Modeling Information-Seeking Expertise on the Web

Diana Tabatabai
Department of Educational and Counselling Psychology
McGill University, Montreal

Thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Educational Psychology

Your file Votre référence
ISBN: 0-612-85746-8
Our file Notre référence
ISBN: 0-612-85746-8

Acknowledgements

My thanks to my supervisor, Dr. Bruce M. Shore who is a true scholar, an ethical, flexible yet firm person with clarity of mind who set clear expectations for me and helped me excel. Thanks to my committee members, Dr. Alain Breuleux, and Dr. Jamshid Beheshti whose diverse background helped me look at this study from different perspectives.

I am grateful to all the participants and also Dr. Mark W. Aulls for his support. Also thanks to Francesca Luconi whose friendship I treasure. My special thanks to Sue Davies whom I have the good fortune of knowing for almost a decade: Thank you for your warmth and wisdom.

I also extend thanks to my family. To my mother, Ozma Khanoom Mortazavi, who is an unconventional woman born and raised in a very traditional society. To my late father, Dr. Mir Jalaladdin-Khan Tabatabai-Shafeii, who instilled the love of research in my mind by conducting research on cancer that got him the Medical degree from l'Université de Lyon in 1935. To my two daughters, Sarah and Mina Etezadi-Amoli, for showing me that if I can survive their teenage years, I can survive a PhD thesis, and to the fourth generation, Kaïa Etezadi-Smith, whose unexpected arrival has enriched my life beyond imagination. I owe the greatest gratitude to my husband, Dr. Jamshid Etezadi-Amoli, who has been by my side and showed unwavering confidence in my abilities and me. He never gave up even when I almost did, which enabled me to finish this work.

Abstract

Searching for information pervades a wide spectrum of human activity, including learning and problem solving. With recent changes in the amount of information available and the variety of means of retrieval, there is even more need to understand why some searchers are more successful than others. This study was undertaken to advance our understanding of expertise in seeking information on the Web by identifying strategies and attributes that will increase the chance of a successful search on the Web. A model that illustrated the relationship between strategies and attributes and a successful search was also created. The strategies were: Evaluation, Navigation, Affect, Metacognition, Cognition, and Prior knowledge. Attributes included Age, Sex, Years of experience, Computer knowledge, and Info-seeking knowledge. Success was defined as finding a target topic within 30 minutes. Participants were from three groups. Novices were 10 undergraduate pre-service teachers who were trained in pedagogy but not specifically in information seeking. Intermediates were nine final-year master's students who had received training on how to search but typically had not put heir knowledge into extensive practice. Experts were 10 highly experienced professional librarians working in a variety of settings including government, industry, and university. Participants' verbal protocols were transcribed verbatim into a text file and coded. These codes, along with Internet temporary files, a background questionnaire, and a post-task interview were the sources of the data. Since the variable of interest was the time to finding the topic, in addition to ANOVA and Pearson correlation, survival analysis was used to explore the data. The most significant differences in patterns of search between novices and experts were found in the Cognitive, Metacognitive, and Prior Knowledge strategies. Based on the fitted survival model, Typing Keyword, Criteria to evaluate sites, and Information-Seeking Knowledge were associated with timely success of the search.

Résumé

La recherche de l'information couvre un large spectre d'activité humaine, y compris la resolution de problème et l'apprentissage. Avec les changements récents dans la quantité d'information disponible et la variété de moyens de recherche, il y a même plus de besoin pour comprendre pourquoi quelque rechercheurs ont plus de succès que d'autres. Cette étude a été entrepris pour avancer notre compréhension d'expertise dans la recherche d'information sur l'internet en identifiant les stratégies et attributs qui augmentera la probabilité de la recherche sur l'internet avec succes. Un modèle qui a illustré le rapport entre stratégies, attributs et une recherche avec succes, a été créé. Les stratégies étaient: Evaluation, Navigation, Affecte, Metacognition, Cognition, et Connaissance Précédent. Les Attributs inclus age, sexe, années d'expérience, connaissance d'ordinateur, et connaissance de la recherche d'information. Le succès a été défini en trouvant un sujet de cible dans 30 minutes. Les participants étaient de trois groupes. Les novices étaient 10 étudiants d'enseignement de pré-service non-diplômé qui ont été entraîné dans la pedagogie mais pas en particulier dans la recherche d'information. Les intermediates était neuf étudiants dans l'année finale de leurs maitrise qui avait reçu l'entraînement sur la recherche d'information mais n'avait pas typiquement mis leurs connaissances en pratique. Les Experts étaient 10 bibliothécaires professionnel travaillant dans une assortiment de montages y compris gouvernement, industrie, et Université. Les protocoles verbaux des participants ont été transmit dans un fichier texte et encodé. Les sources des données etaient ces codes, avec les fichiers d'internet temporaire, un questionnaire des éléments de base, et un entrevue poste-tâche. Puisque la variable d'intérêt était le temps à trouver le sujet, en plus de la ANOVA et corrélation de Pearson, “survival analysis” a été utilisée pour explorer les données. Les différences les plus significatives dans les modèles de recherche entre novices et experts étaient trouvé dans les stratégies cognitives, metacognitives, et connaissance précédent. Basé sur le model "fitted survival," Taper le mot-Clé, les Critères pour évaluer les sites, et Connaissance de la recherche d'information ont été associé avec le succès opportun de la recherche.

List of Tables

Table 1. *A Comparison of Three Groups of Participants* 29
Table 2. *Template for Transcribing the Audiotape* 34
Table 3. *Complete List of Codes Used to Code Verbal Protocols* 35
Table 4. *A Comparison of Three Groups Broken Down by Strategies* 47
Table 5. *Results of Analysis of Variance and Effect Size on Six Strategies* 48
Table 6. *Significant Results for Evaluation: Comparison of three Levels of data* 50
Table 7. *Significant Results for Navigation: Comparison of three Levels of data* 52
Table 8. *Significant Results for Affect: Comparison of three Levels of data* 57
Table 9. *Significant Results for Metacognition: Comparison of three Levels of data* 59
Table 10. *Significant Results for Cognition: Comparison of three Levels of data* 61
Table 11. *Significant Results for Prior Knowledge* 63
Table 12. *Results of Analysis of Variance and Effect Size on Attributes* 64
Table 13. *Summary of Cox Survival Analysis Results: Significant Variables* 66
Table 14. *Information Seeking Model - Variables in the Equation* 69

List of Figures

Figure 1. Survival function at mean of covariates 70

Figure 2. Hazard function at mean of covariates 71

Table of Contents

Acknowledgements 2
Abstract 3
Résumé 4
List of Tables 5
List of Figures 6
Chapter 1: Introduction and Literature Review 9
Literature Review 12
Information-Seeking as Complex Problem Solving 12
Models of Information Seeking 23
Summary 26
Purpose and Research Questions 27
Hypotheses 28
Chapter 2: Method 29
Participants 29
Materials 31
Procedure 32
Data Overview 33
Coding System 35
Data Analysis 42
Chapter 3: Results and Discussion 46
Does the Level of Expertise Impact on How People Search? 47
Evaluation 48
Navigation 51
Affect 54
Metacognitive 57

Cognitive 60

Prior Knowledge 62

Attributes 63

Is the Use of Certain Strategies Associated with Success? 65

Can a Successful Search on the Web be Modeled? 68

Chapter 4: Summary and Conclusions 72

Contributions to Knowledge 76

Limitations 77

Future Research 77

References 78

Appendices

Appendix A: Glossary 84

Appendix B: Certificate of Ethical Acceptability For Funded and non-Funded Research Involving Humans 86

Appendix C: Equipments and Material 87

Appendix D: Consent form 88

Appendix E: Background-Information Questionnaire 89

Appendix F: Post-task Interview Questions 90

Appendix G: Frequencies of All Codes 91

Chapter 1: Introduction and Literature Review

Information is difficult to discuss in isolation from the storage and retrieval systems that provide access for its users. Although searching for information is nothing new, the systems that process and organize it have recently gone through tremendous changes. Over the past century, many technologies such as telephone, television, cars, airplanes, and computers have been developed that have changed how we live and how we work. What is significantly different about new technologies is the exponential rather than linear effect of merging single technologies, where not only is there a greater impact, but perhaps also a different kind of effect on the society. As Cartwright (1982) had predicted, we not only do things differently but also do different things. For example, the shift from mainframe computers to stand-alone computers brought about a change in the primary function of computers, namely from computation to communication. It seems that the effect of merged technologies is greater than the sum of the singular effects of each alone. Specifically, the merger of information databases and communication technology has resulted in electronic information technology with an impact that surpasses both.

The search process has accelerated and searchers can access more information. At the same time, information retrieval and evaluation of its quality has become more difficult because of the explosion in the amount of information, the many different ways it can be accessed, and the difficulty in defining what exactly is information. Information not only means different things to different people but its definition and focus also change over time. Now it seems that information has become a commodity in itself, with numerous organizations having expertise solely in how to organize, select and retrieve it. The information seeker no longer has to submit a request to an intermediary who would submit it to the computer to be processed in batches. In those manual systems, users had to wait for a reply and resubmit the request if they wanted any changes. In contrast, on-line information retrieval (IR) systems are interactive and can be accessed from remote locations. As a result of this easier access, the number of on-line searches grew from almost one million in 1974 to 80 million in 1997 (Walker & Janes, 1999), and continues to increase rapidly.

With the daunting amount of information available, the ability to seek and evaluate the appropriate information has become a key requirement for success in our "digital" society (Tapscott, 1996). Consumers of information have to develop expertise in dealing with information that often challenges the old goals of traditional classrooms. With so much at stake, it is not surprising that recent sources of information such as the World Wide Web (the Web) and the Internet have become targets of research. As more and more users sought information, there was a need for a shift in the focus of research. For decades, IR research was focused on system-centered questions. These types of inquiries still account for a large portion of research, but the trend to explore user characteristics and to understand the cognitive processes used to seek information is gaining momentum (Walster, 1996). As the focus of research has changed from system demands to user characteristics, so has the research methodology. Quantitative studies have been replaced or augmented by qualitative techniques such as the analysis of verbal protocols.

The Web, despite rich data and powerful connections, will not support learning by itself. The challenge is to turn it into a cognitive tool, in the manner that Lajoie and colleagues (Lajoie & Derry, 1993; Lajoie, 2000) have characterized computers. The metaphor of "cognitive tool" implies that there are tools that can help learners with their cognitive tasks. The authors described two computer systems that act as such tools, Sherlock, a computer-based learning environment for avionics troubleshooting, and Bio-World, a computer-based learning environment where high school biology students can practice diagnosing infections (Lajoie, 1993). These environments promote learning for a variety of reasons including support memory and lower level cognitive skills, therefore leaving more resources available for higher order thinking skills. They give the learners an opportunity to engage in cognitive activities, in the context of problem solving, to an extent that they would not be able to experience in real world. Likewise, an important transition for the Web and hypermedia should be from a mere source of information to a partner in learning to amplify and augment human cognitive abilities.

Dillon and Gabbard (1998) conducted a review of quantitative research on the use of hypermedia in education. They noted that, for one or more of the following reasons, the hypermedia capabilities of the Web are advocated as having a powerful cognitive effect on education: non-linear access to large amount of information, ability for the information to be explored deeper and on demand, learners' option to progress at their own pace, the engaging nature of hypermedia, and, finally, the way knowledge is organized in hypertext reflects the organization of knowledge in human mind. The authors continued by stating that, "Such writing is strong on claims but, so far, short on supporting evidence from studies of learners" (p. 323). Similar to computers, hypertext capabilities of the Web alone do not turn the Web into a cognitive tool. What is missing is identifying and incorporating strategies that will support the cognitive processing of the information. In order to exploit these strategies, there is a need to measure the information success of those whom Marchionini (1995) calls information specialists, referred to in the present study as successful information seekers.

According to Chi and Glaser (1985), every problem has three components: initial state, goal state, and a path that connects these two. Problem solving occurs when the current state of information is less than desired and there are barriers between the initial state and goal state. Most problems can be classified into the categories of well-defined and ill-defined problems. In a well-defined problem, the initial state, the path, and the goal state are clear. In an ill-defined problem, any or all of the three components may be vague and unclear. Similarly, most domains can be classified into the categories of well-structured and ill-structured domains. In a well-structured domain such as mathematics, most of the tasks are well-structured and there are restrictions as to what the problem solver can use to carry out the task (for example only using digits or letters, or both given in the problem statement). In an ill-structured domain, most of the tasks are ill-structured, i.e., the components of a task are vague and unclear. The Web can be considered as an ill-structured environment given its vastness and its openness of goals and means to reach them (Hannafin, 1999). Information seeking on the Web is an ill-defined problem-solving task due to the fact that in searching on the Web, at least two of the three components, namely the path and goal state, are changing.

Expertise in seeking information using traditional means (such as catalogues, books, CD-ROMs, and libraries) has been extensively studied for well-defined problems and there is a growing amount of research on how librarians search for information and what constitutes expertise in library studies. However, little is known about strategies and attributes and problem-solving skills of experts in ill-structured environments. Specifically, it is not known whether expertise and strategies and attributes of experts in traditional environments can be successfully applied to problem solving in the Web.

Literature Review

The two theories that frame this study are information-processing theory and theories of expertise. They provide an opportunity to analyze information-seeking processes from a complex problem-solving approach, and identify successful information seekers as experts in their field. The review of literature that follows, therefore, will focus on (a) Information seeking as complex problem solving, especially in Open Ended Learning Environments (OELE) such as the Web, and (b) information-seeking models that are created to measure users' strategies and attributes while searching on the Web.

Information Seeking as Complex Problem Solving

Although information can take several different forms, such as information-as-process, information-as-knowledge, and information-as-thing, most generally information is defined as anything that can change the state of one's knowledge (Marchionini, 1995). Much of human activity is information seeking--a purposeful search for information in order to bridge the gap between what is known and what is unknown--and as such, it demands cognitive processes. In order to become successful information seekers, we develop many kinds of knowledge. For example, general knowledge that can be applied in different situations, domain-specific knowledge that pertains to a particular task or problem, declarative knowledge or knowing "that" something is the case, procedural knowledge or knowing "how" to solve a problem, and finally conditional knowledge meaning knowing "when and why" to apply the declarative and procedural knowledge that has been acquired. Gagné (1985) called the last, cognitive strategies. For many, this is the stumbling block in solving problems. Facts and procedures are known, but due to a lack of appropriate cognitive strategies, people are not able to apply what they know to solve a problem.

Based on information-processing theories (Greeno, 1978; Simon, 1979), problem solving involves the following important cognitive processes: (a) construction of a problem representation or problem space, (b) a solution process or problem-solving space that involves problem solvers' search within the problem space, and (c) control processes to check the success or failure of the solution. In order to understand or represent the problem, the solver may activate certain knowledge in memory, i.e., a strategy or a schema. Problem-solving strategies can be schema-driven or involve the comparison of initial state to goal state and then trying to bridge the gap, as in means-ends analysis (Gick, 1986).

The Web is significantly different from other information media for several reasons. In terms of structure, there is no implicit index or hierarchy to the Web, making it more difficult to create a mental model of the information. As such, it is hard for users to imagine their location in the overall map of the knowledge they are searching. This potentially leads to information overload and the feeling of disorientation. The Web also differs in terms of content. The fact that anyone with a computer and a modem can publish on the Web leads to the question of validity of content and authorship, and makes it difficult to separate out the useful information. Searching in the Web is an ill-defined task, different from searching in the library. The Web's unique parameters raise the question of what kind of different strategies and attributes experts require for successful problem solving in this environment.

Traditionally, library and information studies have focused on providing extensive access to vast sources of information in what Kuhlthau (1993) calls the "bibliographic paradigm" (p. 1). One of the most important concepts in this paradigm has been the retrieval of information from the system's perspective. There is a plethora of research in this field on what matches the system's representation of texts rather than how to address users' problems and process of information seeking. In contrast, open-ended systems or open-ended learning environments (OELE) are learner-centered systems that may include a variety of components such as electronic card catalogues, CD-ROMs, or the Web (Hannafin, Hall, Land, & Hill, 1994). The subtle difference between the Web and CD-ROMs is that while users of the Web operate within a large and ever changing problem space, information seekers using CD-ROMs are interacting within a more limited and static problem space. Nevertheless, lessons learned from how users interact with information on CD-ROMs are invaluable to understanding the Web search process.

Large, Beheshti, and Breuleux (1998) investigated the navigational skills of grade six students by observing them using the *Castle Explorer* CD/ROM to write an essay on the middle ages. Results showed that the majority enjoyed using the computer and even capably explored unrelated areas such as occupations during the middle ages. However, their naïve information-seeking strategies and lack of abilities to create synonyms impeded their success. Expert users employed more analytic search strategies such as thinking, planning, and evaluating. Novices preferred browsing and interacting with the interface without using the on-line help. Marchionini (1989) drew similar conclusions from a fact-retrieval study using full-text encyclopedias and hypertexts. He compared 28 third and fourth graders with 24 sixth graders. Data sources were keystrokes and observers' notes. Results show that both groups used the same number of moves but older students were more successful in their search because they used more analytical or goal-oriented strategies and planning, whereas novices used more heuristic interactive strategies.

As more and more of the traditional systems migrate to the Web, researchers in library and information studies have expanded their scope to include information retrieval on the Internet and the Web, and that has changed some elements of searching for information. To begin with, there is no control on who provides information for the users. With its prolific growth, any commercial or private entity with an account can create its own database and publish it on the Web (high quality or inane). Users of the information are also vastly different in terms of both technical aptitudes and individual characteristics (Borgman, 1989). Finally, the focus of the research is also changing. All these changes put a strain on all the domains dealing with information; even core concepts in information science such as information retrieval and information seeking are being re-examined (Hjorland, 1997).

Marchionini (1995) defined information seeking as "a process in which humans purposefully engage in order to change their state of knowledge" (p. 5). He preferred seeking to retrieval because it is more human-oriented and, in addition, the "re" in "information retrieval" implies finding an item that was known before, and that makes it not applicable to some learning situations. Other concepts such as browsing and searching are also under scrutiny. In general, searching and browsing are considered discrete activities, one purposeful and formal and the other random and informal. However, Cove and Walsch (1988) gave three new definitions of strategies: Searching is equated with Search Browsing, browsing is equated with General Purpose Browsing, and he added a new strategy called Serendipitous Browsing where information seeking is purely random.

In agreement with the above authors, Bates (1989) stated that searching and browsing are integrated and much of our searching is primarily browsing. Marchionini also put more emphasis on browsing, defining three types of browsing: directed, semi-directed, and undirected. Similarly, Wilson (1997) categorized strategies based on how attentive the user was when looking for health-related material: passive attention (e.g., watching TV), passive search (i.e., incidental), active search, and ongoing search.

The concept that helped browsing become a research topic was hypermedia (Large, Tedd, & Hartly, 1999). A hypermedia document is a hypertext document with various media links, such as images and sounds. A hypertext document is an electronic document that contains hyperlinks. Hyperlinks are built-in links that enable a user to access information in a non-linear fashion. Although Ted Nelson is credited with having coined the term hypertext in the 1960s (Large, et al., 1999, p. 187), the concept was introduced twice before. First, it seems that Thorndike, (1912) had a vision of hypertext when he said, "If, by a miracle of mechanical ingenuity, a book could be so arranged that only to him who had done what was directed on page one would page two become visible" (pp. 164-166). Then Vannervar Bush (1945) wrote about our inability to make real use of the records in a library because they are sorted alphabetically or numerically and "the human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain" (Section 6). Browsing by jumping from one link to another, as is done in hypermedia, is considered to be a major impetus behind the popularity of the Web.

Catledge and Pitkow (1995) carried out the first major study of Web browsing behavior that was published on the Web. For a three-week period, 170 users of XMosaic browser consented through a consent window that the log file showing their interaction with the Web is saved on a disk for further analysis. Results were measured based on three task levels: High Level (application, e.g., Open File), Mid Level (interface action, e.g., pull-down menu), and Low Level (interface technique, e.g., mouse click). The study yielded some expected and some unexpected results. As expected, hyperlinks (52%) and the Back command were the most popular actions. Browsers were less patient than searchers and spent less time exploring the sites. They surprisingly found that only 2% of sites were either saved or printed and the sites that were book-marked were not necessarily the most popular ones. The major limitation of this study is that a qualitative analysis was not performed to further explore the reasons behind the findings. In addition, it is not clear whether the users had easy access to disks and printers, or if they were even aware of their availability. Furthermore, there is the question of whether the minimal bookmarking was due to their lack of know-how or to a lack of interest in the sites.

One of the most comprehensive studies investigating cognitive strategies and learning from the Web was by Hill and Hannafin (1997). Fifteen educators initially participated in the study. They searched the Web for a topic they chose while thinking aloud. Sources of data were a pre-search survey, think-aloud protocols, an audit trail, post-search questionnaires, and stimulated post-search interviews. Five attributes were studied in relation to five research questions: Metacognitive knowledge, perceived orientation, perceived self-efficacy, system knowledge, and prior subject knowledge. Four participants were chosen for an in-depth qualitative analysis. Disorientation, a problem often cited in the use of OELEs, played a significant role. However, the participants who possessed metacognitive strategies and had a broader knowledge base were more successful and reported much less disorientation. This study provided valuable insight into the search strategies of information seekers but, due to its limited sample size, the results cannot be generalized.

Land and Greene (2000) focused on the process of seeking, locating, and integrating information retrieved from the Web. This was a project-based study in which nine undergraduate pre-service teachers were asked to create Internet projects to be integrated into the curriculum. Think-aloud protocols and video observations were used. Although the focus of this study was on seeking, locating, and integrating resources in instruction, their findings are critical in developing useful information-seeking strategies. The first finding was that progressing from a data-driven search to a goal-driven approach is a strategy that helped the participants narrow down their search. The next finding was that once they were challenged in finding relevant information on the Web, some of them scaled back their projects to make the task more manageable. In the last finding, the authors reported that metacognitive, domain, and system knowledge were critical to the success of the participants. It seems that the best way to help information seekers is to use several forms of scaffolding to help them reflect and articulate their understanding while they search in a complex environment such as the Web.

There has been a change in the focus of research on information seeking from "bibliographic paradigm" to "user paradigm." Studies on user characteristics, especially users on the Web, are in their infancy but growing rapidly. What will help future studies is to consider the Web as an ill-structured medium and information seeking as a complex problem-solving activity. As such, more qualitative studies directed towards understanding the cognitive aspects of Web users should be incorporated in research that focuses on the process of information seeking and how it can be improved through education and training. In treating information seeking as a cognitive activity, it is important to explore the concept of expertise, as it exists in the literature and how it relates to information seeking.

Expertise and information seeking. The book, *Thought and choice in chess* (de Groot, 1965) provided a rare impetus for an impressive amount of research on expertise during the past three decades. The focus of subsequent studies was also on well-structured domains of science such as chess and physics. The results of these studies are mostly explained using the principles of the information-processing approach. The exceptional performance of experts in terms of knowledge and skills that are acquired through experience is compared to other learners sometimes labeled as "novice," who are not yet well defined. The experts are described in terms of how their knowledge is organized and how they acquire skills in comparison to novices. Among the expert-novice differences, the followings have been cited more often: (a) novices' knowledge of a domain is structured around the main phenomena in a domain, experts' knowledge is compiled in a condition-action form, (b) experts perceive their domain as connected meaningful chunks or patterns, (c) experts knowledge becomes proceduralized, (d) experts develop automaticity of basic operations, (e) experts have superior memory, (f) experts are faster at solving problems in their domain and make fewer mistakes, (g) novices' representation of the problem is based on the givens in the problem statement, experts tend to use their procedural knowledge to represent problem and arrive at solutions, (h) experts spend more time examining the relevant information, perhaps chunk information differently, (i) experts are goal driven and use forward reasoning, (j) expertise develops from knowledge initially acquired from methods such as means-ends analysis, but increases steadily with practice, (k) expertise can be taught by means of teaching expert rules, therefore, we can predict the performance of experts based on the rules they use, (l) expertise is domain-specific, experts' performances outside their domain mirrors non-experts, (m) experts display more self-regulatory processes (Coleman & Shore,1991; Jaušovec, 1994, Pelletier & Shore, in press).

Meanwhile, studies of problem solving in ill-structured domains such as law (Cole & Kuhlthau, 2000), medicine, philosophy, psychology, education, and art, pose a new challenge because the methods and rules applied in well-structured domains do not easily apply here. What is agreed upon is the notion that both quantitative abilities such as speed of recall and qualitative abilities such as forward reasoning, characterize the difference between novices and experts. Shore and Kanevsky (1993) examined the thinking processes of three groups, children with high IQs, children who do very well in school, and adult experts. In exploring their differences, the authors argued that there is a need to shift from ability-oriented research to research that emphasizes processes in order to understand expertise. As a result, they extracted seven principal characteristics that differentiate the thinking processes of experts from novices and gifted vs. non-gifted: memory and the knowledge base, self-regulatory processes, speed of thinking process, problem representation and categorization, procedural knowledge, flexibility, and preference for complexity.

When it came to experienced searchers, Bates (1979) admitted early on, "We know discouragingly little about just what those skills are and how they develop" (p. 205). She introduced the concept of search tactic, any move that is made to search a topic further. She grouped 29 such tactics into four categories: monitoring, file structure, search formulation, and term. She also defined the concepts of strategy and tactic in a military manner, with strategy referring to long-term, large-scale planning, and tactic to short-term goals and maneuvers. Later (Bates, 1987), she focused more on how to use search tactics on-line, at the beginning of a search or at the end, and finally (Bates, 1989) she introduced the model of "berry picking" for on-line search that will be discussed later.

As on-line resources expanded, classical models for information search came under attack. These models were based on a perfect match between the need of the information seeker written in "query" language, and a document representation written in controlled vocabulary. If there was no match, it was the information seeker who had to revise the query to make it understandable by the system. Soon, more and more researchers began to ask for a more flexible and user-centered system that would allow them to solve their problems quickly and effectively.

Librarians as expert information-seekers have developed substantial knowledge related to the factors of information seeking. Some of these factors are: (a) knowledge of how various domains are organized, (b) individual cognitive and perceptual characteristics, personal experience in making inferences and evaluating the appropriateness of certain types of reasoning, (c) spatial visualization, (d) superior memory and visual scanning abilities, (e) recognition of different types of problems that drive searching tasks, and (f) awareness that an information-seeking task that drives the information-seeking process, may change as the search progresses (Marchionini, 1995, pp. 68-69).

Notwithstanding the importance of these factors, they have been based on on-line bibliographic database searching, and this might differ somewhat from seeking information on the Web.

In an attempt to understand information seeking on the Web, Tabatabai and Luconi (1998) compared three experts and three novices, all graduate students of the Faculty of Education at McGill University. Their task was to search for inquiry-based instruction in order to write a research paper. The criterion that differentiated experts from novices was the hours they spent on the Web. Experts, on average, spent 15 hours a week on the Web, novices spent 3 hours. Performance measures were the differences between how each group searched the Web. Cognitive task analysis and verbal protocols were used to gather data. Differences between the experts and novices were analyzed in terms of (a) knowledge base, (b) problem space, (c) strategies, and (d) affect. Experts used more key words and evaluated the sites based on established criteria (wider knowledge base). They also had a different problem space. Novices used the Back key more often, used fewer numbers of search engines, and missed some highly relevant sites. In terms of strategies, experts were more goal-driven and used more metacognitive strategies (e.g., monitoring) than novices who mostly relied on trial-and-error. Also, affect played a much bigger role for novices than experts. Novices felt more nervous, tired, frustrated, lost, and overwhelmed by the amount of information available on the Web. Shortcomings of the study are its small sample size (three novices, three experts), and lack of specific criteria on how "experts" were chosen.

Lazonder, Biemans, and Wopereis (2000) observed 25 grade four students with various levels of Web experience while performing three tasks on the Web. The research question was whether the Web experience had any effect on two components of the search process, namely locating a Web site and subsequently locating information on that site. Performance measures were based on success, time, efficiency, and effectiveness. Log files and questionnaire results were used to analyze the data. For the first part of the task, i.e., locating the site, the difference between experts and novices was statistically significant in the following ways: experts were faster, completed more tasks, and were more efficient and effective. For the second part of the task, i.e., locating information within the site, there were no significant differences between the groups. These findings suggest that the differences found in hypertext searching may be due to locating the sites and not the information in the sites. Therefore, teaching how to use search engines to locate the wanted sites is likely to improve performance. Although the study showed that novices differ from experts, it provided little insight into the kinds of search strategies that contributed to each group's success. More qualitative in-depth analysis would increase our understanding of user information-seeking habits on the Web.

Palmquist and Kim (2000) also found that on-line database-search experience improved performance of the participants in their study. Forty-eight undergraduate students searched their university's Web page to perform two tasks. First, they had to find a specified site and bookmark it. Next, they had to find relevant pages in a site and bookmark them. There were two independent variables, cognitive style and prior experience. Cognitive style was identified by the *Group Embedded Figures Test* and prior search experience was determined by questionnaire. Performance measures were the average length of time spent on retrieving the information and the average number of nodes visited. Field-dependent participants spent more time retrieving information than field-independent participants. Although the paper went into a detailed discussion of the difference between field-dependent and field-independent participants, it is all based on searching, not on the Web, but just in the university's home page. Again, as with other studies, there were no qualitative analyses and it is difficult to extend the results to the search on the Web at large.

Research in other media such as CD-ROMs (Large, Beheshti, & Breuleux, 1998; Breuleux, Renaud, Large, & Beheshti, 1993) shared some qualities with research on the Web. For example, Breuleux, et al. (1993) found that children, who had more alternative representations in the form of graphics or animation, were more distracted and recalled less information than those who had received only text. Similar to the Web, the availability of multiple sources of information by itself did not lead to better acquisition of knowledge and its complexity added to the level of confusion, much as for the claimed potential of hypermedia (Dillon & Gabbard, 1998). However, participants in the aforementioned studies were children and although the results are related to the findings of research involving adult participants, it is at times harder to compare these two groups. Large and Beheshti (2000) agree that when it comes to information-retrieval systems, children are distinct due to their lack of language sophistication, lack of ability to manipulate search strategies, and their cognitive overload that results in disorientation (p. 1070).

There is obviously a shift in information-seeking research from the system-centered to the user-centered perspective. The above studies pay more attention to the process and the users' characteristics than to the technology itself, recognizing that users engage in information seeking as a dynamic problem-solving activity in which their idiosyncratic attributes affect the process of search and the outcome. Novices applied naïve strategies, relying on trial-and-error most of the time, and had unreasonable expectations from the system due to their misconceptions about technology. Experts, on the other hand, found information faster and more efficiently. A review of the users' characteristics (in the form of varied models) that affect information-seeking outcomes is presented in the following section.

Models of Information Seeking

With the shift of focus from system to individual user needs, the effect of psychological and cognitive processes on the information-seeking process and how to model these effects became a viable research topic.

In a model called the Anomalous State of Knowledge (ASK), Belkin, Oddy, and Brooks (1982), described the constructive process of information seeking in terms of knowledge abnormality in the users' minds. An information need, based on this model, is the discrepancy between what the user already knows and what he or she needs to know. ASK is dynamic and changes as the search goes on. Thoughts evolve from being unclear, vague, and uncertain, to clear and more focused as users progress from an anomalous state of knowledge to understanding their topics and solutions.

Bates (1989) described a model of information retrieval called "berry picking" that differs from classical models in four areas: flexible query, flexible search process, wider range of search techniques used, and wider information territory in which the search is conducted. Similar to picking berries in the field, information is accumulated "a bit-at-a-time" (p. 410) instead of in a single best set. Her model is one of the first models that emphasized user-centered search and initiated a discussion on the similarities and differences between browsing and searching.

Kuhlthau (1993) described a more holistic view of information seeking, encompassing the classic triad of thoughts, actions, and feelings or cognition, physical activities, and affect. Her Information Search Process (ISP) model has six stages composed of high level goals: task initiation, topic selection, prefocus exploration, focus formulation, information collection, search closure, and starting writing (p. 42). In her qualitative methodology she used journals, search logs, short written statements, case studies, concept maps, teachers' assessments, and a perceptions questionnaire. Her methodological principles were impressively vast and provided important knowledge from the point of view of the individual's experiences. At the same time, ISP is a process model that can be used to differentiate users based on how they set information-seeking goals.

The Personal Information Infrastructure (PII) model proposed by Marchionini (1995) is a collection of mental models that includes "abilities, experience, and resources to gather, use, and communicate information" (p. 11). These mental models interact for specific information systems. His often-cited model of the information-seeking process, includes eight sub-processes that can develop in parallel: (a) recognize and accept an information problem, (b) define and understand the problem, (c) choose a search system, (d) formulate a query, (e) execute the search, (f) examine the results, (g) extract information, and (h) reflect/iterate/stop (pp. 49-60). Although Marchionini did not make a direct connection between PII and the eight sub-processes, he considered information seeking to be dynamic and the information seeker to be in control of these sub-processes. Marchionini's PII model can be used as a framework in studying information seeking, and it is not intended to be used as a stand-alone model like Kuhlthaus's model.

The first mathematical model of surfing on the Web was developed by Huberman, Pirolli, Pitkow, and Lukose (1998). They looked at patterns of Web surfing by developing a mathematical model based on the laws of probability. The probability of a user going from site to site depends on the value of the current site versus the anticipated value of the next site. The user stops Web surfing when the cost of moving to the next site is more than its expected value. They analyzed data from a representative sample of America Online (AOL) Web users for five consecutive days. This is the most staggering amount of data that was relatively easily collected and analyzed so far. For example they reported that on one day (December 5, 1997) 23,692 users accessed more than one million Web sites. Their results validated the model they were proposing. Users show strong regularities in their search for information, and these regularities can be statistically described. Although this sophisticated model does not directly address users' strategies, its use of a mathematical model establishes a firm foundation for further research.

Sutcliffe and Ennis (1998) acknowledged the need for an integrated theory of information retrieval and proposed a cognitive framework for modeling information searching called Process Model. Contrary to most models of information retrieval, this model uses only two agents: the user and the information search-support system (leaving the expert intermediary out). The purpose of the model is to display expert user behavior in four major activities: problem-identification, need-articulation, queries-formulation, and results-evaluation. They provided a comprehensive and, at times, complicated set of rules that govern user behavior. This was the first model that sought to predict user behavior in different task stages, based on the users' needs and their knowledge on one hand, and the system facilities on the other. The actual behavior of users cannot be accurately predicted. Yet, this model can be best used in developing a tutoring or intelligent help system. It also has potential as a decision-support system or as an adaptable user interface.

Another model was proposed by Choo, Detlor, and Turnbull (1999), who integrated browsing and searching. This model was a combination of two other models: scanning (Aguilar, 1967) and information seeking (Ellis, 1989). Thirty-four knowledge workers from seven companies volunteered for this study. A Web Tracker application was installed on their computers at work to register all their moves for two weeks. They were also interviewed twice and recalled critical incidents of using information from the Web. The model was based on a plot of 61 critical-incident episodes. On one axis, episodes were plotted according to the model's four modes of scanning: undirected viewing, conditioned viewing, informal search, and formal search. On the other axis, episodes were plotted according to six categories of information seeking: starting, chaining, browsing, differentiating, monitoring, and extracting. The highest number of episodes (23) was informal search (e.g., company names, products), 18 episodes fell into the category of conditioned viewing (e.g., bookmarked sites), 12 were undirected viewing (e.g., general news Websites), and the least number of episodes (8) were categorized as formal or intentional searches seeking specific information. The study suggested that investigating strategies and tactics of users might be helpful in analyzing information seeking on the Web. The study also used both quantitative and qualitative methods of data collection that gave the results more validity.

Summary

Research in ill-defined areas including information seeking on the Web is augmenting research in traditional systems such as libraries. The few studies that focused on users' characteristics provide a starting point for further research. However, experts and novices in these studies were defined *a priori* based either on their prior performance or on some measure of ability or talent. Novices' attributes were not defined very well except that they were different from experts. Furthermore, expertise in these studies was portrayed statically--without examining the process of becoming an expert.

Information seeking processes need to be explored to reveal differences between experts' and novices while searching on the Web. There is also a need to define expertise in a new manner, i.e., *a posteriori*, in relation to the strategies and attributes that have a positive effect on the outcome of the search. Finally, expertise in information seeking should be modeled to isolate strategies and attributes, as well as other relevant factors that have a combined effect on the success of the search.

Purpose and Research Questions

The primary purpose of this study is to advance our understanding of expertise in information seeking on the Web by identifying a set of (or a systematic account of) strategies and attributes that will increase the chance of a successful search on the Web. The strategies are defined as any maneuver that will help search a topic further: Evaluation, Navigation, Affect, Metacognition, Cognition, and Prior knowledge. Attributes are varied data on the participants and include Age, Sex, Years of experience, Computer knowledge, and Information-seeking knowledge. (Full definitions are provided in Appendix A, p. 85). Success is defined as finding a definition of Expository Teaching Model within 30 minutes and it is assumed that, on average, experts will find this information faster than non-experts (e.g., novices). In this study, the degree of expertise will be assumed to be related to the employed strategies and attributes and the time it takes to find the required information. Expertise will therefore be defined empirically in relation to the outcome. This is a quasi-experimental study because the selection and assignment of participants is not randomized, there is no control group, and independent variables, (i.e., strategies and attributes) cannot be manipulated by the researcher. The outcome (dependent) variable considered in this question is the timely success of the search. Participants will be from three groups of novices, intermediates, and experts.

The secondary purpose of the study is to develop and determine the effectiveness of a model to predict what strategies and attributes will increase users' chances of a successful search.

This study hopes to answer the following research questions:

1. Are there differences in the patterns of search among users with varying degrees of expertise? Do expert searchers use different strategies than intermediates and novices?

2. Is there a relationship between users' strategies and attributes and timely success of the search on the Web?

3. Can a successful search on the Web be modeled?

Hypotheses

1. Expert searchers use different strategies than intermediates and novices. Experts would have higher prior knowledge and would rely more on Metacognitive strategies; in contrast, novices would mostly rely on trial-and-error.

2. Possession of prior knowledge and use of certain strategies such as Metacognition are associated with success.

3. There are certain strategies and attributes that affect the outcome of the search and these factors can be modeled using survival analysis.

Chapter 2: Method

Participants

In order to detect and account for the effect of individual attributes and strategies on the search outcomes, a total of 29 participants were observed in three groups of novices, intermediates, and experts. Demographic information about these groups is presented in Table 1.

Table 1

A Comparison of the Three Groups of Participants

Attributes	Participants		
	Novices	Intermediates	Experts
	Mean (SD)	*Mean (SD)*	*Mean (SD)*
Gender (F, M)	7F, 3M	7F, 2M	8F, 2M
Age	24 (7.50)	31 (9.00)	44 (8.10)
Yrs of University	1.5 (1.01)	8.1 (5.73)	5.3 (2.06)
Yrs of Experience	5.5 (4.76)	6.9 (5.7)	17.3 (9.14)
Yrs on This Job	1.7 (1.03)	2.3 (1.17)	10.4 (6.50)
Computer Literacy	2.6 (1.17)	3.8 (.44)	3.7 (.82)
Info-Seeking Know.	2.8 (0.92)	4.2 (0.67)	4.2 (1.03)
Web Knowledge	1.7 (1.06)	3.4 (0.53)	3 (1.56)
Hrs/day on Computer	1.2 (0.47)	4.2 (1.30)	6.5 (2.22)
Hrs/day on Web	1.3 (1.76)	1.39 (.49)	3.1 (1.38)

Novices were 10 undergraduate pre-service teachers from the Faculty of Education at McGill University. They were chosen because they were trained in

Pedagogy but not specifically in information seeking. Intermediates were nine final-year master's students in the Graduate School of Library and Information Studies (GSLIS), also part of the Faculty of Education at McGill University. These students were chosen as intermediates because they had received training on how to search the Web as librarians, but typically had not put their knowledge into extensive practice. Experts were 10 highly experienced professional librarians working in a variety of settings including government, industry, and university, and whose institutions provided internship opportunities for Library and Information Studies graduate students.

Experts were older than novices with more years of university, more job experience, and more computer knowledge. They rated themselves higher on information-seeking and Web knowledge, spent more hours a day in front of a computer, and spent more hours a day searching the Web than did novices. Intermediates were selected to fill the gap between these two groups but many times they resembled experts more than they resembled novices. For example, one intermediate and one expert had five and eight years of teaching experience, respectively.

There were some differences between the participants with regard to how they learned to use the Web. All participants, except for two experts, reported that personal interest primarily motivated them to learn the Web, alone or in combination with formal or job training. Expert 7 reported learning the Web solely as a requirement on the job, and Expert 9 learned it through a combination of formal training in a course and on-the-job learning. Personal interest played a more important role for novices as eight of them and only one expert reported it as the prime motive for learning how to use the Web. A combination of personal interest and partial requirement for a course helped two novices, seven intermediates, and one expert learn the Web. Five experts reported that they learned through a combination of personal interest and training on the job. Only two intermediates and one expert used all three (i.e., personal interest, formal training in a course, and on-the-job training) to learn how to use the Web. Formal training, whether part of a course or a requirement on the job, played an important role in how participants learned to use the Web; all intermediates and all experts except one reported being trained formally, whereas most of novices did not.

Novices and intermediates similarly reported that their top activity on the Web was e-mail. One expert reported using e-mail but only as the third activity. Experts used the Web mainly for research; novices and intermediates used it for personal reasons such as communication, entertainment, travel planning, or music.

Novices and intermediates were first approached in their university classrooms where the researcher explained the nature of the study and asked for volunteers to write their names, phone numbers, and e-mail on a piece of paper so they could be contacted later. Experts were individually contacted from a list of names provided by a faculty member of the GSLIS. As a gesture of appreciation for their time, one participant from each group, by the luck of a draw, was given a mug with the University logo.

This study received Certificate of Ethical Acceptability for Research Involving Humans from the Faculty of Education, McGill University (Appendix B).

Materials

The materials consisted of consent forms, instructions to participants, a background-information questionnaire, researcher's notes, and Internet temporary files. The equipment used to conduct the study consisted of a laptop computer, VCR, portable audio system with microphone, telephone with extension cords, timer, Dragon Naturally Speaking software, and a transcriber (Appendix C).

All participants signed a consent form (Appendix D) acknowledging that they have been informed about the purpose of the research, that they could withdraw at any time, how long the session would take, what procedures were involved, and how their anonymity would be guaranteed.

The background-information questionnaire (Appendix E) collected data about name, age, gender, years of university, years of experience, years of teaching, how they learned to use the Web, how many hours a day they spent in front of a computer and using the Web, their favorite search engine, their prime use of the computer, which part of a Web site they read first, and what criteria they use to evaluate a Web site. They were also asked to rate themselves on three areas of computer literacy: information-seeking knowledge, computer knowledge, and knowledge of the Web (ranging from 1 = poor to 5 = excellent).

The task (see below) was typed and handed to each participant. After reading it, any misconceptions were clarified. The Web browser, Microsoft Internet Explorer, created Internet temporary files automatically. At the end of each session, these files were printed and used later on to fill in any missing data.

Procedure

Novices and intermediates came to the researcher's lab to participate in the study. Experts performed the task in their work office to minimize the demand on their busy schedule. Participants were observed individually. Everyone started from McGill's Faculty of Education Home Page and used Internet Explorer version 4 that was setup before they arrived.

The participants were given the consent forms to sign and background-information forms to fill out while the laptop, VCR, and audio recorder were started. The researcher asked the participants to read the task and then explained what was expected of them. The instructions to the participants were the following:

> Your task is to find as much information as you can on a definition of "Expository Teaching Model" on the Web during the next half hour. The primary purpose of the task is to find a definition of the model. If you find a definition and have more time left, you may try to find out who initiated the model.

Special attention was paid to explaining and stressing the importance of talking aloud in collecting verbal protocols as data. To make sure they understood how to talk aloud, the researcher did two exercises with them. First they were asked to count aloud the number of windows in their place and then to describe how they tied their shoelaces (Ericsson &

Simon, 1993). Any misconceptions were corrected before going on to the next stage. Then the researcher started the laptop, the VCR, and the audiotape.

Before continuing, participants were asked if they had any questions or needs and were told that, during the next half hour, the researcher would be quiet except for prompts like "keep talking." The researcher wrote down the start time and told the participants to begin the search. While they were searching, the researcher wrote down any observations and questions to ask them at the end. The time, at which the participant found the first definition of Expository Teaching Model, if it was found, was recorded. After the task was completed, participants were interviewed about various topics related to Web searching (Appendix F). Their responses were triangulated with the results of the questionnaires and the verbal protocols to increase reliability of the data. Member checks for all participants were sent before the data analysis; no corrections were requested. To assure their anonymity, participants were numbered. For example 020712 meant the session was on February 7th at 12 o'clock. After each session, the history and favorites of the participant were saved in a separate file.

Data Overview

The method for data collection in this study was verbal protocol; it allowed for in-depth detailed data collection from the participants while they were searching the Web for a definition of Expository Teaching Model. Other sources of data were Internet temporary files, background questionnaires, and post-task interviews. Following data collection, a template was created (Table 2) that tracked the processing and analysis of the data. First, the audiotapes were transcribed verbatim into a text file for each participant. Second, after the text files were created, each file was segmented into clauses and entered in an Excel file where each line was a clause ready to be coded. The codes were directly entered into the Excel file, creating 29 files. Third, the column with the codes from each Excel file was copied into a separate file. This file included the complete data and included all the codes from 29 participants in the order that they appeared in the text. Finally, the codes were rank ordered and, using Microsoft Excel PHStat©, a one-way frequency table for all the codes was created individually for each participant. These codes were manually entered in the Table of all codes (Appendix G). This table represented the total processed data ready for analysis.

Table 2

Template for Transcribing the Audiotape

ID: 020712	
Date/Sources of data:	
1st Source – Verbal Protocols, audiotapes:	Date done
Transcription using Dragon Naturally Speaking	
Second listening for editing	
Watch the videos to insert commands and missing actions using the following format: Search Engines, (Key words), COMMANDS, http://NameOfSitesVisited.	
2nd Source - Internet files:	Date done
Insertion of URLs and names of sites (and the description as it appears under the name if possible) from Internet Temporary Files into a separate file called List of URLs.	
3rd source - my notes and observations	
Insert any other explanation deemed necessary for understanding the actions in [ ]	
4th Source - Questionnaires	
When did they find the answers? (Time in minutes)	
Answered question 1 (definition of etm)	Time
Answered question 2 (Ausubel)	Time
Coding:	
Document ready for coding:	Date
Coding done:	Date

Coding System

The coding started as model-based coding (Ericsson & Simon, 1993) that was first developed by the researcher and one of her colleagues (Tabatabai & Luconi, 1998) for a study that investigated expert-novice differences in searching on the Web. It included 66 codes and the reliability as measured by percentage of agreement was 90% when the authors first applied these codes. The coding for the present study evolved as the researcher gained familiarity with the literature and with the data. At the end, it included six strategies (plus two categories of Vague and Not Relevant), all identified by a one-digit prefix; 25 tactics identified by a two-digit prefix; and 127 moves identified by a three-digit prefix. A complete list of codes and their definitions are presented in Table 3. A *strategy* is defined as a tool that a participant uses in order to solve a problem. It usually involves large-scale broad maneuvers that are a combination of several sub-processes or tactics. *Tactics* are short-term sub-processes that are a collection of several moves. *Moves* are fine-grain actions, feelings, knowledge that are consciously selected, applied, and monitored to solve a problem (Marchionini, 1995; Bates, 1987). The transcribed texts of participants' utterances are coded based on their Moves. Tactics and strategies are then derived from these codes.

Table 3

Complete List of Codes Used to Code Verbal Protocols

1. Evaluation

1.1 Criteria to evaluate sources (currency, authorship, accuracy, etc..)

1.1.1	Evaluates sites but criteria are tacit therefore we use cues to infer them
1.1.2	Clearly uses criteria to evaluate sites
1.1.3	Uses surface criterion i.e., how a site looks
1.1.4	Other (can't evaluate, not enough info)

1.2 Evaluation of one source - access/open a site

Negative evaluation

1.2.0 Based on its URL (Unified Resource Locator)

1.2.1 Based on its title, headers

1.2.2 Based on its summary/content (descriptions or abstracts)

1.2.3 Based on the organization of the document/design of the page/absence of link/hypertext/font

1.2.4 Other (personal page, lack of criteria in the site)

Positive evaluation

1.2.5 Based on its URL (Unified Resource Locator)

1.2.6 Based on its title, headers

1.2.7 Based on its summary/content (descriptions or abstracts)

1.2.8 Based on the organization of the document/design of the web page/presence of link/hypertext

1.2.9 Other (appear more than once, seen it before)

1.3 Evaluation of other sources (i.e., Search Engines, Web, McGill home page, work, this experiment, or this computer)

1.3.0 Negative evaluation

1.3.1 Positive evaluation

1.3.2 Neutral

1.4 Evaluation of search engines

1.4.0 Negative evaluation

1.4.1 Positive

1.4.2 Comparison of search engines

1.5 Evaluation of immediate results or hits

1.5.0 Generally negative

1.5.1 Generally positive

1.5.2 Not sure

1.5.3 Neutral, talking factually about results or hits (I've got so many # of hits)

1.6 Evaluation of final results (overall)

1.6.0 Generally negative

1.6.1 Generally positive

1.7 Wrong Evaluation – Missed opportunity

1.7.0 Missed opportunity (not notice a good site OR stay in a bad one that he has seen and judged before)

2. Navigation

2.1 Moves

2.1.0 Backward moves (one step or more)

2.1.1 Access a search engine for the first time

2.1.2 Access a search engine for the second time and more

2.1.3 Scroll, skim, glance (not counted the sources)

2.1.4 Add to favorites, or Book marking (shared with 5.4.1)

2.1.5 Skip a site previously seen

2.1.6 Close navigation window (copy/paste into file 5.4.1)

2.1.7 Access to a site (click)

2.1.8 Click to access other than a site, i.e., Word, Thesaurus in Word, web page matches

2.1.9 NEXT page

2.1.10 Using MSN as it appears on the left side of screen

2.2 Typing the search term

2.2.0 Type the search term as is (Expository Teaching Model)

2.2.1 Type the search term truncated (expository teaching)

2.2.2 Type the search term truncated (teaching model)

2.2.3 Type the search term truncated (expository)

2.2.4 Type the search term truncated (teaching)

2.2.5 Type the search term truncated (model)

2.2.6 Typing mistakes

2.2.7 Typing "etm"

2.3 Typing other words

2.3.0 Add a new key word to the search term or add + sign

2.3.1 Type a new word or term (without any of the search terms)

2.3.2 Typing mistakes and other mistakes

2.3.3 Use synonyms

2.3.4 Use quotations

2.3.5 Use SEARCH key or looking for it

2.3.6 Use of ADVANCE search key

2.3.7 Further moves to refine search, narrow it down

2.3.8 Submit

2.3.9 Other (type MSN, use of STOP, HOME keys)

2.4 Waiting any type of activity/comment (except feelings: 3.3) while the user waits for a response

2.4.0 Decides not to wait because it took too long for the system to respond

2.4.1 Waits

2.4.2 Comment as to slowness

2.4.3 Any other comment as to status quo (I'm waiting, or I'm at)

2.4.4 Any action while they wait/opening another search engine/ switching between them

3. Affect

3.1 Personal feelings expressed towards oneself during the search (I'm no good)

1.1.0 Negative feelings towards oneself

1.1.1 Positive

1.1.2 Neutral, not sure how it feels

1.1.3 Change of mind

3.2 Expectations from the search or the search results (this link is going nowhere)

3.2.0 Negative, unfulfilled

3.2.1 Positive, a hope or projection

3.2.2 Neutral (there it is, I'm at)

3.2.3 Ambivalent (I guess this is library)

3.3 Overall perception of how the task is going (Labyrinths)

3.3.0 Negative: impatient, frustrated, can't solve problem

3 3.1 Positive: patient, relaxed, focused, curious

3.3.2 Neutral

3.3.3 Lost (don't know . . . no idea)

3.4 Information overload

3.4.0 Expression of information overload, too many hits

<u>4. Metacognitive strategies</u>

4.1 Reflections Only if they verbalize it themselves otherwise not a reflection.

Reflecting on themselves

4.1.0 Negatively on their own knowledge, ability, and interest

4.1.1 Positively on their own knowledge, ability, and interests

Reflecting on the search process (and meaning of etm)

4.1.2 Negatively (that is not a good definition)

4.1.3 Positively

4.1.4 Neutral

4.1.5 Not sure

4.1.6 Difficult process

4.1.7 Wrong inferences (from one concept to the other)

4.2 Monitoring Finger on pulse of how the search is generally progressing

4.2.0 Monitoring on how they are doing (talking to themselves)

4.2.1 Monitoring on how the search is going (double check with task)

4.2.2 Ask for how much time is left

5. Cognitive Strategies

5.1 Search strategies - Tools and rationale used to start and carry out the search

5.1.0 Rely only on memory

5.1.1 Using pencil and paper

5.1.2 Printing for later perusal

5.1.3 Rational/comments for chosen strategies

5.1.4 Change in strategy/Search

5.1.5 Using Ctrl F

5.1.6 Other (cut URL, Feeling Lucky option)

5.1.7 Look for something different

5.1.8 Opening multiple search screens

5.2 Reading priorities/habit

5.2.0 Reading Title of hits

5.2.1 Reading URL of hits.

5.2.2 Reading summary of hits (under URL)

5.2.3 Reading the results

5.2.4 Scanning/Reading results pages

5.2.5 Lazy in reading (too little time to read)

5.2.6 User is inside the site and reads the content thoroughly

5.2.7 User is inside the site but scans the content and reads quickly

5.2.8 Reading the first 10 hits or so

5.2.9 Long pause, read silently

5.3 Planning Includes intentions.

5.3.0 No plan (trial and error)

5.3.1 Expression of an intention to do something that is not covered by the following categories

5.3.2 To navigate

5.3.3 To evaluate

5.3.4 To continue search after the present task (go to library for personal curiosity)

5.3.5 Other (don't want to do something, don't use pen)

5.4 Keep track of search - when find a good source, what do you do with it?

5.4.0 Copy/paste download in a text file

5.4.1 Book marking or add to favorite

5.4.2 Write down words, synonyms, steps of task, sources, steps of the search

5.4.3 Creating links between what they find on the Web and the topic of the search

5.4.4 Print for later perusal

5 .5 Tangent – whether actually goes off track or just talk about it

5.5.0 Distractions

5.5.1 Curiosity stop

5.5.2 Copy/paste/e-mail/bookmark of sources of interest to them

5.5.3 Discussion about time to make for curiosity stop, to quit and time they'll spend in each cite

5.6 Metaphor Use of metaphorical knowledge

5.6.0 Use of metaphors

<u>6. Prior knowledge</u>

6.1 Domain– Directly or indirectly related to topic

6.1.0 Lack of domain knowledge

6.1.1 Assumptions and biases about the domain

6.1.2 Possessing some domain knowledge

6.2 System - (dragging mouse, knowledge. about characteristics of search engines) recognition of features

6.2.0 Lack of system knowledge

6.2.1 Biases about system, i.e., the Web is all-commercial, a site sponsored by a computer co. is not reliable

6.2.2 Possessing some system knowledge

6.2.3 Other, i.e., knowledge of Word

7. Vague Statements that are not clear due to either low voice or poor quality of tape. This code will be excluded from the analysis.

8. Not Relevant Any small talk that is not related to the task such as talks about the weather or ride in the Metro.

The coded data created in this study were extensive. They could be visualized as a matrix of 29 rows (number of participants) by 160 columns (number of codes used), more than 4500 cells. An attempt to perform a Pearson correlation resulted in more than 20,000 cells. To organize these data for a more manageable analysis, the data were divided into three levels from a very specific three-digit level (moves) to the most aggregated one-digit level (strategies) as follows:

Level 3 data (moves) were the actual codes applied to the transcribed text, for example, 1.1.3 (Surface Evaluation), or 2.1.0 (Backward Move).

Level 2 data (tactics) were derived by adding all related level 3 data, for example, 1.1 (Evaluation Criteria) was a combination of four level 3 data ranging from 1.1.1 (Tacit Criteria) to 1.1.4 (Other Evaluations).

Level 1 data (strategies) were derived by adding all related level 2 data, for example, level 1 (Evaluation) included seven level 2 data ranging from 1.1 (Evaluation Criteria) to 1.7 (Wrong Evaluation). Level 1 data were the most aggregated and included the six strategies and attributes. The main analysis in this thesis was done with level 1 data, drawing explanations from levels 2 and 3 whenever warranted.

Data Analysis

In order to answer the first research question (Are there differences in the patterns of search among users with varying degrees of expertise?), primarily descriptive analysis was used. Analysis of variance was also performed on level 1 data to look at group differences in strategies and attributes. Pearson correlation coefficients were calculated on all three levels of data.

In order to answer the second research question (What is the relationship between users' strategies and attributes and timely success in the search?), Cox survival analysis (1972) was performed on all levels of data. The outcome (dependent) variable considered in this question was the timely success of the search and the explanatory (independent) variables were strategies and attributes.

In order to answer the third research question (Can successful search on the Web be modeled?), backward stepwise Cox survival analysis was used. The dependent variable was the same as in question two; the independent variables were similar except that in this question they were all put together in the model, whereas in question two their effect was investigated individually.

Survival analysis. A number of studies have used regression or other methods such as analysis of variance to examine the reasons behind success or failure of searches. However, as Willet and Singer (1991) have pointed out, there are some methodological concerns. One is about the nature of the research question asked. Researchers have traditionally asked "whether or not" an outcome happens and although that is an important question, Willet and Singer argued that we should also ask "when" the outcome happens. Statistical techniques such as regression analysis or analysis of variance cannot address the "when" question because they are not capable of handling "censored" data. Censored data (e.g., in survival analysis) are those observations that do not experience the event of interest during the period of data collection. For example, in this study, participants who did not find the required information within the half hour were considered as censored data.

The data collected in this study fell into the category of survival-time data, which have also been called lifetime or failure time data (Lawless, 1982). This is the type of data in which the variable of interest is the time to occurrence of an event, in this case, data in which the variable of interest is the time to occurrence of an event, in this case, the time to finding a definition of Expository Teaching Model (ETM). In other words, survival analysis is useful whenever we ask research questions about *whether* an event occurs, and also *when* it occurs. Statistical analysis of these types of data is referred to as survival analysis. The use of the word survival refers to its origin in the health domain in which the success of a treatment is measured in terms of how long patients survive after the treatment. In this study, survivors are those participants who have not found the definition after 30 minutes.

In survival analysis, the time until an event happens is called the survival time, t, with a probability density function denoted by $f(t)$. Survival time varies from zero to infinity. In this study, the occurrence of the event (finding ETM) was observed within the first 30 minutes. Survival probability is the proportion of those individuals who have survived the event, i.e., participants who have not found the definition. A distribution of these probabilities or proportions over time forms the survival function, $S(t)$. A graph of $S(t)$ versus t is called a survival curve. At time zero, when participants start to search, they are all surviving (no one has yet found the definition), and the value of the survival function is 1.00. As the search process progresses, some find the definition and the value of the survival function drops. Since not every participant finds the definition before the half hour is over (censored data), the curve never reaches zero (Willet & Singer, 1991). By definition, all survivor functions have a similar shape dropping from a value of 1 to zero over time, but their rate of decline, obviously, is different.

The hazard function $h(t)$ is defined as the ratio of $f(t)$ over $S(t)$ and reflects the instantaneous risk that the event (finding the correct definition) occur at exactly time t. While survival probability can be compared to a glass that is partially empty (how many participants did not find the definition), hazard probability can be compared to a glass as partially full (how many found the definition). Due to the fact that we would like the event to happen (participants find the definition), it will be easier to explain the event in terms of the hazard function.

Hazard is the reciprocal probability of survival. When there are no survivors, the hazard is fully realized. At the outset, when everyone is a survivor (i.e., no one has yet found the definition), no hazard has yet been realized. The hazard function depends, in general, on both time and a set of covariates. The proportional hazards model (Cox, 1972) used in this study separates these components by specifying that the hazard at time *t* for an individual whose covariate vector is *x* can be written as:

$$h(t;x) = h_0(t) \exp(\beta x)$$

In this model the impact of an explanatory variable (predictor) *x* is exponentially related to the baseline hazard function (h_0). In other words, hazard at time *t*, for a given value of *x* equals baseline hazard times exponential of βx, where β is the regression coefficients for *x*. The baseline hazard function is the hazard function for an individual for whom all values of explanatory variables take the value of zero. The higher the value of exponential β, the higher is the hazard and the more chance of finding ETM.

In this study we were primarily interested in determining whether the distribution or patterns of the dependent variable (response times) in one group differed from that of another. Second, can this difference be modeled? It is important to note that, as a requirement for the model, it was necessary to clearly define the criteria that specify the occurrence of the outcome of interest, in this case the exact time that the participant found the definition. Due to the participants' differing strategies and attributes, some will find it sooner than others. The criterion in this study was the time that the first definition of Expository Teaching Model was retrieved. Specifically, it was important for the participant to notice at least one aspect of the definition. For example, Intermediate 2 met the criteria when she said, "Expository really promotes using lectures as a style of teaching" and Novice 7 said, "I think we struck gold, this method can teach and address [a] large number of the students, ignoring individual differences."

Chapter 3: Results and Discussion

The results are presented in three sections, addressing each of the three research questions in turn. The first section addresses the question of differences in the patterns of search among users with varying degrees of expertise. Based on in-depth analyses of verbal protocols from participants in three groups, novices, intermediates, and experts, this section attempts to draw a picture of a typical information seeker from each group. This question is explored primarily descriptively but at times the results of quantitative analyses are interjected to strengthen the points.

The second section addresses the relationship between users' strategies and timely success of the search. Based on results of performing survival analysis on the data, this section attempts to identify the strategies and attributes that have an effect on timely success of the search. Timely success was defined as finding a definition of "Expository Teaching Model (ETM)" in less than 30 minutes, assuming that some will find it sooner and some will find it later. In this section, level of expertise is defined, not by *a priori* categories of novice, intermediate, and expert, but by participants' choices of strategies that helped them find ETM more quickly than others. The strategies and attributes were those that were associated with timely success, regardless of the group, and were determined after examination of the data.

The third section proposes a predictive model of success based on the effect of those strategies and attributes that were identified in the previous sections.

To measure the reliability of coding, three profiles were chosen at random, one from each group. Despite the difficulty for the second rater to code without having listened to the tape, and unaware of voice connotation, the percentage of agreement was 80% for level 3 and 85% for level 1 data.

Does the Level of Expertise Impact on How People Search?

The first research question addresses the possibility of differences in search patterns or processes among users with varying degrees of expertise in Web searching. Although expertise can be represented across a continuum, the participants were categorized into three broad categories and a variable called Group was created with three values: 1 = Novice, 2 = Intermediate, and 3 = Expert. The data gathered from these groups respond to the question at the most aggregated level (level 1). It is sufficient to look at the six strategies to find differences; therefore the principal results are expressed in terms of level 1 data. Level 2 and level 3 data will be used to illustrate and elaborate on the meaning of level 1 data. For a comparison of these six strategies, see Table 4. Numbers are total frequencies of codes used for each strategy.

Table 4

A Comparison of Three Groups Broken Down by Strategies

Participant	Evaluation	Navigation	Affect	Meta-cognitive	Cognitive	Prior Knowledge
Novice	176 (10%)	632 (35%)	451 (25%)	76 (4%)	441 (25%)	19 (1%)
Intermediate	156 (9%)	534 (30%)	370 (21%)	151 (8%)	524 (30%)	34 (2%)
Expert	253 (12%)	596 (27%)	438 (20%)	149 (7%)	686 (31%)	253 (12%)

Analysis of variance (ANOVA) was performed on level 1 data to look at group differences and effect sizes. For results see Table 5. In addition to ANOVA, in order to determine the strength of the relationships among variables, Pearson correlation coefficients were calculated on all three levels of data. All the reported significances are at the .05 level and two-tailed. For tables 6 to 11, numbers in column two (level 1 analysis) are for either ANOVA or correlation, numbers in columns three and four are the actual value of correlation coefficients (*r*).

Table 5

Results of Analysis of Variance and Effect Size on Six Strategies

Strategy	Source	*df*	*F*	*P*	η
Evaluation	Between Groups	2	0.98	0.387	0.27
	Within Groups	26			
	Total	28			
Navigation	Between Groups	2	0.14	0.870	0.11
	Within Groups	26			
	Total	28			
Affect	Between Groups	2	0.11	0.901	0.09
	Within Groups	26			
	Total	28			
Metacognitive	Between Groups	2	3.84	0.035	0.48
	Within Groups	26			
	Total	28			
Cognitive	Between Groups	2	3.95	0.032	0.48
	Within Groups	26			
	Total	28			
Prior knowledge	Between Groups	2	4.12	0.028	0.49
	Within Groups	26			
	Total	28			

Evaluation

The first of the six strategies to be examined in detail is Evaluation. One of the key elements of any search is that the information seekers must be able to correctly judge the relevance of the sources they find with respect to the task at hand. The evaluation of sources is affected by the participants' prior knowledge and ability to monitor the application of that knowledge. Novices were different in this regard from the outset. When the participants were asked in the Background Questionnaire to list the three most important criteria to judge a site (the anticipated correct answer was authorship, currency, and content), half of the experts and half of the intermediates correctly listed them and the other half named at least two of these criteria. Only half of the novices cited one criterion, one novice sited two criteria, and the rest would judge a site based on its appearance, graphics, looks, etc. When it came to applying their knowledge, the difference was striking; 70% of intermediates and 80% of experts made use of their knowledge by applying it to evaluate a site whereas only 20% of novices actually applied their knowledge or clearly talked about it (variable 1.1.2).

There is a high correlation between Evaluation and Metacognitive strategies ($r = .66, p < .001$) that supports the assertion that the ability to evaluate is part of metacognitive abilities. Significant correlations at level 2 data between Reflection and the use of Criteria ($r = .48, p = .003$) and between Monitoring and the Use of Criteria ($r = .53, p = .008$) tell us that use of Criteria is an important contributor to the level of correlation between Evaluation and Metacognition. Evaluation is also significantly correlated with prior knowledge ($r = .60, p = .001$) suggesting that, in order to be able to use evaluation, one should be aware of the evaluation criteria and able to articulate and apply that knowledge. Lack of knowledge (or perhaps bias) about the Google search engine caused Novice 1 not to go to Google right away because she said, "of the topic that had to do with education." In her mind, Google would not have education-related materials. Expert 2, on the other hand, retrieved a document that was called "Expository Teaching and Reception Learning" and, without even scanning the document, she commented, "I found a document that answers the question, so I'm done." She had not found a satisfactory definition of Expository Teaching Model. Most experts spent time reading and understanding the text they wanted to evaluate. They also made sure they understood what Expository Teaching Model meant before settling on a definition. Experts and intermediates typically first read the document and ended up evaluating more sources positively after they had read it, whereas novices evaluated more negatively just by scanning the content or having a dislike for how a certain Web site looked.

From the category of Evaluation, the most important variable that set experts and novices apart was Criteria to evaluate a site.

Table 6

Significant Results for Evaluation: Comparison of Three Levels of Data

Strategies	Level 1, ANOVA and Correlation	Level 2, Correlation	Level 3, Correlation
Evaluation			
.	ANOVA – Not Sig.	Criteria & Results:.40	Criteria & Neg_4: .58
	Correlation:	Criteria & Other: .40	Criteria & Possess: .45
	Metacognitive: .66	Criteria & Ref: .48	622 & Criteria: .42
	Prior knowledge: .60	Criteria & Mon: .53	
	Word: .41	Criteria & Dom: .40	
	Find: .37	Source & Other: .47	
		Source & Ref: .54	
		Source & Mon: .59	
		Source & Domain: .53	
		Source & System: .46	
		Other & Domain: .49	
		Other & System: .50	
		S_eng & Domain: .40	
		Result & O_load: .52	
		Result & Ref: .41	
		Result & Mon: .41	

Navigation

Novices spent more time interacting with the system than engaging in any other strategy. For intermediates, Navigation and Cognition tied as the most used strategies. In comparison, experts spent most of their time in Cognitive activities. At level 1 of the data, Navigation correlated only with one other variable, namely, Satisfaction ($r = .42$, $p = .03$). Participants who used more Navigation strategies were more satisfied with the results of their search. Since novices navigated more, it seems that there is a certain point-and-click without evaluation, which gives them the impression that they are effectively searching the Web and brings them satisfaction. The reverse was true for experts. Expert 6 who found the definition after only 12 minutes was not happy with the results of his search because, he said, "I feel there was something out there, I was in the right direction but I was looking for a site dedicated to that topic."

The participants' choice of search engine was also different. Only 20% of novices used Google compared to 70% of intermediates and experts. MSN, Yahoo, DogPile, and Northern Light were the participants' second choices. Novices used the Back key more often than the other groups but the difference was not significant. It is interesting to note that based on the field observations and notes, although all groups used the Back key, many times their reasons were different. When novices felt lost and did not know what else to do, they would click on the Back key, sometimes more than five times to see where it would take them, hoping to get out of a labyrinth. On the contrary, intermediates and experts would mostly click on it to get back to where they wanted to go. Their move was purposeful and intentional, unlike that of the novices whose use of the Back key was mainly based on trial-and-error.

In order to get to the right site, the participants typed a variety of search terms with some interesting differences. The total number of times novices typed "Expository Teaching Model" and any combination of these words (e.g., expository teaching, expository model, or even expository alone) was greater than the total for intermediates or experts (frequencies = 55, 29, 35, respectively). However, the difference was due to the number of times they typed variations of "Expository Teaching Model" whereas the experts' strategy was to use the search term as-is before trying different combinations.

Table 7

Significant Results for Navigation: Comparison of three Levels of data

Strategies	Level 1, ANOVA and Correlation	Level 2, Correlation	Level 3, Correlation
Navigation	ANOVA – Not Sig.	Move & Type: .43	Back & Access: .65
	Correlation:	Type & Ref:.37 Neg.	Back & Trunc: .46
	Satisfy: .42	Type & Mon:.40 Neg.	211 & Se_21: .73
		T_Oth & Search: .42	213 & 527: .44
		T_Oth & Plan: .40	213 & Scroll: .73
		Wait & O_all: .51	217 & Change: .48
		Wait & O_load: .53	220 & 543: .43 Neg.
			220 & Neu_3: .42
			220 & Etm: .82
			220 & Neu_t: .43
			Etm & Wait: .40

For example, novices typed "Teaching Model" 11 times versus zero for intermediates and twice for experts. Experts and intermediates did not type "Teaching" or "Model" alone as a key word at all, whereas novices typed them five times. If an expert did not type "Expository Teaching Model" as a whole, it was not due to lack of knowledge about the importance of keeping the search term together, but due to a misunderstanding. For example, Expert 7 did not type "Expository Teaching Model" as a whole right away because she said, "I thought you [the researcher] have done the search and if I do it [type ETM], I would get it right away and it will be too easy!" She knew the advantage of using the search term as-is before trying any of its truncated forms. In addition to using the search term as-is and its combinations, all the participants used new search terms, too, but experts and intermediates were more likely to add a new word to any combination of ETM whereas novices more likely introduced a totally new search term. Once the search term was submitted, the novices and intermediates were more impatient than experts for a response from the system. Four novices and five intermediates did not wait for the site to load whereas all the experts waited patiently for their site to load. While waiting, novices were more likely to make a comment about the length of waiting period than intermediates or experts. Novice 9 commented, "Oh, actually it is taking a while to load so I don't like that" and clicked on Stop button. This lack of patience was a common characteristic among novices: 60% of them reported themselves as being impatient compared to 20% for intermediates and 50% for experts. The interesting contrast is that 100% of the novices told the researcher that they do not put a time limit on their search whereas all the intermediates said they do and half of the experts limit the time they spend searching on the Web. It seems that novices know they are impatient but they also know their limitations and assume they need a lot of time on the Web to get reasonable results. They have not yet discovered a routine that will speed up their search. One-half of them said they don't follow a routine in striking contrast to all intermediates and all except one expert who usually have a routine they have found that works for them, and they stick to it.

There are some other interesting, albeit non-significant, correlations that are worth mentioning between Navigation and some other variables. For example, Navigation is negatively correlated with Metacognition ($r = -.20, p = .29$); participants who used more navigation strategies used fewer metacognitive ones as if they were just clicking and scanning without much evaluation and metacognitive thinking. Navigation is also negatively correlated with Prior Knowledge ($r = -.17, p = .38$), participants who navigated more knew less or had misconceptions about the domain or the system. For example, Novice 2 commented, "If it's a dot org and I'm looking for something educational, I might be more motivated to go there." She thought she could differentiate educational sites by looking at .org at the end of their URL. Navigation had negative correlations with Years of University ($r = -.31, p = .10$), Years of Experience ($r = -.10, p = .62$), Information-Seeking ($r = -.05, p = .78$), and Web Knowledge ($r = -.05, p = .79$). In other words, participants who navigated more frequently had fewer years of university, had fewer years of experience, and rated themselves lower on information-seeking and Web knowledge. As was mentioned before, these were not statistically significant correlations but they may be worth investigating in future studies.

A variety of strategies were used by the participants to retrieve the relevant sites. Although not significant, the negative correlation between Metacognitive and Navigation plus the negative correlation between Prior Knowledge and Navigation appeared to reveal the effect of these strategies on how participants search. Point-and-click may have impaired use of metacognitive strategies and brought more disorientation and impatience. The interplay between Metacognitive knowledge and Prior subject Knowledge influenced strategy use. Participants with high metacognitive knowledge appeared more capable to reflect on their actions and be less impulsive and more patient. It also seemed that novices and experts spent time on different aspects of the navigation. Novice 9 spent nine minutes in one site alone but would not wait one minute for a site to load; experts were much quicker to judge the relevance of a site but would patiently wait for a response from the system.

Affect

Affect was significantly correlated with Cognition ($r = .49, p = .007$) confirming the construct validity, as the study of emotions has become an important part of research in cognitive psychology, referred to as "hot cognition" (Lepper, 1988). This strategy was present all the time during the search and, as such, it is hard to evaluate its effect in isolation. What really differentiated novices from experts was seen in the positive comments of experts and in the frustration of novices. Since there was no significant difference in mean scores of Affect between the three groups at the most aggregated level of data, it seemed that all participants expressed their feelings somewhat equally. However, if we look further at levels two and three, we see that the kinds of feelings expressed were different for experts than for the other two groups. In order to measure the total positive statements, a variable called Pos_4 was created. Its value was the total of all positive statements made during affect and metacognitive activities (3.1.1 + 3.2.1 + 3.3.1 + 4.1.1 + 4.1.3). Total negative statements were similarly calculated (Neg_4 = 3.1.0 + 3.2.0 + 3.3.0 + 4.1.0 + 4.1.2). The ANOVA between these two new composite variables and Group variable were calculated. There was no significant difference in how participants used Neg_4. Experts made significantly more positive statements ($F(2,28) = 7.56, p = .003$) about themselves and everything else than novices or intermediates. For example, Expert 1, talking about Google said, "I, I like the Google search engine the best and I always start with it; it has not let me down yet" and E10 commented: "I actually like Scout Report [Archives]." Novice 2, reflecting negatively, said, "Well you see I am not going to find an answer in the next three minutes." And Novice 4 was disappointed: "I don't know, basically some kind of a model . . . but not the one we're looking for." Experts and intermediates also experienced Unfulfilled Expectations (3.2.0) as shown in the raw data (frequencies = 133, 129, 148, respectively), but perhaps experts' expectations from the Web were more realistic. For them, this was business as usual, and they did not take the shortcomings of the Web personally. They made six times more positive statements towards themselves and their abilities than did novices. Novices, on the other hand, felt lost, disoriented, and caught in a labyrinth more often than experts, as shown in the negative significant correlation of Lost with Group ($r = -.46, p = .01$). Interestingly, although experts made more positive comments than others, they themselves were not the most relaxed of the groups. Intermediates were the most relaxed during the search, followed by novices. Experts reported some nervousness after the search, perhaps due to the fact that they have been practicing information seeking at their respective companies for an average of more than 10 years. They had acquired routine skills-and-strategies as a result of years of experience to perform a task rapidly and accurately. (For a comparison of routine-reproductive and adaptive-creative experts, see Syer, Jad, Pelletier, & Shore, 2002.) The present task was out of their domain and challenged their flexibility. Intermediates, on the other hand, were recently trained on how to search in a general domain and had aptly applied their knowledge of information seeking to the task.

As participants were faced with difficulties finding the relevant site, they changed their minds frequently. For example Novice 5 changed her mind without having a clear plan: "Don't seem to be going anywhere. Like I might go back to Excite and [thought not completed]. How do I get back? Or, shall I just click here or 'mmm.' Maybe I should just. Research. Actually I might go back to Excite."

Experts changed their minds about a strategy more often than the other two groups, but their actions seemed more purposeful and usually a result of considering where they were and what their plan was. For example, Expert 9 wanted to change strategies but decided to click on "tips" and get advice on how to search to make sure she was not missing important information. She said, "What about, oh no, I will go back to Britannica and will look at models. No, I guess I could go on on the tips [clicks on tips and advice on how to search] just to see if I'm researching [right]." Or Intermediate 7 who said, "So I guess, OK, maybe I will try Google again. Wait, wait, wait, there is going to be some kind of. I'm going back to the electronic BACK to the library homepage [by cutting the rest of URL in location]. I wonder if anything, ha, electronic reference shelf house." Expert 10 said, "I could try going back to the just cutting down the URL to University of Calgary but one, it could take more time, and, I didn't look yet at others."

All three groups expressed their Positive and Negative Feelings about themselves and everything else throughout the process. However, experts and intermediates were more generous with Positive statements whereas novices felt more lost and changed their mind, without having a clear option, just to see what will happen next.

Table 8

Significant Results for Affect: Comparison of three Levels of data

Strategies	Level 1, ANOVA and Correlation	Level 2, Correlation	Level 3, Correlation
Affect	ANOVA - Not Sig.	Person & Expect: .44	310 & 322: .55
	Correlation:	Person & Planning: .47	310 & 532: .46
	Cognitive: .49	Expect & Rdg: .43	310 & Srch: .41 Neg.
	Satisfy: .38	Expect & Plan: .54	310 & Neu_t: .56
		O_all & O_load: .42	320 & 322: .45
			320 & 532: .49
			320 & Neu_3: .47
			320 & Neu_3: .53
			321 & 526: .46
			321 & Etm: .37 Neg.
			321 & Rdg: .46
			322 & 527: .45
			322 & Neg_3: .61
			322 & Scroll: .47
			322 & Trunc: .55
			322 & Neg-t: .48
			Pos_4 & Group: .54
			Pos_t & Group: .41

Metacognitive

Among the six strategies examined, Metacognitive was one of the three strategies for which the ANOVA was significant in the search process. Participants who used more Metacognitive strategies reflected more on themselves and on the search process, and monitored their own cognitive processes and the progress of the search more frequently. Intermediate 7 put the task paper on the keyboard in front of her and thought the task was not easy: "It's not an easy exercise." Intermediate 4 was the only one who took his pocket watch out and kept more than one search screen open at the same time. He did not give up until he was 100% sure. Intermediate 6 monitored his progress by making sure he did not stray off the topic, "Well this is getting off a little bit." Expert 1 was fully aware of her limitation when she said, "This is an area that I never, I never know, I have no knowledge of it whatsoever." Expert 2 reflected on her knowledge by saying, "I don't have a frame of reference to figure out whether, what I'm looking at is actually is going to explain the concept to me." An ANOVA on level 1 data showed that there was a significant group difference on the use of Metacognitive strategies ($F(2, 28) = 3.84$, $p = .04$). In order to find out where this difference lay, a contrast was used and the results showed that (assuming equal variances between means of the group) novices used significantly fewer Metacognitive strategies than intermediates ($t(26) = -2.6$, $p = .02$) or experts ($t(26) = -2.2$, $p = .04$). There was no significant difference between intermediates and experts. For complete ANOVA results see Table 9.

The strong correlation between Metacognition and Evaluation strategies discussed previously explained why some researchers included evaluation as part of Metacognition. The significant correlation between Metacognitive and Prior Knowledge ($r = .54$, $p = .003$) showed that participants who used metacognitive strategies also possessed more Prior Knowledge. The interplay between metacognitive strategies and Prior Knowledge was confirmed by the significant correlations between Reflection and Domain knowledge ($r = .45$, $p = .01$), Reflection and System Knowledge ($r = .41$, $p = .03$), Monitoring and Domain Knowledge ($r = .45$, $p = .01$), and between Monitoring and System knowledge ($r = .55$, $p = .002$). Since there was not a significant difference in mean scores of Reflection strategies between the three groups, it was concluded that the difference in Metacognitive strategy between novices and experts was due to the difference in their Monitoring strategies. Experts monitored themselves and the search process better than did the others

($r = .48, p = .02$). They were in control by having their finger on the pulse of how they were (i.e., self-monitoring) and also how the search was progressing. For example, Expert 6 dismissed a site saying, "It's not where I want to be" as if he knew or had a mental model of where he should be. He compared where he was with where he should be and concluded that he was in the wrong place.

Older participants used more Metacognitive strategies ($r = .38, p = .04$,) and participants who ranked themselves higher on Information-Seeking Knowledge also used more Metacognitive strategies ($r = .41, p = .03$). Satisfaction had a negative correlation with Metacognitive strategies ($r = -.38, p = .05$). Experts were less satisfied with their search results. For example, Expert 5 found a definition after 14 minutes, but said, "I think that is pretty much an answer but more than one would be nice" and she went back for more definitions. This strategy proved to be fatal for Expert 2 who found the relevant site, but said, " I don't want to spend an hour reading a document" and went to another site. At the end of the session when I asked her why she did not read the whole text, she told me "I never trust the first thing I find, never."

Table 9

Significant Results for Metacognitive: Comparison of three Levels of data

Strategies	Level 1, ANOVA and Correlation	Level 2, Correlation	Level 3, Correlation
Metacognitive	ANOVA - .04	Ref & Mon: .86	Find: .61
	Correlation:	Ref & Domain: .45	Age: .38
	Evaluation: .66	Ref & System: .41	Info-seeking kn.: .41
	Prior knowledge: .54	Monitor & Domain: .45	Satisfaction: -.38
		Monitor & System: .55	

Cognitive

Experts spent more of their time in this strategy than in any other (31% of total). An ANOVA showed a significant difference in mean scores on Cognitive strategies between the three groups ($F(2, 28) = 3.95, p = .03$). In order to find out where this difference occurred, a contrast was used and the results showed that novices used significantly less Cognitive strategies than experts ($t(26) = -2.8, p = .01$). There was no significant difference between novices and intermediates or intermediates and experts.

Experts used Search strategies significantly differently than others ($F(2,28) = 4.17, p = .03$). While searching, they changed their strategies more frequently but also had the highest score for knowing the rationale behind what they wanted to do next. Expert 3 said, "OK, fine, so that strategy is not going to work." Expert 4 gave a reason for changing the search key: "OK, I just want to do the same search but without model and see if I can get anything." Expert 2 changed search engines because "I just try to get a different kind of search engine to see if I can just put a different twist on the top of it because the other guys are all throwing the same document at me over and over again." None of the participants relied on Memory as a strategy. When asked later, they said they would normally either write their search results up, or print them.

One of the most important search strategies was using the Ctrl F key (Find) to find a word in a text. Experts used this strategy more often than novices did ($F(2,28) = 3.51, p = .05$) but the interpretability of this significance is in doubt because the assumption of homogeneity of variance was not met. Only one of the novices used this strategy and most of them did not even know about it. Half of the intermediates and 70% of experts frequently used Ctrl F to search for a term. Expert 5, in particular, used it in an efficient way as she explained, "however to find out [a definition] I will have to sit in here and read this thing or I could go to the edit Find on this page [typed "defin"], now I have got defin and it will find either definition or not [define or definition]."

Within Cognitive, the next significant strategy was Planning ($r = .47, p = .01$). Among the different variables in planning, experts significantly expressed their Intentions to either navigate ($r = .42, p = .02$) or take Another Action ($r = .43, p = .02$). Expert 9 expressed her intention to be faster: "This time I won't be. I'll I will check right away if there is phrase searching."

Table 10

Significant Results for Cognitive: Comparison of three Levels of data

Strategies	Level 1, ANOVA and Correlation	Level 2, Correlation	Level 3, Correlation
Cognitive	ANOVA - .03	Search & Plan: .60	513 & 543: .45
	Correlation:	Search & Track: .57	513 & Criteria: .42
	Affect: .49	Plan & System: .40	526 & Pos_3: .52
	Word: .68	Plan & Group: .47	526 & Error: .54
			526 & Pos_t: .48
			527 & Neu_3: .40
			527 & Scroll: .47
			527 & Wait: .44
			531 & Group: .43
			531 & 532: .51
			531 & Word: .41
			531 & Change: .41
			532 & Group: .42
			532 & Neg_3: .60
			532 & Neu_3: .42
			532 & Neg_t: .57
			532 & Neu_t: .45
			543 & Etm: .40 Neg.

Novice 4 planned to click on a site: "I'm going to Timothy series" and N3 said, "There is this one article that has come up with all the search engines that I haven't looked at so I've decided I'm going to look at it."

There was not a significant difference in Reading, how to keep Track of Search, Curiosity Stop, and use of Metaphors between our groups of participants. One of the important cognitive strategies, namely, Linking New Knowledge to old was used more often by intermediates and experts than by novices and is worth investigating in future studies.

Prior Knowledge

Similar to Metacognitive and Cognitive strategies, Prior Knowledge was significantly predictive of how participants searched for information on the Web. Experts had higher Prior Knowledge than other groups ($F(2, 28) = 4.12, p = .03$). The difference was primarily between Prior Knowledge of novices and experts ($t(26) = -.2.8, p = .009$). There were no significant differences between Prior Knowledge of novices and intermediates or between Prior Knowledge of intermediates and experts. As none of the participants had Prior Knowledge about Expository Teaching Model (Domain knowledge), the level of System knowledge had a significant influence on strategy use ($r = .50, p = .006$). Inadequate System knowledge limited the use of the system. As Novice 9 admitted, "I know there is some way of searching on the Internet. I think you can search keywords and then move on to the area but I don't know that, I'm not an advanced searcher." Novice 1's comments reflected on novices' lack of trust in what they knew: "I don't know, to see if every definition is more or less the same or there is [sic] large gaps." Experts were also faced with their knowledge limitations but they usually expressed these as a matter of fact, as Expert 6 said, "I'm not really familiar with that [MSN home page]." Most of the participants did not take advantage of the help key that is provided by all search engines. Intermediate 3 was the only participant who used AltaVista's "cheat sheets." In order to check the significance of total Prior Knowledge, i.e., Domain and System knowledge together, a variable called Possess was created by adding the three values of variables 6.1.2 + 6.2.2 + 6.2.3. Experts used more Possess than others ($F(2,28) = 3.97, p = .031$), but the interpretability of this significance is in doubt because the assumption of homogeneity of variance was not met. Participants with higher Prior Knowledge were older ($r = .47, p = .01$), had more job experience ($r = .43, p = .02$), and spent more hours a day in front of a computer ($r = .50, p = .006$).

Table 11

Significant Results for Prior Knowledge: Comparison of three Levels of data

Strategies	Level 1, ANOVA and Correlation	Level 2, Correlation	Level 3, Correlation
Prior Knowledge	ANOVA - .03 Correlation: Evaluation: .60 Metacognition: .54 Word: .56	Domain & System: .41 System & Track: .40	622 & Group: .54 622 & Pos_3: .41 622 & Pos_4: .65 622 & Lack: .58 622 & Pos_t: .51 Possess & Group: .48 Age: .47 Job: .43 H/day: .50

Participants' Attributes

As can be seen in Table 12, there was a significant mean score difference between the three groups. Although experts were more educated, had more job experience and higher computer and information-seeking knowledge, they expressed a certain amount of anxiety even after finding the relevant data. Novices, on the other hand, did not find as many good sites as experts, and felt lost more frequently, but they were more satisfied with the results of their search.

Table 12

Results of Analysis of Variance and Effect Size on Participants' Attributes

Variable	Source	*df*	*F*	*P*	*η*
Age	Between Groups	2	4.78	.017	0.74
	Within Groups	26			
	Total	28			
Year	Between Groups	2	9.16	.001	0.65
	Within Groups	26			
	Total	28			
Experience (Total)	Between Groups	2	9.10	.001	0.64
	Within Groups	26			
	Total	28			
Experience (This job)	Between Groups	2	12.23	< .001	0.72
	Within Groups	26			
	Total	28			
Computer knowledge	Between Groups	2	5.51	.010	0.55
	Within Groups	26			
	Total	28			
Info-Seeking knowledge	Between Groups	2	8.17	.002	0.62
	Within Groups	26			
	Total	28			
Web knowledge	Between Groups	2	6.02	.007	0.56
	Within Groups	26			
	Total	28			
H/day on computer	Between Groups	2	30.44	< .001	0.84
	Within Groups	26			
	Total	28			
H/day on the Web	Between Groups	2	5.43	.011	0.54
	Within Groups	26			
	Total	28			

Is the Use of Certain Strategies Associated with Success?

The second research question addressed the possibility of a relationship between strategies and attributes and timely success of the search on the Web. Success was defined as finding a correct definition of Expository Teaching Model within 30 minutes and timely success was defined as finding it relatively sooner within that period. In other words, data responded to two questions: (a) *whether* the event happened (Did the participant find the definition?), and (b) *when* did the event happen (How long did it take to find the answer?). For this question, the data were analyzed using survival analysis (Cox, 1972), more specifically, the Cox Proportional Hazard model. In this model the impact of an explanatory variable (x) is exponentially related to the baseline hazard function. After exploring all levels of data, the following explanatory variables (Table 13) were found to have a significant relationship with timely success of the search.

Results for the first research question had indicated that experts outperformed the other two groups on three strategies: Metacognition, Cognition, and Prior Knowledge. However, with regard to the second research question, when all the participants were pooled together, results of the survival analysis showed that several variables, listed in Table 8, from all three levels of data, had a significant impact on finding a correct definition of Expository Teaching Model. In this analysis, a good strategy could have been employed by a novice or an expert. At level 1 data, the two most important strategies were Evaluation (*Beta* = .03, $p = .04$) and Metacognition (*Beta* = .07, $p = .004$). The only two novices who found the definition of ETM, i.e., Novices 2 and 7, made the two highest numbers of Evaluation and Metacognition statements in their groups. Novice 2 showed her clear evaluation criteria when she said, "So I also sometimes look at who made the site, when I see things about advertisers, 10 cents, I am less inclined to look at them." The other novice reflected on the content of some Web pages and said, "How can they make Web pages like that where like [there is] information on nothing?" She was also aware of the effect of verbal protocols on her performance when she commented, "It's hard to talk and think of what I'm trying to say, very difficult."

performance when she commented, "It's hard to talk and think of what I'm trying to say, very difficult."

Table 13

Summary of Cox Survival Analysis Results: Significant Variables

Level	Variables	*Beta*	*Standard Error*	*df*	*p*	*Exponential Beta*
Level 1	Evaluation	.03	.02	1	.04	1.03
	Metacognition	.07	.03	1	.004	1.08
Level 2	Use of Criteria	.14	.05	1	.007	1.15
	Source	.06	.03	1	.04	1.06
	Typing	-.28	.12	1	.02	0.76
	Reflection	.11	.04	1	.007	1.11
	Monitoring	.19	.06	1	.002	1.21
Level 3	Content	.10	.05	1	.04	1.11
	Criteria	.15	.05	1	.007	1.16
	Pos_3	.09	.04	1	.008	1.10
Attributes	Info-seeking	.62	.30	1	.04	1.86
	Age	.05	.02	1	.05	1.05
	Feeling lost	-1.41	.65	1	.03	0.25

The importance of evaluation and Metacognition tells us that regardless of other strategies and attributes that helped experts become more successful on this task, if information seekers evaluated and monitored much as experts did, their chances of success would increase.

Exploring level 2 data, there were five variables that had a significant impact on hazard function. The two categories that made evaluation significant were Use of Criteria address right on to decide to choose or not because depending by what type of address they have, what ending it is, whether it is .org or .edu or you know there is new ones .TV." It is interesting to note that Novice 2 had the highest score in 1.1 but Novice 7 scored zero. This was compensated by variable 1.2 (Evaluation of one source) where Novice 7 scored more than twice as highly as Novice 2.

Since Evaluation and Metacognition are intertwined, it was not surprising to see that both categories within Metacognition, i.e., Reflection and Monitoring, also had significant relationship with the success of the search. Participant 2 had the highest number of metacognitive strategy statements in his group, for example, "Let me think, none of these are what I want. OK I'm not really finding anything else here."

The surprising variable that had a significant relationship to success at level 2 was Navigation, specifically, Typing the search term (ETM) or any of its combinations (2.2). Navigation at level 1 was not significantly related to success, perhaps due to the fact that navigation scores were aggregated with other values and that their total was not significantly related to success. We found that typing the search term either as-is or in truncated form improved the chances of success more than typing a new search term. However, in survival analysis results, the negative *Beta* shows that more navigation, and specifically typing the search term and its combinations more often, had a negative relation to the success of the search, that is, it reduced the chances of success.

In order to further explore the effects of level 2 variables, several composite variables were created by combining groups of level 3 data. The variable Criteria was the sum of 1.1.1 (Tacit Criteria) +1.1.2 (Clear Criteria) +1.1.3 (Surface Criteria). It was significantly related to the success of the search. Among all the variables that comprise Evaluation, Use of Criteria (1.1) proved to be most related to the success of the search.

Another surprising significant variable was 5.2.6 that was defined as "when user is inside the site and reads the Content thoroughly (instead of scanning it quickly)." For example, Intermediate 8 spent six minutes in one site, and Intermediate 9 read one site thoroughly before moving to the next.

The last strategy that was significantly related to success was a composite variable called Pos_3. The value for this variable was the sum of three positive feelings when they expressed them toward themselves, the search, or the overall task: 3.1.1 (Personal) + 3.2.1 (Experimental) and + 3.3.1 (Overall). It had a significant relationship with the success of the search. Among participants' attributes, older participants were more successful. There was no gender difference and training was not significant either. However, more years of university were significantly related to success and knowledge of information seeking proved to be more important than computer knowledge.

When conducting a successful search on the Web, therefore, the searcher (a) clearly used criteria to evaluate the sites, (b) did not navigate too much, (c) reflected on strategies and monitored progress, (d) had background knowledge about information seeking, and (e) approached the search with a good attitude and enjoyed the process.

Can a Successful Search on the Web be Modeled?

The third question examined whether a successful search on the Web can be modeled. That is, using the findings of the previous section, can we obtain a model that can be used to describe the dependent variable *t*.

The outcome variable considered in this analysis was the time it took participants find a definition of ETM on the Web. To investigate the effect of the explanatory variables on the outcome variable and to build a model for the data, the significance of each individual variable on its own was first tested. From the answer to the second research question, we know that there were 13 variables from all levels of data that had a significant relationship with the time to complete the task. To reveal which variables were most reliable in the presence of other variables, the 13 variables were entered into the Cox model together, and similar to the backward elimination technique in regression, the effect of omitting each of the variables in the model was evaluated. Level 1 variables, Evaluation and Metacognition, were not entered because they were represented by levels 2 and 3 data. Evaluation was represented by 1.1, 1.2, and Criteria; Metacognition was represented by 4.1 and 4.2. A subset of significant variables that remained in the Cox model was identified. This final survival model contained only statistically significant variables.

In the final count, three variables, namely, Information-Seeking Knowledge, Criteria, and Typing search term were left. These variables together have a significant relationship with the success of the search.

Table 14

Information-Seeking Model

Variables	*Beta*	*Standard Error*	*df*	*p*	*Exponential Beta*
Information-Seeking Knowledge	.502	.298	1	.092	1.652
Criteria	.133	.059	1	.025	1.142
Typing search term	-.305	.139	1	.028	.737

This model that represents the pattern of search by experts is derived entirely empirically from the data. The variables are, in order of impact, to conduct a successful search on the Web, the information seeker is expected to possess high information-seeking knowledge, know the criteria and clearly use them to evaluate sites, and refrain from too much navigating (especially typing search terms too often).

Figures 1 and 2 give the graphs of survival and hazard functions for the above model at the mean of covariates. That is, in these graphs the value of Information-seeking knowledge is fixed at its mean (3.72), value of Criteria is fixed at its mean (3.59), and value of Typing search term is fixed at its mean (4.10). The survival curve shows the proportion of participants who did not find the meaning of ETM at the outset, 10 minutes, 20 minutes, and up to 30 minutes. The hazard function shows proportion of participants who found the meaning of ETM.

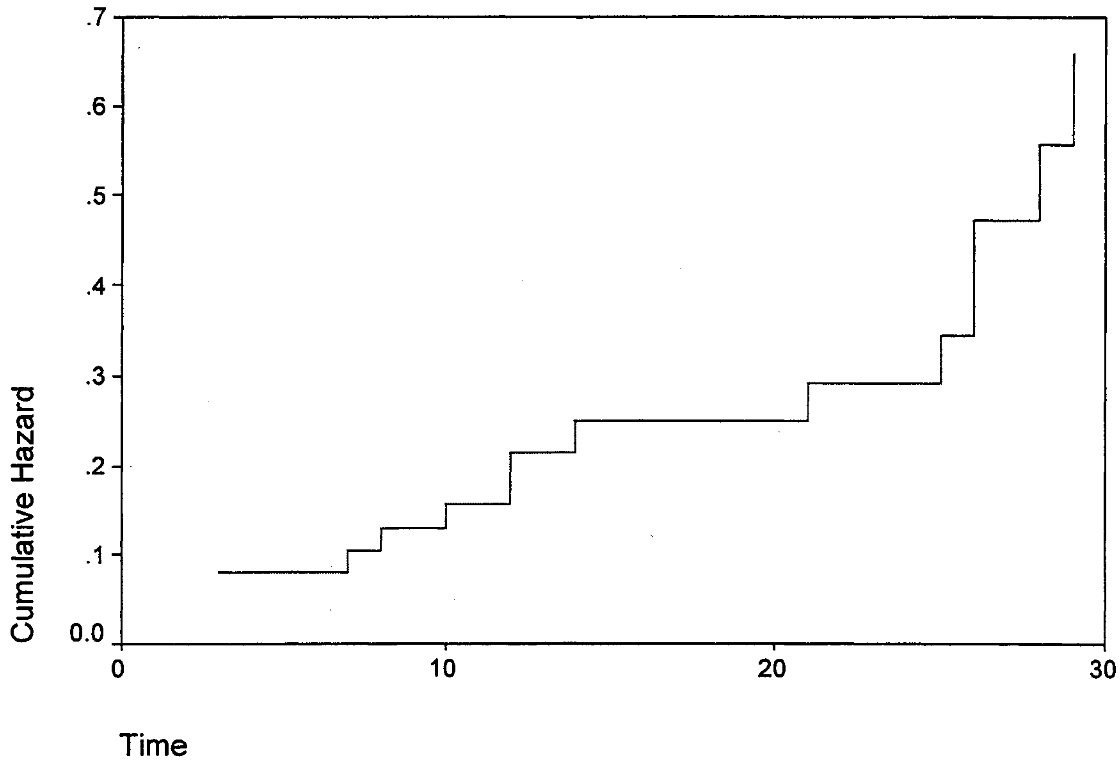

Figure 1. Survival function at mean of covariates.

These graphs describe the impact of use of information seeking, Criteria, and Typing on the outcome of the search. For example, an information seeker whose values of the top three variables are 3.72, 3.59, and 4.10 respectively, as in Figure 1 and 2, has about 25% chance of finding the definition at 20 minutes.

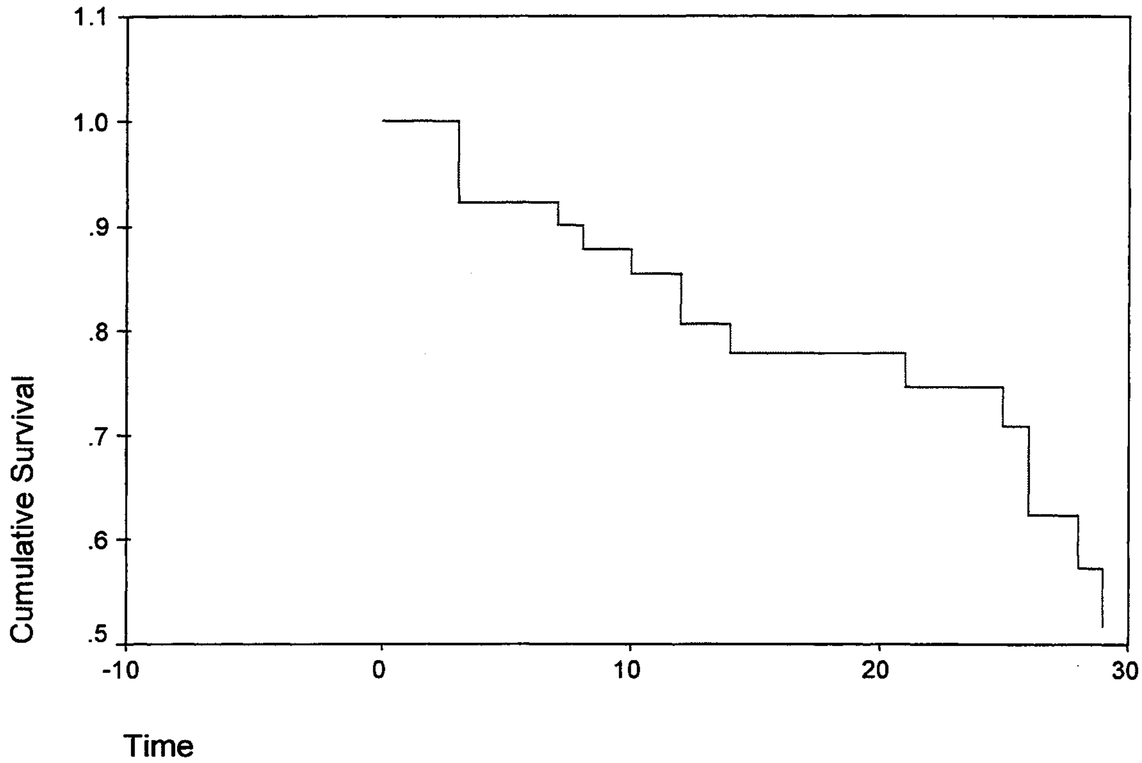

Figure 2. Hazard function at mean of covariates.

Chapter 4: Summary and Conclusions

This study investigated the effect of strategies and attributes on information-seeking processes on the Web. This was a quasi-experimental study in which verbal protocols were collected from three groups of participants who searched for a definition of Expository Teaching Model on the Web. The findings will be interpreted in three sections, each addressing one of the research questions in the context of the hypotheses and previous research.

The participants were chosen with certain differences in mind and the results showed that they did systematically differ, but the difference was greater between novices and the others than between intermediates and experts. Experts had domain-specific knowledge (Keating, 1990) and had mastered performance skills specific to the content area. The task presented a problem outside their domain of expertise. Intermediates' principal knowledge, on the other hand, was domain-general, an outcome of their initial professional education. A quick look at the content of the Web courses they were taught in their Master of Library and Information Studies degree revealed that they received common Web Training, but they also entered graduate school with highly diverse undergraduate backgrounds. Their cumulative GPA minimum was 3.0, the dropout rate was low, and their motivation was high. Of their eight core courses, six dealt with information and two specifically with on-line searching. Almost all students had summer internship experience in their field. Novices' knowledge about the Web, in contrast to both intermediates and experts, was un-integrated and scattered. There were no compulsory technology course in their curriculum, eight out of 10 reported personal interest as the prime motive for learning how to use the Web. Their lack of knowledge about ETM showed that this topic was either not covered or perhaps was covered early in their education and forgotten.

The first research question addressed the possibility of differences in search patterns among users with varying degrees of expertise. This study confirmed the hypothesis that experts' patterns of search on the Web are different from novices. The most significant differences were found in the cognitive, metacognitive, and Prior Knowledge categories. Within cognitive strategies, experts used different kinds of planning strategies and they expressed their intentions to either navigate or take another action. Differences in the use of trial-and-error were not significant.

These results echo research done by Hill and Hannafin (1997) in which participants on a Web search chose their own topics. Those who used metacognitive strategies and possessed a broader knowledge base were more successful and experienced less disorientation than those who did not. Land and Greene (2000) also reported similar results: Use of metacognitive strategies and possession of domain and system knowledge were critical to the success of the Web searching.

Use of the *Back* key as a popular command among participants was also found in other studies such as Catledge et al. (1995), Rumpradit (1998), Tabatabai and Luconi (1998), and Tauscher and Greenberg (1997). Novices use backtracking to see where it will take them, hoping to get out of a labyrinth and find a comfort zone. On the contrary, intermediates and experts use it to go where they want to go. Similar to findings of this study, Cove and Walsch (1988) also found that novices are less patient, and rely more on trial-and-error. Impatience led them to navigate more, to click more, and to execute before spending enough time exploring or planning. Novices in this study used the highest number of navigation moves in the group. Their sometimes aimless clicking was called "serendipitous browsing" by the above authors. Shore and Lazar (1996) reported similar findings: A high-IQ school-age group spent relatively more time on exploration of a complex problem, whereas a control group was anxious to start executing the solution steps.

In terms of knowledge and memory, the novices were at a disadvantage. Many studies found that experts have better memory, more prior knowledge, and more know-how to use it (Coleman & Shore, 1991; Shore & Kanevsky, 1993; Tabatabai & Luconi, 1998). Despite their lack of knowledge, novices in this study did not use the on-line help at all. Large and Beheshti (2000) were also surprised to find that "In fact, on-line help was almost totally ignored" (p. 1077).

Although there was no overall difference between groups vis-à-vis affect, it had an umbrella effect on the search process. To borrow terminology from cognitive science, the difference between the groups was in "meta-affect" or awareness of one's feelings. Experts were more aware of their feelings, were not surprised by them, and relied on them for modifying their own perceptions. Their positive outlook on the search gave them patience to wait for a response from the system.

Novice participants move through the stages of information-seeking processes making relatively random choices and using their knowledge to navigate the system, but typically do not successfully reach their goal. Intermediates and experts, in contrast, integrate responses from the system with their own knowledge and monitor their decisions, changing their strategies as needed.

The second research question addressed the effect of certain strategies and attributes on the timely success of the search. Level 1 data revealed that timely success of the search was associated with the use of Metacognitive and Evaluation strategies. These findings are in line with prior findings that one of the characteristics of experts is that they finish more tasks, are more efficient, and make fewer mistakes (Coleman & Shore, 1991; Lazonder et al., 2000). These results confirm part of the hypotheses that Metacognition is associated with success but refutes the association of success with Prior Knowledge.

The third research question was an attempt to model the successful search on the Web. It was hypothesized that there are certain strategies and attributes that impact upon the outcome of the search and these factors can be modeled using survival analysis methods. The data and design by themselves did not support a causal conclusion, but the model helps to initiate a conversation about causality. The hypothesis was confirmed and a causal or partially causal model that is consistent with the data and included three factors (Information-seeking knowledge, Criteria, and Typing search term) was created.

The fact that only three variables (from a list of 13 significant variables) were in the model may be due to the similarity between intermediates and experts. According to this model, possession of Information-Seeking Knowledge, Use of Criteria to evaluate sites, and refraining from too much navigation are associated with success. This survival model is based on a unique process and it is difficult to compare it to the other models reviewed in Chapter 2 because of discrepancies between the models. Marchionini's Personal Information Infrastructure is much broader than this model and Huberman's is based on laws of probability, not survival analysis. However, this model does share its origin with Kuhlthaus's (1993) Information Search Process because it incorporated thoughts, actions, and feelings.

The response to the first research question painted a picture of difference in patterns of search among experts and novices. The most significant differences were found in three of the six strategies (Metacognitive, Cognitive, and Prior Knowledge). The second question explored those differences deeper by looking at the effect of strategies on the success. Two strategies were found to be associated with success (Metacognitive and Evaluation). The common strategy found by results of both questions, i.e., Metacognitive, echoes conclusions of other studies: Use of this strategy could be one of the criteria that separates experts from novices.

The responses to the second and third questions revealed that user-centered studies could shed light on the differences in the Web searching between experts and novices. Expert participants in this study shared a few characteristics with other experts, such as, on average, they found the definition of Expository Teaching Model faster, had higher Prior Knowledge, had different search strategy, and made more metacognitive statements. Investigating success of the search brought expected and unexpected results. Metacognitive strategies were associated with success but Prior Knowledge was not. Modeling the Web search with survival analysis was an effective way to isolate the most important variables that are associated with timely success of the search.

Contributions to Knowledge

This study contributes to knowledge in the following areas:

1. The conceptual contribution of this thesis to the field of expert-novice differences is two-fold. The first part is through the definition of what constitutes expertise on the Web in order to create the survival model. Traditionally, experts and novices have been defined either by personal attributes or measures such as hours of Web use. The researchers then compared experts and novices as two distinct and well-defined groups, and reported the differences. In this study, expertise in information seeking has been defined as a continuum of observable strategies that encompasses novices, intermediates, and experts. In this manner, strategies and attributes of experts are objectively identified and related to the outcome of the search.

The second unique contribution is achieved by creating a model that begins to explain users' behavior beyond their idiosyncratic differences. This model is a unique contribution because it is the first time that these factors are brought together to create a model that could describe or possibly predict a user's Web-searching performance.

2. The methodological contribution of this thesis is also two-fold. The first is the introduction of survival analysis and development of a survival model of Web-searching strategies and their relative relation to success. It allows a simultaneous analysis of whether and when a participant finds the required information. The second methodological contribution is the extensive coding schema that consisted of eight categories in three levels of aggregation. The codes are numbered instead of labeled, and have the potential for use in other domains.

3. The practical contribution of this study is to highlight that there are specific actions associated with success in Web searching, namely refraining from navigating too much without thinking, learning the evaluation criteria and using them to judge the relevance of one site before making a decision to go to another, reflecting on steps to take and monitor progress, increasing their background knowledge about information seeking, and taking steps to prevent getting lost. This information may be useful in building

curricular interventions for novice groups, such as student teachers, and should be a part of a checklist for the content of web courses designed to train expert Web information seekers.

Limitations

There are several limitations to this study. Methodological limitations are related to the use of verbal protocols to collect data. According to Ericsson and Simon (1993), because thinking processes are much faster than talking, when participants try to verbalize, their thinking processes slow down (the automaticity argument). Furthermore, verbalization can change the decision-making process (interference-competition). With respect to the sample, the expert participants turned out to be slightly different than we thought. Their similarity to intermediates may have limited the scope of the model. Experts were observed in their place of work, but intermediates and novices were not; this raises issues about generalizability of results in the natural environment.

Future Research

Information seeking on the Web is a relatively recent endeavor and research about it is still in its infancy and in need of development. The participants were chosen with certain differences in mind and the results showed that they did systematically differ, but the difference was more between novices and the others than between intermediates and experts. Expertise related to domain-specific knowledge implies a degree of automaticity, usually in a familiar and repetitive situation. The task in this study presented a problem known to be outside the experts' and intermediates' specific domain of expertise. Future research could explore different kinds of expertise and its effect on seeking information on the Web. One group of experts could have domain-specific knowledge (or expert-in-context as defined by Hoffman et al., 1996), another group, domain-general. The results of the research could shed more light on what constitutes expertise on the Web. A more complex task and a longer time to work on it might also help to elucidate the differences between the experts and intermediates of the present study, and among all levels of expertise.

References

Aguilar, F. J. (1967). *Scanning the business environment.* New York: Macmillan.

Allison, P. D. (1995). *Survival analysis using the SAS system, a practical guide.* Cary, NC: SAS Institute.

Bates, M. J. (1979). Information search tactics. *Journal of the American Society for Information Science, 30,* 205-214.

Bates, M. J. (1987). How to use information search tactics online. *Online*, May, 47-54.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review, 13*, 407-423.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval. Part I: Background and theory. Part II: Results of a design study. *Journal of Documentation, 38,* 61-71, 145-164.

Borgman, C. (1989). All users of information retrieval systems are not created equal: An exploration into individual differences. *Information Processing & Management, 25,* 237-251.

Breuleux, A., Renaud, A., Large, A., & Beheshti, J. (1993, April). *Effects of multimedia presentation on children's acquisition of information.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Brown, A. L., Bransford, J. L., Ferrara, R. L., & Campoine, J. C. (1983). Learning, remembering, and understanding. In P. H. Mussen (Ed.), *Handbook of child psychology* (pp. 77-166). New York: Wiley.

Bush, V. (1945). *As we may think.* Retrieved July 31, 2000, from http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm.

Cartwright, G. F. (1982). *The impact of symbionic technology on education* (Paper 82-02). Toronto: Research and Evaluation, Ministry of Education, Government of Ontario.

Catledge, L. D., & Pitkow, J. E. (1995). *Characterizing browsing strategies in the World Wide Web*. Retrieved August 1, 2000 from http://www.igd.fhg.de/archive/1995_www95/papers/80/userpatterns/UserPatterns.Paper4.formatted.html.

Chi, M. T. H., & Glaser, R. (1985). Problem solving ability. In R. J. Sternberg (Ed.), *Human abilities: An information processing approach* (pp. 227-250). New York: Freeman.

Choo, C. W., Detlor, B., & Turnbull, D. (1999). *Information seeking on the Web - An integrated model of browsing and searching*. Paper presented at the annual meeting of the American Society for Information Science, Washington DC.

Cole, C., & Kuhlthau, C. C. (2000). Information and information seeking of novice versus expert lawyers: how experts add value. *The New Review of Information Behavior Research, 1,* 103-115.

Coleman, E. B., & Shore, B. M. (1991). Problem-solving processes of high and average performers in physics. *Journal for the Education of the Gifted, 14,* 366-379.

Cove, J. F., & Walsch, B. C. (1988). Online text retrieval via browsing. *Information Processing & Management, 24,* 31-37.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, 34,* 187-202.

de Groot, A. D. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.

Dillon, A., & Gabbard, R. (1998). Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research, 68,* 322-349.

Ellis, D. (1989). A behavioral model for information retrieval system design. *Journal of Information Science, 15,* 237-247.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: verbal reports as data*. Cambridge, MA: MIT Press.

Gagné, R. M. (1985). *The conditions of learning and theory of instruction*. New York: Holt, Rinehart, & Winston.

Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist, 21,* 99-120.

Greeno, J. (1978). Natures of problem-solving abilities. In W. Estes (Ed.), *Handbook of learning and cognitive processes* (pp. 239-270). Hillsdale, NJ: Erlbaum.

Hannafin, M. J. (1999). *Learning in open-ended environments: Tools and technologies for the next millennium.* Retrieved May, 20, 1999 from http://dragon.ep.usm.edu/~dsurry/ole.htm.

Hannafin, M. J., Hall, C., Land, S., & Hill, J. (1994). Learning in open-ended environments. Assumptions, methods, and implications. *Educational Technology, 34,* 48-55.

Hill, J. R., & Hannafin, M. J. (1997). Cognitive strategies and learning from the World Wide Web. *Educational Technology Research & Development, 45,* 37-64.

Hjorland, B. (1997). *Information seeking and subject representation.* Westport, CT: Greenwood Press.

Hoffman, R. R., Feltovich, P. J., & Ford, K. M. (1996). A general framework for conceiving of expertise and expert systems in context. In P. J. Feltovich, K. M. Ford, & R. R. Hoffman (Eds.), *A preliminary tour of human and machine expertise in context* (pp. 543-580). Cambridge: MIT Press.

Huberman, B. A., Pirolli, P. L. T., Pitkow, J. E., & Lukose, R. M. (1998, April 3). Strong regularities in World Wide Web surfing. *Science, 280,* 95-97.

Jaušovec, N. (1994). *Flexible thinking: An explanation for individual differences in ability.* Cresskill, NJ: Hampton Press.

Keating, D. P. (1990). Charting pathways to the development of expertise. *Educational Psychologist, 25,* 243-267.

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

Lajoie, S. P. (Ed.) (2000). *Computers as cognitive tools*, volume two: No more walls: theory change, paradigm shifts, and their influence on the use of computers for instructional purposes. Mahwah, NJ: Erlbaum.

Lajoie, S. P. (1993). Computer environments as cognitive tools for enhancing learning. In S. P. Lajoie & S. Derry (Eds.) *Computers as cognitive tools* (pp. 261-288). Mahwah, NJ: Erlbaum.

Lajoie, S. P., & Derry, S. J. (1993). *Computers as cognitive tools*. Hillsdale, NJ: Erlbaum.

Land, S. M., & Greene, B. A. (2000). Project-based learning with the World Wide Web: A qualitative study of resource allocation. *Educational Technology Research & Development, 48* (1) 45-68.

Large, A., & Beheshti, J. (2000). The Web as a classroom resource: Reactions from the users. *Journal of the American Society for Information Science, 51,* 1069-1080.

Large, A., Beheshti, J., & Breuleux, A. (1998). Information seeking in a multimedia environment by primary school students. *Library and Information Science Research, 20,* 343-376.

Large, A., Tedd, L. A., & Hartley, R. J. (1999). *Information seeking in the online age: Principles and practice.* West Sussex, UK: Bowker-Saur.

Lawless, J. F. (1982). *Statistical methods for lifetime data.* New York: Wiley.

Lazonder, A. W., Biemans, Harm, J. A., & Wopereis, I. G. J. H. (2000). Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science, 51,* 576-581. Lepper, M. R. (88) cog & inst, 5,4, p. 289

Lepper, M. R. (1988). Motivational considerations in the study of instruction. *Cognition and Instruction, 5,* p. 290

Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge, UK: Cambridge University Press.

Marchionini, G. (1998). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science. 40,* 54-66.

Palmquist, R. A., & Kim, K. (2000). Cognitive style and on-line database search experience as predictors of Web search performance. *Journal of the American Society for Information Science. 51,* 558-566.

Pelletier, S. & Shore, B. M. (in press). The gifted learner, the novice, and the expert: Sharpening emerging views of giftedness. In D. C. Ambrose, L. Cohen, & A. J. Tannenbaum (Eds.), *Creative intelligence: Toward theoretic integration.* New York: Hampton Press.

Rumpradit, C. (1998, August). *From novices to experts: User performance, confidence, and satisfaction on the World Wide Web.* Paper presented at the annual meeting of the Association for Information Systems, Baltimore MD.

Shore, B. M., & Kanevsky, L. S. (1993). Thinking processes: Being and becoming gifted. In K. A. Heller, F. J. Mönks, & A. H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 131-145). Oxford, UK: Pergamon.

Shore, B. M., & Lazar, L. (1996). IQ-related differences in time allocation during problem solving. *Psychological Reports, 78,* 848-850.

Simon, H. A. (1979). Information processing models of cognition. *Annual Review of Psychology, 30,* 363-396.

Sutcliffe, A., & Ennis, M. (1998). Towards a cognitive theory of information retrieval. *Interacting with Computers, 10,* 321-351.

Syer, C. A., Jad, R., Pelletier, S., & Shore, B. M. (2002). *Adaptive-creative versus Routine-reproductive expertise in hypermedia design: An exploratory study.* Manuscript in preparation.

Tabatabai, D., & Luconi, F. (1998, August). *Expert-novice differences in searching the Web.* Paper presented at the annual meeting of the Association for Information Systems, Baltimore, MD.

Tapscott, D. (1996). *The digital economy: promise and peril in the age of networked intelligence.* New York: McGraw-Hill.

Tauscher, L., & Greenberg, S. (1997). How people revisit web pages: Empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies, 47,* 97-137.

Thorndike, E. L. (1912). *Education.* New York: Macmillan.

Walker, G., & Janes, J. (1999). *Online retrieval--A dialogue of theory and practice.* Englewood, CO: Libraries Unlimited.

Walster, D. (1996). Technologies for information access in library and information centers. In D. H. Jonassen (Ed.), *Handbook of research for educational and communication technology,* (720-752). New York: Macmillan.

Willet, J. B., & Singer, J. D. (1991). From whether to when: New methods for studying student dropout and teacher attrition. *Review of Educational Research, 61,* 407-450.

Wilson, T. D. (1997). Information behavior: An interdisciplinary perspective. *Information Processing & Management, 33,* 551-572.

Appendix A: Glossary

Affect	Emotions and feelings expressed during the search. May be broken down further into personal feelings and system-oriented feelings.
Attributes	Varied data on the participants: Age, Sex, Years of University, Major area of study, Years of experience, Years of teaching (if applicable), Computer knowledge, Info-Seeking knowledge, Web knowledge, How they learned to use the Web, Hours a day spent on computer, and Hours a day spent on the Web.
Cognitive Strategy	Mental activities such as thinking, reading, planning, learning, processing, retrieving, and problem solving.
Evaluation	Assessment of relevancy of sources, search engines, and results.
Domain Knowledge	Knowledge that users have of the broader frame of reference or organized knowledge. In this study the domain is field of education
Hot cognition	Cognition that deals with affect.
Hypermedia	Non-sequential linked text, sound, and images on the Web.
Information Seeking	A purposeful activity intended to change the state of knowledge.
Intermediary	The search may be carried out by the person who needs the information (end user), or it may be carried out by an intermediary, usually a librarian or information scientist who knows the language of the system being used and can interpret or "translates" the request into the necessary language.
Intermediate	A person whose knowledge, skills, and abilities are between a novice and an expert.

Metacognitive Strategy	Reflection and monitoring, or self-control/regulation exercised when engaged in a cognitive activity.
Move	Fine-grained actions, feelings, and knowledge that are consciously selected, applied, and monitored to solve a problem
Navigation	Any action using the mouse or the keyboard, and waiting for a reply from the system.
Open-Ended Learning Env. (OELE)	Learner-centered systems that facilitate problem solving by supporting learners' efforts to understand that which they determine to be important.
Prior Knowledge	Knowledge that users have of either the domain or the system.
Quasi-Experiments	Experiments that have treatments, outcome measures, and experimental units but do not use random assignment to create the comparisons from which treatment-caused change is inferred (Cook & Campbell, 1979, p.6).
Strategy	One of six processes: Evaluation, Navigation, Affect, Metacognition, Cogitation, and Prior Knowledge. It is made up of several tactics.
System Knowledge	Prior knowledge about the Web and the computer.
Tactic	Short-term sub-processes that include several moves.

Appendix C: Equipments and Material

The laptop used was a Dell Inspiron 5000e model with 128 MB of RAM with a 15-inch active matrix screen.

A voice recognition software called Dragon Naturally Speaking© was used to transcribe verbatim. It usually takes about a half hour to train the software so that it can recognize our speech patterns. In this study, since the data consisted of 29 different voices, the researcher listened to each participant's voice and repeated the words uttered by him or her.

The VCR was a Panasonic PV-8554 with 4 head Hi Fi Stereo Omnivision.

The audio system was a portable Sony with microphone.

Transcriber was a Sanyo Compact Cassette Dictating/Transcribing system.

Researcher's notes consisted of a Spiro booklet where observations, names and codes of participants and any questions were recorded.

Appendix D: Participant Consent Form

This is to state that I agree to participate in the research project entitled: *Information- Seeking Expertise: Modeling a Web Search.* This study is part of Ph.D. dissertation conducted by Diana Tabatabai in the Faculty of Education, McGill University. She can be reached at: dtabat@po-box.mcgill.ca for any questions. Her faculty advisor is Dr. Bruce Shore, and he may be reached at shore@education.mcgill.ca. I have been informed that the *purpose* of the research is to investigate how we search for information on the Web. I understand that *participation* in this research is voluntary, anonymous, and confidential. I may withdraw from participation at any time for any reason and without any consequences. My compliance or refusal of participation will not affect my grade or my job in any way. My responses are confidential and results of analysis will be grouped to maintain anonymity. There is no penalty for not participating or withdrawing anytime after I agree to participate. I understand that *procedures* of collecting the data are:

1. Filling out a background questionnaire – 5 to 10 minutes.
2. Performing the task. I will be asked to search for information on the Web and talk aloud as I search. This session will be audio and video-recorded for later analysis. If I want, transcripts will be sent to me for feedback and final approval. This part takes ½ hour.
3. Semi-structured interview following the data collection – 10 to 15 minutes.

It is anticipated that the entire time requirement will be approximately 1 hour scheduled at my convenience. There are no foreseeable risks except getting tired. I can take a break anytime. Refreshments will be available. The personal benefits for participation include receiving feedback from the researcher on my search strategies, learning about how my peers use the Web, and a chance to receive a McGill mug. The researcher will be the only one with access to information collected on the identity of participants. I have carefully studied the above and understand my participation in this agreement. I freely consent and voluntarily agree to participate in this study

Name: ______________________________

Signature: ___________________________ Date: ________________

Appendix E: Background Information Questionnaire

Name ________________________________ **Age**____ **Gender**___

If student: Total # of years of university ______What is your major? ________

If working: Total # of Years of Experience ____Years at present job: ____Title: _____

If teaching: # of years of teaching: _____

How would you rate your level of computer literacy? (Please circle only one)

Novice 1_____2_____3_____4_____5 Expert

How would you rate your knowledge of information seeking? (Please circle only one)

Poor 1_____2_____3_____4_____5 Excellent

How would you rate your knowledge of the Web (i.e., creating Web pages, JAVA, etc..)

Poor 1_____2_____3_____4_____5 Excellent

How did you learn to use the Web? Please give as much detail as you can.

1. Self interest (self taught or formal training?)
2. Required (in a course or on the job?)
3. Other (please explain)

On average, how many <u>hours a day</u> do you work with a computer? ________________________

On average, how many <u>hours a day</u> you spend on the Web? ____________________________

What type of browser do you usually use?

Do you have a favorite search engine?

1. .....
2. .....
3. ....

What is the first thing you read in a site? ____

Give 3 criteria you use to judge a site:

1. .....
2. .....
3. ....

Appendix F: Post-task Interview Questions

1. How did you feel during the search/how do you feel now?
2. Are you satisfied with the results of the search?
3. Did you feel lost or confused at any time? How do you deal with any frustration that may come as a result of search on the net?
4. How would you finish these sentences?
    a. I learned that…
    b. I was very happy when…
    c. I didn't know that …
    d. I had the most difficulty when …
5. Do you have a routine you follow when you search the Web? What comes automatic to you without thinking about it?
6. Do you put a time limit for your search?
7. Are you patient? How long do you wait for a site to load? What do you do when you have to wait?
8. Are there distractions on the Web? How do you stay focused? On task? How do you deal with curiosity stops?
9. How do you keep track of your search?
10. What are your usual sources of information, the Web or the library?
11. How do you make a decision on which search engine to use? Do you use more than one search engine?
12. When you are in a site, what are the three most important criteria that help you judge its quality to separate relevant from irrelevant material?
13. When you get a list of hits, how do you select which one to focus on? What part do you read?
14. What are the most difficult problems for you when searching the Web?
15. If you meet a Martian today, how would you explain the concept of the Web, what is it?
16. How would you categorize the problem types most novices face?
17. Where does your understanding of the Web come from?
18. If you were going to do the search again, would you change anything?
19. Was this search a typical (representative) of what you usually do in your everyday search?

Appendix G: Frequencies of All Codes

Code	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10		I1	I2	I3	I4	I5	I6	I7	I8	I9		E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	
1	3	41	23	8	17	9	35	20	16	4	176	19	38	11	18	9	21	19	11	10	156	49	37	7	33	61	9	8	17	20	12	253
11	0	15	5	0	2	1	0	1	0	0	24	1	12	6	1	0	4	1	0	6	31	8	2	1	16	10	2	0	7	9	2	57
111		6				1		1			8		2							2	4	4	1		2	1			3	4		15
112		9	3								12	1	10	6	1		3	1		4	26	3	1		12	8	2		3	5	2	36
113					2						2						1				1											0
114			2								2										0	1		1	2	1			1			6
											0										0											0
12	0	15	12	3	9	6	32	5	13	1	96	11	14	3	9	7	11	13	6	4	78	27	13	1	10	28	2	4	6	4	7	102
120			1				1				2						1	2			3	1				2						3
121		2			1	1	5	1			10	2							1		3					1			1		1	3
122		4	1		5	3	8	2	2		25	1	1			4	3	2			11	5	2	1	1	4	1		1	1		16
123		4	1		1		4		2	1	13		2		1		1	1	1	2	8	5				1		1				7
124			1			1			2		4	2	1				1		1		5	1	2		1		1					5
	0	10	4	0	7	5	18	3	6	1	54	5	4	0	1	4	6	5	3	2	30	12	4	1	2	8	2	1	2	1	1	34
125			2						1		3	1			2		2	2	1		8	4			1	5				1	2	13
126		2			1		6		1		10		3								3	3	4			3					1	11
127		1	2	1		1	1	1	1		8	3	5	1	3	2	3	2	1	1	21	8	1		6	11			2	1	1	30
128			1				2	1	4		8				1			2	1		4		1			1					1	3
129		2	3	2	1		5				13	2	2	2	2	1		2		1	12		3		1			3	2	1	1	11
	0	5	8	3	2	1	14	2	7	0	42	6	10	3	8	3	5	8	3	2	48	15	9	0	8	20	0	3	4	3	6	68
13	0	0	0	0	0	0	0	1	0	1	2	0	1	0	0	0	1	0	0	0	2	1	0	0	2	8	0	0	0	0	0	11
130								1			1		1				1				2				1	3						4
131											0										0	1			1	5						7
132										1	1										0											0
											0										0											0
14	0	1	4	0	0	1	1	0	1	0	8	0	0	0	2	1	3	1	0	0	7	1	12	1	0	6	0	0	1	0	1	22
140		1	2								3				1	1					2		8			4			1			13
141			2			1					3						2	1			3	1				2						3
142							1		1		2				1		1				2		4	1							1	6
											0										0											0
15	3	9	2	3	6	1	2	13	2	2	43	5	11	2	6	0	2	3	4	0	33	12	10	4	5	9	5	3	2	7	2	59
150	1	2			1	1	2	1		1	9	2	2	1				1			6	4	4	2	1	5	1			3		20
151		1						3	1		5	2	5		3				1		11		2		2	1	1	2				8
152		2			1						3			1							1	2								1	1	4
153	2	4	2	3	4			9	1	1	26	1	4		3		2	2	3		15	6	4	2	2	3	3	1	2	3	1	27
											0										0											0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
160											0										0											0
161											0							1			1											0
											0										0											0
170		1		2							3	2				1			1		4							1	1			2
											0										0											0
2	48	40	44	75	58	43	93	93	84	54	632	85	73	80	42	65	43	67	44	35	534	34	77	62	52	63	81	61	46	64	56	596
21	35	33	37	62	43	40	73	72	66	42	503	67	50	70	29	52	38	57	39	34	436	30	68	56	39	57	63	49	29	43	52	486
210	6	9	8	17	11	7	29	18	21	16	142	17	15	17	7	15	7	22	14	12	126	7	17	19	12	8	15	13	6	16	10	123
211	1	2	1	1	2	1		1	2	2	13	4	2	2	1	1	1	2	1	1	15	1	4	1	1	3	1	3	2	1	2	19
212		1					1	2			4					3					3			1			1	1	2			5

213	6	8	8	18	7	11	4	9	18	6	95	7	4	14	5	6	7	6	6	2	57	4	5	15	10	12	13	9	3	4	11	86
214											0										0											0
215											0										0	3				7	1					11
216			1					2			3										0		1				1		1	1		4
217	20	10	15	24	14	17	36	24	23	14	197	35	24	32	15	21	10	24	17	16	194	14	33	18	10	22	25	17	9	20	29	197
218	1				4	2		10	1		18	4									4		2	2	1	3			1			9
219	1	3	4	1	3	2	3	6	1	4	28		5	5	1	6	13	3	1	3	37	1	6		5	2	6	5	5	1		31
2110				1	2						3										0							1				1
											0										0											0
22	4	1	1	5	6	1	5	15	8	9	55	6	4	5	5	1	1	3	3	1	29	1	7	2	4	3	1	5	6	4	2	35
220	4			3	4	1	1	4	4	2	23	2	2		5		1	2	2	1	15	1	2	2	1		1	4	3	4	1	19
221		1		1			3			3	8		1	5		1		1			8		4		2				2		1	9
222				1	1		1	2	2	4	11										0							1	1			2
223								4			4								1		1		1									1
224					1				1		2										0											0
225								2	1		3										0											0
226								3			3										0											0
227			1								1	4	1								5				1	3						4
											0										0											0
23	5	2	4	5	5	2	14	4	5	3	49	7	17	5	6	10	3	7	2	0	57	2	2	3	5	2	14	5	10	16	1	60
230		2		1			3	1	1	1	9		6	1	1	5	1	2	1		17		1				7	1	9	7	1	26
231	2		2	2			5	1	1		13	1	5	1	2			2			11	1			1		3	1		1		7
232			2		2	2	4	2	1	1	14	2				1					3	1	1	3	1	2	1	1		1		11
233											0	1		1							2							1		1		2
234											0										0											0
235				2	2				2		6	3	1		1		1				6				1		2					3
236										1	1			1			1	3			5							1				1
237					1						1		3		2	3			1		9				1		1		1			3
238	3										3		1	1							2				1					6		7
239							2				2		1			1					2											0
											0										0											0
24	4	4	2	3	4	0	1	2	5	0	25	5	2	0	2	2	1	0	0	0	12	1	0	1	4	1	3	2	1	1	1	15
240					1				3		4	2	1			2					5											0
241	1	2	1	3				2			9		1				1				2	1		1	4			1	1			8
242	1	1	1		1		1		2		7	2									2						1	1			1	3
243	2	1			2						5	1			2						3					1	1			1		3
244											0										0						1					1
											0										0											0
3	12	45	31	81	54	30	37	83	29	49	451	55	40	32	53	66	53	22	22	27	370	48	30	25	36	53	65	23	41	79	38	438
31	2	8	0	8	3	6	1	13	0	8	49	4	0	6	11	5	2	6	0	7	41	10	7	3	6	8	3	1	10	21	6	75
310		8		7	1	6	1	11		7	41	3		4	3	1	1	1		3	16	7	3	2	4	4	3		3	13	3	42
311	1							1			2				3	1		2		3	9	3	2	1		1			2	3		12
312								1		1	2	1		1	5		1	2		1	11		2		2	1			4	3	1	13
313	1			1	2						4			1		3		1			5					2		1	1	2	2	8
											0										0											0
32	5	28	31	59	42	19	35	64	24	33	340	38	30	26	41	56	49	14	21	19	294	36	20	22	27	40	53	20	29	58	29	334
320	5	7	19	26	13	8	18	16	8	13	133	24	14	11	20	25	13	4	10	8	129	8	8	10	11	24	23	12	11	30	11	148
321		16	3	5	12	4	5	9	5	3	62	4	5	8	8	17	16	6	7	4	75	24	5	8	9	8	10	2	6	10	7	89
322		4	8	26	16	7	10	39	11	17	138	6	6	3	12	9	7	1	3	5	52	4	7	4	7	6	14	4	8	15	6	75
323		1	1	2	1		2				7	4	5	4	1	5	13	3	1	2	38					2	6	2	4	3	5	22
											0										0											0
33	4	6	0	13	8	5	1	3	4	7	51	11	8	0	1	5	1	2	1	0	29	1	2	0	1	5	7	2	1	0	3	22

330					2			2	1	5	10	2	1			1					4	1				1	1	2			3	8
331											0					1					1		1		1		2		1			5
332								1			1	1	1			1	1	1			5		1				2					3
333	4	6		13	6	5	1		3	2	40	8	6		1	2		1	1		19					4	2					6
											0										0											0
340	1	3		1	1			3	1	1	11	2	2				1			1	6	1	1		2		2		1			7
											0										0											0
4	9	22	9	4	0	4	15	4	9	0	76	12	25	13	18	24	14	18	14	13	151	38	19	12	19	19	5	2	8	17	10	149
41	8	15	7	3	0	3	10	4	8	0	58	6	17	11	12	16	10	13	9	9	103	24	13	7	10	12	4	2	5	13	6	96
410		2	1	1		1	1		2		8	1	5	5	2	4		1	1		19	7	5		3	3				1		19
411			1				1	2	1		5					1	1	1			3	2	1	1		3			1	2		10
	0	2	2	1	0	1	2	2	3	0	13	1	5	5	2	5	1	2	1	0	22	9	6	1	3	6	0	0	1	3	0	29
412	1	9				1	2		2		15	1	7	1		3	3	4		1	20		3	1		2			1	2	1	10
413							1		2		3		1	1			1	1	1		5	2	2	1	5	1	2		1		2	16
414	1	2	1	1		1	3				9		2	1	6	4	3	5	6	5	32				1	2	2	1	1	7	1	15
415		2	1	1			1				5	2	1	1	1	1	1		1	1	9					1		1	1		2	5
416							1		1		2					1	1	1			3											0
417	6		3					2			11	2	1	2	3	2				2	12	13	2	4	1					1		21
	8	13	5	2	0	2	8	2	5	0	45	5	12	6	10	11	9	11	8	9	81	15	7	6	7	6	4	2	4	10	6	67
42	1	7	2	1	0	1	5	0	1	0	18	6	8	2	6	8	4	5	5	4	48	14	6	5	9	7	1	0	3	4	4	53
420		2	1	1			1		1		6	3	3		3	3		1		2	15			2	3	1				1	1	8
421		5	1				3				9		5	1	3	4	4	4	5	2	28	14	6	2	5	6	1		3	3	2	42
422	1					1	1				3	3		1		1					5			1	1						1	3
											0										0											0
5	31	51	57	48	58	23	40	52	36	44	440	46	52	55	89	88	65	45	45	39	524	67	96	61	56	56	100	43	33	112	62	686
51	3	12	14	1	6	2	10	2	6	2	58	8	12	14	20	15	7	9	1	10	96	0	25	8	21	17	29	4	6	32	15	157
510											0										0											0
511											0										0							1			2	3
512											0	1									1				2		2				1	5
513		11	14		3	2	6	2	5	1	44	3	4	6	7	6	6	9		8	49		14	4	9	8	5	2	4	19	8	73
514	1	1		1	1	0	2		1		7	1	1	3	1				1	1	8		5	3	1		1	1		1	1	13
515										1	1	1	5	5	8	8					27		4		7	9	21		1	8	2	52
516	1										1	1				1					2		2	1	2				1	3	1	10
517	1				2		2				5	1	2				1			1	5											0
518											0				4						4									1		1
											0										0											0
52	17	18	20	26	27	10	9	22	9	19	177	11	15	9	27	34	30	13	12	17	168	34	23	25	18	15	9	16	8	22	7	177
520	4	1	3	6	6		4			8	32			1	4	15	14	2	3	1	40	5	3		1	4				4	3	20
521											0			1							1								1		1	2
522			2								2					1	1	1			3	1	2		3						1	7
523	1	2	1		6	3	1		1		15	1	1			1				1	4		1							1		2
524		1			2						3	1	1					1			3				1							1
525		1	1								2				1					3	4	1	1								1	3
526	8	9	12	2	4	6	3	12	4	4	64	5	13	4	20	17	12	5	8	10	94	17	9	16	5	8	3	2	5	11		76
527	4	4	1	18	8	1	1	10	4	6	57	4		3	1		2				10	10	6	8	8	3	6	14	2	2		59
528					1						1						1				1		1									1
529										1	1				1			4	1	2	8			1						4	1	6
											0										0											0
53	10	16	9	21	23	8	14	25	15	23	164	26	21	23	30	32	23	16	27	11	209	25	40	23	11	17	48	19	18	52	28	281
530				1	1			4			6																1					1
531	3	4	2	2	10	3	2		4	6	36	7	10	10	15	11	6	5	5	1	70	10	16	4	3		16	7	6	19	6	87
532	7	11	6	18	12	5	10	21	11	17	118	16	11	12	13	18	15	11	18	9	123	15	23	19	8	17	23	11	9	29	22	176

533		1	1				2				4				1	1			3		5		1				1		1			3
534											0										0											0
535											0	3		1	1	2	2		1	1	11						7	1	2	4		14
											0										0											0
54	1	5	14	0	1	3	7	3	5	0	39	1	4	8	9	7	4	7	5	1	46	8	8	5	3	7	14	4	1	6	12	68
540							1				1										0		2				9					11
541						2					2				1						1									2		2
542											0				1						1										4	4
543	1	5	11		1	1	6	3	5		33	1	4	8	6	7	3	7	5	1	42	5	6	5	1	7	5	3	1	3	7	43
544			3								3				1		1				2	3			2			1		1	1	8
											0										0											0
55	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
550											0										0											0
551		1									1										0											0
552											0										0											0
553											0										0											0
560					1				1		2			1	3		1				5				3							3
											0										0											0
6	1	1	4	1	1	3	4	1	2	1	19	5	4	3	4	4	9	1	2	2	34	16	5	0	9	16	3	0	0	8	15	72
61	1	1	2	0	0	0	2	0	0	0	6	3	2	2	0	1	5	1	0	0	14	2	2	0	2	4	0	0	0	2	1	13
610	1		1								2		1								1	1	2			2						5
611											0										0											0
612		1	1				2				4	3	1	2		1	5	1			13	1			2	2				2	1	8
											0										0											0
62	0	0	2	1	1	3	2	1	2	1	13	2	2	1	4	3	4	0	2	2	20	14	3	0	7	12	3	0	0	6	14	59
620				1		2			1		4	1									1					3	2			2	5	12
621			1							1	2					1	1				2	7				1					1	9
622			1		1	1	2		1		6	1	2	1	4	2	3		2	2	17	7	3		7	8	1			4	7	37
623								1			1										0										1	1
	104	200	168	217	188	112	224	253	176	152	1794	222	232	194	224	256	205	172	138	126	1769	252	264	167	205	268	263	137	145	300	193	2194