUMTOA		1	-0.00039540	0.00022606	-1.75	0.0864
PIITSUMDISPO		1	0.00025541	0.00048162	0.53	0.5982
LOCDISPO		1	-0.00008662	0.00073713	-0.12	0.9069
LOCTOA		1	-0.00068173	0.00035911	-1.90	0.0634
DISPOTOA		1	-0.00002182	0.00011459	-0.19	0.8498

The REG Procedure
Model: MODEL1
Test 1 Results for Dependent Variable RECALL

Source	DF	Mean Square	F Value	Pr > F
Numerator	5	0.01005	8.57	<.0001
Denominator	50	0.00117		

Table 13.1: SAS 9.2 Printout for Precision Variables

The REG Procedure
Model: MODEL1
Dependent Variable: PRECISIO PRECISION

Number of Observations Read	60
Number of Observations Used	60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	0.61878	0.06875	82.91	<.0001
Error	50	0.04146	0.00082926		
Corrected Total	59	0.66024			

Root MSE	0.02880	R-Square	0.9372
Dependent Mean	0.61600	Adj R-Sq	0.9259
Coeff Var	4.67482		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.38182	0.11131	3.43	0.0012
LOC	LOC	1	-0.01392	0.00974	-1.43	0.1589
TOA	TOA	1	-0.00226	0.00334	-0.68	0.5006
DISPO	DISPO	1	0.00607	0.00476	1.28	0.2082
PIITSUM	PIIT SUM	1	0.00063453	0.00662	0.10	0.9240
PIITSUMTOA		1	0.00014418	0.00019014	0.76	0.4518
PIITSUMDISPO		1	-0.00018739	0.00040508	-0.46	0.6457
LOCDISPO		1	0.00006220	0.00061998	0.10	0.9205
LOCTOA		1	0.00033756	0.00030204	1.12	0.2691
DISPOTOA		1	0.00017096	0.00009638	1.77	0.0822

The REG Procedure
Model: MODEL1
Test 1 Results for Dependent Variable PRECISIO

Source	DF	Mean Square	F Value	Pr > F
Numerator	5	0.00367	4.42	0.0021
Denominator	50	0.00082926		

Table 14: SAS 9.2 Printout of Multi-Collinearity Analysis

Pearson Correlation Coefficients, N = 60
Prob > |r| under H0: Rho=0

	PIIT	LOC	TOA	DISPO
PIIT	1.00000	-0.89706	-0.00623	-0.12841
PIIT		<.0001	0.9623	0.3282
LOC	-0.89706	1.00000	-0.22654	-0.07217
LOC	<.0001		0.0818	0.5837
TOA	-0.00623	-0.22654	1.00000	0.91590
TOA	0.9623	0.0818		<.0001
DISPO	-0.12841	-0.07217	0.91590	1.00000
DISPO	0.3282	0.5837	<.0001	

As indicated previously, a third model has been developed to investigate whether a significant relationship exists between the behavioral and exploration variables. The behavioral variables have been set up as the independent and the exploration variables have been set up as the dependent. A summary of the results appear in Table 15 below. The null and alternative hypotheses are the same from the behavioral model given that the same betas are being investigated (TOA, LOC, DISPO, and PIIT).

Table 15: Summary of Behavioral-Exploration Model Results

Independent Variables	Alpha	Dependent Variable Effected
TOA	Not Significant	
LOC	**.01**	**Number of Documents**
	.05	**Time Per Document**
DISPO	Not Significant	
PIIT	Not Significant	

In terms of the impact of the behavioral scales upon exploration behavior, LOC was significant for *Time per Document* at alpha .05, and for *Number of Documents* viewed at alpha .01. The other three scales were not significant for either *Recall* or *Precision*. SAS 9.2 printouts for these results are reproduced in Table 16 and Table 17 on the next page.

Table 16: SAS 9.2 Printout for Number of Documents Viewed

The REG Procedure
Model: MODEL1
Dependent Variable: **TOTAL DOCS VIEWED**

Number of Observations Read 60
Number of Observations Used 60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	36717	9179.18102	47.37	<.0001
Error	55	10657	193.75653		
Corrected Total	59	47373			

Root MSE	13.91965	R-Square	0.7751
Dependent Mean	43.66667	Adj R-Sq	0.7587
Coeff Var	31.87705		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	25.05723	14.35748	1.75	0.0865
LOC	LOC	1	-3.41746	0.60807	-5.62	<.0001
TOA	TOA	1	0.02524	0.46524	0.05	0.9569
DISPO	DISPO	1	0.97680	0.91530	1.07	0.2905
PIITSUM	PIIT SUM	1	-0.35271	0.38716	-0.91	0.3663

Table 17: SAS 9.2 Printout for Time per Document

The REG Procedure
Model: MODEL1
Dependent Variable: **PER DOC TIME**

Number of Observations Read 60
Number of Observations Used 60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1.47097	0.36774	6.64	0.0002
Error	55	3.04387	0.05534		
Corrected Total	59	4.51483			

Root MSE	0.23525	R-Square	0.3258
Dependent Mean	0.56833	Adj R-Sq	0.2768
Coeff Var	41.39313		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.59497	0.24265	2.45	0.0174
LOC	LOC	1	0.03403	0.01028	3.31	0.0016
TOA	TOA	1	0.00260	0.00786	0.33	0.7425
DISPO	DISPO	1	-0.00908	0.01547	-0.59	0.5595
PIITSUM	PIIT SUM	1	0.01221	0.00654	1.87	0.0674

The behavioral variables have been analyzed for interactive effects upon *Time per Document and Number of Documents Viewed*. The SAS 9.2 results report for interactive effects has been reproduced in Table 18 and Table 19 on the next page.

An interactive effect has been found to exist between DISPO and TOA upon *Time per Document*. This effect is supported at alpha .05. No other interactive effects were supported.

Table 18: SAS 9.2 Printout for Interaction (Time per Document)

The REG Procedure
Model: MODEL1
Dependent Variable: PERDOCTI PER DOC TIME

Number of Observations Read 60
Number of Observations Used 60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	2.09396	0.23266	4.81	0.0001
Error	50	2.42088	0.04842		
Corrected Total	59	4.51483			

Root MSE	0.22004	R-Square	0.4638
Dependent Mean	0.56833	Adj R-Sq	0.3673
Coeff Var	38.71668		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-1.70402	0.85049	-2.00	0.0505
LOC	LOC	1	0.20550	0.07439	2.76	0.0080
TOA	TOA	1	0.06818	0.02550	2.67	0.0101
DISPO	DISPO	1	0.06719	0.03638	1.85	0.0707
PIITSUM	PIIT SUM	1	0.07675	0.05057	1.52	0.1354
PIITSUMTOA		1	0.00020629	0.00145	0.14	0.8877
PIITSUMDISPO		1	-0.00242	0.00310	-0.78	0.4381
LOCDISPO		1	-0.00641	0.00474	-1.35	0.1824
LOCTOA		1	0.00055853	0.00231	0.24	0.8098
DISPOTOA		1	-0.00211	0.00073644	-2.87	0.0060

The REG Procedure
Model: MODEL1
Test 1 Results for Dependent Variable PERDOCTI

Source	DF	Mean Square	F Value	Pr > F
Numerator	5	0.12460	2.57	0.0379
Denominator	50	0.04842		

Table 19: SAS 9.2 Printout for Interaction (Total Documents Viewed)

The REG Procedure
Model: MODEL1
Dependent Variable: TOTALDOC TOTAL DOCS VIEWED

Number of Observations Read 60
Number of Observations Used 60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	39295	4366.08555	27.02	<.0001
Error	50	8078.56338	161.57127		
Corrected Total	59	47373			

Root MSE	12.71107	R-Square	0.8295
Dependent Mean	43.66667	Adj R-Sq	0.7988
Coeff Var	29.10932		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	58.61794	49.13050	1.19	0.2385
LOC	LOC	1	-0.73322	4.29734	-0.17	0.8652
TOA	TOA	1	-0.80526	1.47287	-0.55	0.5870
DISPO	DISPO	1	-0.88471	2.10130	-0.42	0.6755
PIITSUM	PIIT SU	1	-0.76073	2.92120	-0.26	0.7956
PIITSUMTOA		1	0.01383	0.08393	0.16	0.8697
PIITSUMDISPO		1	-0.00732	0.17880	-0.04	0.9675
LOCDISPO		1	-0.01698	0.27366	-0.06	0.9508
LOCTOA		1	-0.06714	0.13332	-0.50	0.6168
DISPOTOA		1	0.04745	0.04254	1.12	0.2700

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable TOTALDOC

Source	DF	Mean Square	F Value	Pr > F
Numerator	5	515.60918	3.19	0.0141
Denominator	50	161.57127		

Our analysis found no significant correlation between *Recall* and *Precision*. A printout of the Pearson Correlation appears in Table 20. Conventional wisdom has always been that *Recall* and *Precision* have an inverse relationship, in so far as, when one increases, it does so at the expense of the other. The reader will remember that this assumed relationship has fostered the alternative F-measures which discount for particularly lopsided Recall-Precision performance

trade-offs. The findings here are limited in that our sample size is only 60. However, we believe that the results produced here certainly warrant further study into the Recall-Precision relationship especially in light of our experiments in the next chapter which support a finding that precision can be enhanced without significant reduction in recall.

Table 20: Recall-Precision Correlation

Pearson Correlation Coefficients, N = 60

Prob > |r| under H0: Rho=0

	RECALL	PRECISIO
RECALL	1.00000	0.07847
RECALL		**0.5512**
PRECISIO	0.07847	1.00000
PRECISION	**0.5512**	

6.9 Discussion

Perhaps the most interesting and significant result produced in this study is that although *Total Time Spent Exploring* (TTE) is significant for both *Recall* and *Precision*, it is positively correlated for recall but negatively correlated for precision. This supports the claim that more time spent exploring the corpus leads to greater recall, but also leads to less precision. This result is consistent with prior research establishing the inverse relationship between *Recall* and *Precision* however, prior to this study no empirical explanation has been put forth. The result produced in this study provides a possible explanation for why this relationship is this way. The beta associated with *Total Time* for *Recall* was .009 and -0.097 for *Precision*, suggesting that for

every minute increase in *Total Time* we should expect to see an increase in *Recall* by almost .01 and a decrease in *Precision* by almost .10.

However, the study found that *Precision* is positively correlated with *Number of Documents Viewed*; the associated beta of .005, suggests that for every additional document viewed we should expect to see an increase in *Precision* by .005 units (a two document increase will produce a .01 increase in *Precision*).

The study found an interactive effect upon *Precision* by *Total time and Number of Documents Viewed* with a beta of -.016 for *Total Time,* a beta of -.013 for *Number of Documents Viewed,* and a beta for the interactive effect of .0003. This implies that for every 1 minute increase in *Total Time, Precision* will increase (or decrease) by -.016 + (.0003*number of documents viewed), and for every 1 document increase in the Total documents viewed precision will increase (or decrease) by -.013 + (.0003*time explored).

The linear equation looks like this:

$$\text{Precision} = B_0 + B_1T + B_2N + B_3T*N$$

$$\text{Effect of Time on Precision} = (B_1 + B_3N)$$

$$\text{Effect of Documents on Precision} = (B_2 + B_3T)$$

Where:

T = Total Time Explored

N = Number of Documents Viewed

In terms of behavioral factors impacting *Precision*, TOA reports a beta value of .005. The TOA inventory used in this study is scored based upon a person's lack of tolerance, the higher someone scores, the less tolerant they are. This suggests that for every 1 point increase in an

individual's TOA score *Precision* will increase by .005 units. This intuitively makes sense, given that people less tolerant of ambiguity are going to focus their search narrowly, resulting in less non-relevant documents being returned. However, TOA was not significant in *Recall*. DISPO was significant in precision at alpha .05. The associated beta of .002 suggests that for every 1 point increase in DISPO score an individual will produce .002 more units of *Precision*.

In terms of *Recall*, the only significant behavioral variable was LOC, at alpha .01. The associated beta of -0.01 suggests that for every 1 point increase in LOC score an individual will produce .01 less units of *Recall*. A lower LOC score indicates the individual believes he/she controls their fate rather than external factors. Therefore, a higher LOC should lead to less recall and a lower LOC should lead to greater recall.

The results produced are consistent with our original hypothesis that people with greater internal LOC will be inclined to search broader and therefore produce higher recall. One example of perceived control and its effect upon IR came up during our post-task interviews. Subject PG1 indicated that he was; "less concerned about missing documents." Whereas subject MG2 indicated that; "I feel I may miss 'the smoking gun.'"

A list of the hypotheses with their measured variables and associated betas is listed in Table 21 below.

Table 21: List of Hypotheses Supported and Not

Hypothesis	Supported/Not	Variable	Alpha	Relationship to Recall/Precision
H1a	Supported	TTE	.01	Recall: Direct and Pos
H1b	Supported	TTE*	.01	Precision: Direct and Neg*
H2a	Not	NUM		
H2b	Supported	NUM*	.01	Precision: Direct and Neg*
H3a	Not	PER		
H3b	Not	PER		
H4a	Not	TOA		
H4b	Supported	TOA	.01	Precision: Direct and Pos
H5a	Supported	LOC	.01	Recall: Direct and Pos
H5b	Not	LOC		
H6a	Not	DISPO		
H6b	Supported	DISPO	.05	Precision: Direct and Pos
H7a	Not	PIIT		
H7b	Not	PIIT		

*- **Interactive effect upon Precision supported**

As previously mentioned, possible links between the behavioral scales and exploratory behavior were also evaluated for significance. The only significant behavioral variable was LOC. *Time per Document* was affected by LOC with a beta of .034 at alpha .05. This means that for every 1 point change in LOC, an individual will, on average, spend .034 more minutes per document; a significant, but perhaps not meaningful amount of time differential. However, the real insight comes in the form of recognizing that a relationship exists between these variables that can be exploited by the eDiscovery practitioner. Remember that a higher LOC score translates into less internal and greater external LOC. This may be important in situations where the length or complexity of documents in a corpus is of particular criticality. Persons with less internal and greater external LOC will, on average spend more time per document.

The opposite effect was found in *Total documents viewed* (NUM), where the effect had a beta of -3.41 at alpha .01. In this instance, individuals with higher scores on the LOC scale

having less internal locus will, on average, view 3.4 fewer documents for every 1 point increase in their LOC score. This result is both significant and meaningful. The practitioner will be informed that users with higher LOC scores can be expected to view fewer documents.

An unanticipated interactive effect upon *Time spent per document* by DISPO and TOA was discovered to exist. This effect is supported at alpha .05; with a beta of .067 for DISPO, .068 for TOA, and a beta for the interactive effect of -.002. This implies that for every 1 unit change in DISPO score, *Time per document* will increase (or decrease) by .067 + (-.002*TOA score), and for every 1 unit change in TOA, *Time per document* will increase (or decrease) by .068 + (-.002*DISPO score).

The linear equations look like this:

$$\text{Time per document} = B_0 + B_1D + B_2T + B_3D*T$$

$$\text{Effect of DISPO upon Time per document} = (B_1 + B_3T)$$

$$\text{Effect of TOA upon Time per document} = (B_2 + B_3D)$$

Where:

D = DISPO

T = TOA

A beta value of .002 minutes is a rather small number, so this effect although significant, may not be meaningful. Further studies will need to be conducted to determine the impact of this relationship.

Table 29 contains a complete list of the Independent and Dependent Variables used in this study, along with their scales, ranges, means and standard deviations. It has been included in Appendix – K.

Table 30 contains a printout of the Pearson Coefficient Correlations among the Independent Variables. It has been included in Appendix – L.

6.10 Limitations

This study like all studies has limitations. The first limitation lies in the sample size. Several variables were found to not be significant. One possible reason for this is our small sample size; N=60. We plan to address this by collecting more data in future studies.

A second limitation in this study is the use of law students as an approximation for legal professionals such as lawyers and paralegals. In this case, the use of law students was helpful to us because they had the requisite understanding of legal terminology and strategies in litigation, but they were not jaded by years of legal experience that may impact the study. We plan to conduct future studies with paralegals and lawyers to determine if legal experience matters in this form of IR. For that reason we have designed covariates to track data such as this. We plan to implement this design in our next study.

6.11 Contribution

This study has demonstrated the feasibility of an exploration method instantiated through an automated tool that allows users to acquire knowledge about the context of a corpus and apply that knowledge in their search strategy, thereby addressing a major issue researched in eDiscovery IR – how to resolve the dilemma of context to improve recall and precision in large corpora.

The study reported in this chapter makes several significant contributions to theory. The main contribution is the investigation into how exploration can be useful for large collection information retrieval. The results produced by our experiment support the finding that user exploration of a small portion of a collection will yield improvement at various time intervals. There is clearly a relationship between time and number of documents and IR results produced.

How much time and how many documents are needed for a minimum effect will be investigated in future experiments. The results that have been produced by this experiment indicate that there are ranges within which performance improves and ranges within which performance suffers.

We have investigated behavioral and exploration relationships and discovered some new relationships. First, this study has provided insight into how IR behavioral variables may be used to predict a user's result. Second, this study demonstrates how an exploration model approach to IR can improve performance, specifically when measured against random and non-function word methods.

This study provides insight into which exploration variables can be used to predict IR outcomes and more importantly, how these variables can be used to enhance user productivity.

This study has investigated the underlying constructs of IR in the eDiscovery domain and reported on initial relationships that appear to be present. The next step will be to develop the unresolved questions in future experiments that have come out of this study.

6.12 Future Work

We are encouraged by the results we have produced in this study, particularly in the possible explanation for the recall-precision inverse relationship and in the differing retrieval results for the ranges of time and the number of documents explored. We plan to continue with an additional series of experiments using alternative document collections to cross-validate the results produced here; our next data set will involve a medical records database. We also plan to conduct further behavioral experiments in IR using (RFT) regulatory focus theory. The goal with RFT is to determine whether individuals can be primed to prefer recall or precision preferences;

this will provide an additional tool for practitioners who wish to design an eDiscovery strategy that favors recall or precision.

6.13 Conclusion

Study One is designed to measure the significance of the relationship between exploration of a sample collection and the IR result, and how exploration impacts user performance. Conventional wisdom suggests there should be a direct and positive correlation between exploration and result. This study produced results that showed that the relationship is not linear and in fact, at some ranges performance suffers and exploration should be avoided.

The results produced by this study help explain which behavioral preferences have significant impact on exploration and on IR performance. This study also provides an explanation for why recall and precision are correlated in an inverse relationship. The measured variables used in this study help explain user actions and strategies developed during corpus and document exploration and their significance upon IR performance.

The IT artifact developed for this study is a prototype system designed to support the exploration process for eDiscovery IR. Proof of concept is instantiated via the Design Science Paradigm. The contribution of this study lies in its insights of how differences in exploration variables *Total Time* invested in exploring, the *Number of Documents* viewed and *time spent per documents*, and behavioral variables *locus of control*, *tolerance for ambiguity* and *disposition toward innovativeness* impact the IR result as evaluated by *Recall* and *Precision*.