The Impact of Concept Map Visualizations on the Information Behavior, Perceptions of Performance, Learning and Use with Novices in the Information Retrieval Context

A Thesis

Submitted to the Faculty

of

Drexel University

By

Jodi Christine Williams

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

November 2007

Dedications

To my Mom and Dad; Carol J. & George E. Williams - thanks for life.

Acknowledgements

Michael Atwood, PhD, thesis advisor, who intrigued me with system design, introduced me to Simon's ants on the beach, and worked with me each step of the way.

The invaluable members of my committee each contributing differently and working so wonderfully together:

- Mary Brown, PhD, who started me on this path long ago and throughout it all encouraged and supported me in many ways.

- Xia Lin, PhD, who opened doors of possibilities through IR interfaces.

- Scott Robertson, PhD, for statistical help, support and an always sympathetic ear.

- Howard White, Ph.D. for willingness to work with me, high standards, and always telling it like it is.

Bri-Bri, Frew, Tricia, Lisa, Yolanda, & Min W., the best support group from the very first day.

Most importantly, for the support of family and friends and everyone who listened, encouraged and supported me each in their own unique way: Mooma, Poopa, Tracey, Sean, George, Jesse, Grandma-ma, Man, Beaner, Jennifer, Queenie, Kipper, Di-Di, Captain, Duncan, Donna, Jerry, Ken, Grace, Tom^5, Vivian, Mark, Patrick, Christine, Candy, Constantine, Karl, Vera, NCC, UMA, the Katz Crew, Jay & and all the wonderful students I have worked with over the years. Thanks!

Table of Contents

List of Tables

List of Figures

Abstract

The Impact of *Concept Map Visualizations* on the
Information Behavior, Perceptions of Performance,
Learning and *Use* with Novices in the Information Retrieval Context

Jodi C. Williams
Michael Atwood, Ph.D.

In examining undergraduate students in the information retrieval environment for the impact of computer generated concept maps, two primary research questions were considered: 1) what is the impact of display type on the novice searcher's information behavior; and 2) what is the impact of different display types on the user's perceptions of performance, knowledge and overall use of the system.

Sixty participants in this experiment were given hypothetical information needs on two different medical topics (cholesterol, depression). Participants' explored one of three interactive visualization displays using these medical topics, answered a pre- and post-test instrument and then completed a final questionnaire on their perceptions of the displays. Different types of inferential statistical tests were used to examine the research questions. When appropriate, factorial ANOVAs, mixed between-within ANOVAs, and chi square tests of independence were conducted.

Five main findings resulted from this research: 1) for all display types (LIST, SOM, PFNET) there is an increase in the number of participant search terms and in the incorporation of MeSH terminology from the visualizations following exposure to those displays; 2) there is a relationship between the display type and the interface level from which PFNET participants chose terms; 3) searchers' feelings of confidence, satisfaction, success, and relevance increased across all groups after system interaction; however, pre-test feelings of confidence and satisfaction seem to be dependent upon the participant's self-reported prior knowledge of the search topic; 4) feelings of confidence and satisfaction on the topic participants reported less pre-test knowledge on (cholesterol) shifted to match post-test ratings of confidence and satisfaction on the topic they had more pre-test knowledge on (depression); and 5) participants rated the PFNET system more visually appealing, easier to understand and more likely to be used in the future if given the option. Overall findings suggest that all displays were useful to the participants in this experiment and that the PFNET display was particularly useful for the novice searcher.

CHAPTER 1: INTRODUCTION & GOAL OF THIS RESEARCH

INTRODUCTION

Information Visualization is the “use of computer-supported, interactive, visual representations of abstract data to amplify cognition” (Card, Mackinlay, & Shneiderman, 1999). A basic underlying assumption (Koshman, 2006) and one of the claims by researchers and developers of visualization systems is that by using the brain’s perceptual system for processing information, visualizations can present complex and abstract concepts in an intuitive and more readily understood visual manner, while supporting large numbers of perceptual inferences that are easy for humans (Larkin & Simon as quoted in Card, Mackinley, & Shneiderman, 1999, p. 16). Coupled with an information retrieval system, it is possible that visualizations can allow users to move past retrieval of simple bibliographic entries toward the discovery of semantic relationships within and among documents and related concepts; it might impact behavior, assist with choosing better terminology, support confidence and support learning in the topic area they are searching. Imagine not only being able to search for text, but to also have a visual to aid in your searching; much like reading a picture book where the picture conveys part of the story and the text fills in gaps. These visualization systems can expose both experts and novices to relationships typically hidden behind the system’s standard interface. Data visualizations are intended to help reveal structures that cannot easily be recognized in any other way (Cleveland, 1993). If a visual information retrieval interface with computer-generated concept maps can reveal conceptual patterns, organize information in a meaningful manner, and connect the searcher with the information landscape as an expert might view it, then understanding how that visualization display impacts the novice is important.

Visualization techniques are typically interface displays which present information in a graphical manner which enhance cognitive capabilities (Card, Mackinley, & Shneiderman, 1999). Visualization graphics may take, for example, the form of a puzzle-like structure of boxes, each labeled to suggest the content it represents, or a spider-web-like network of lines connecting content labels. Systems like NewsMap developed by Marcos Weskamp use a puzzle-like visualization called a treemap to reflect the changing landscape or density of articles on a topic covered by Google News (Newsmap. Retrieved from the World Wide Web on December 1, 2007. http://marumushi.com/apps/newsmap/.)

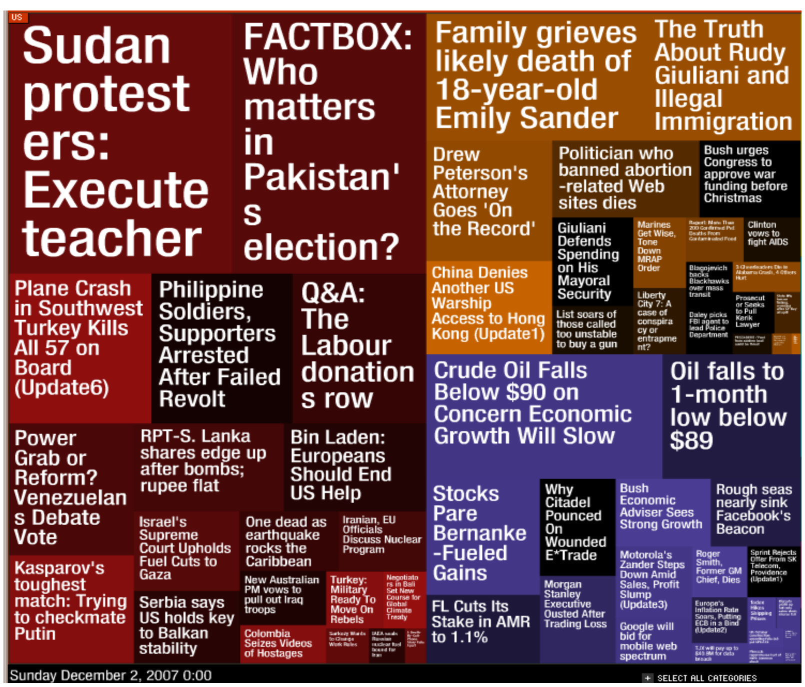


Figure 1. NewsMap is a Web-Available Visualization Application

Figure 2. Kartoo is a Web-Available Searching Tool

Retrieval systems with visualization components like meta-search engine KartOO display the organizational structure of search results in the form of a concept map with two-tone dark blue amoeba-like areas behind small to medium-sized iconic representations of the web page. Within these amoebic structures overlapping like fish strung on a line, words connect the web pages to the terms and when your mouse hovers over the area, lines are displayed showing the interconnectedness of the topics. According to Norman, “real powers come from external aids that enhance cognitive abilities” (Norman, 1993, p. 43). These cognition-enhancing tools have become a focused area for research not only for applications to information retrieval, but to the broader applications of information science and systems.

For a moment let us consider novices and experts searching with the same information retrieval system in an academic library. Bransford, Brown, and Cocking (1999) highlight key principles in experts' knowledge: experts notice meaningful patterns of information; experts have acquired a great deal of content knowledge; that knowledge is organized and reflects a deep understanding of their subject matter; and domain experts have specific skills that allow them to assess, process, and understand a problem differently than that of a novice (Bransford, Brown, & Cocking, 1999). An expert notices, organizes, processes and interprets information in their environment differently than a novice. In an information retrieval environment, those skills in turn make them more successful in their searching (Hsieh-Yee, 1993; Shute & Smith, 1993; Sihvonen & Vakkari, 2004; Wildemuth, 2003). This is in contrast to the novice searcher who, by virtue of their beginning state, does not have the skills or content knowledge to assess, process, or understand a problem in the same manner the expert does. To reconnect with the visualization discussion, we might ask how the skills and traits of the expert can be shared with the novice. An interactive visualization, in the form of a computer generated concept map, can display how an expert might perceive the document space and can provide important proximal and distal clues to the novice. Hence the "real power" Norman is talking about is in those aids for enhancing cognition (Norman, 1993, p. 43).

The role and importance of providing different kinds of cues to the novice searcher can be illustrated by the following scenario. As a child I used to gather hickory nuts in the autumn with my Grandfather. We foraged for nuts much like prehistoric ancestors would have foraged using the clues nature provided to aid us in that task. My Grampa told me what to look for and how to successfully go about this. He didn't just show me a hickory nut, rather he pointed out the hickory trees by their shape, their shaggy bark, wide oval pointed leaves, and giant green shuck casings at the base of the trees. Our competitors for hickory nuts were gray bushy-tailed squirrels who spend their time nibbling, chewing, and sorting through their stores. The squirrels drop the chewed shell and husk fragments close to the trunk which obscures the good nuts close to the base of the tree and make it VERY painful to kneel in that area. Because the pain, sifting process, and overall workload is more energy than it is worth the closer you get to the trunk of the tree, my Grampa pointed out the best concentration and yield of good nuts was ¾ a distance from the tree trunk and we would begin foraging there and work in a circle around the tree. This was our nut-foraging landscape and all the clues I mentioned above were important to our success. It helped my Grandfather and me to minimize effort and receive maximum results.

Chewed shell and husk fragments were "proximal clues" we recognized, and acted upon in order to get the most gain for our output energy. While we were foraging for food, information foraging is a theory by Pirolli and Card adapted from and grounded in anthropology and computational theories of human cognition. It has been used to explain human information seeking behavior. In an article by Rachel Chalmers in "New Scientist," experts who have studied human foragers agree that "foraging on the web presents trade-offs analogous to those of hunter-gatherers"… it is a different context but a similar cost-benefit analysis. In the information seeking behavior area, Card et al. suggest that the "proximal perception of information scent is used to assess the profitability and prevalence of information sources". Those assessments also inform decisions about which items are pursued in order to maximize the "information diet" of the forager (Pirolli & Card, 1998).

The nut-gathering story about my Grandfather can help to highlight important facets of the information seeking context. The distal and proximal clues we relied upon provided the landscape which allowed us to locate and maximize our results. If I asked someone who had never gathered hickory nuts, or any kind of nut to go gathering, would they know where to start? Knowing the characteristics of a hickory nut might be helpful, but it isn't enough. If we think of the hickory nuts as documents in an information retrieval system, my Grandfather might be considered the librarian. The distal and proximal cues we followed while foraging for nuts in this analogy would be the cues traditionally followed in the traditional library setting; indices, thesauri, card catalogs and bibliographic data like author, titles and dates.

To place the need for better retrieval systems within a specific information seeking context, for a moment let us picture the traditional library of the early 1980's. Common information tools and resources historically available to aid searchers in navigating and understanding the domains of retrieval systems included librarians, card catalogs, indices and thesauri. Card catalogs supported browsing habits, and (when used) thesauri and indices exposed domain organizational structure and search terms for broader, related or narrower topics. These tools provided conceptual maps helping with not only term selection but also with structure and articulation of the search query (Vakkari, 2002). If those tools were not known or used, the librarian was another guide for the user to connect her to the information stored in the library.

Now let us fast forward to imagine a modern academic library filled with computer systems readily accessible to any user walking through the door and available from home. The card catalog, indices, and other tools have been either been removed from the library, circumvented, or are hidden behind the structure of a computer-based retrieval system. The librarian is frequently bypassed as users now have direct remote access to information retrieval systems from home (Carlson, 2001). Your average academic librarian would agree paper-based indices are typically no longer used and their existence unknown to the average searcher. Even if searchers are aware of such tools, there is very little use of controlled vocabulary or of the database dictionaries as Efthimiadis found in users searching a CD-ROM database (1994). Seeing and having access to tools which would provide the context of a search topic within the wider information domain used to be much more accessible. Resources which would help the user place their need within a context, which would in turn help them better understand their need and state it in more appropriate terms for the system, are bypassed or not used when present partially because many users are accessing the library from remote locations (Franklin & Phum, 2004).

Complicating the barren landscape is the proliferation of available electronic information. Recall my Grandfather and I stayed away from the base of the tree because it was obscured by shells, fragments and worm-eaten nuts. Finding a good nut within that area would have been more time-consuming than it was worth. The "Principle of Least Effort will minimize the effort required to obtain information, even if it means accepted a lower quality or quantity of information" (Case, 2002, p. 143). Research has demonstrated that people rely on close friends and relatives and oral channels for information (Dervin, as quoted in Case, 2002, p. 142; Durrance, 1988). Because the yield of documents returned on a search using the term diabetes might be over 543,879, the good nuts become obscured. We know from Simon's Nobel winning work in economics that people have a limited ability to process and evaluate information, that given all possibilities presented, and the cognitive limitations, a person will opt for a "good-enough" solution (Simon, 1996, p. 27; Agosto 2001; Prabha, Connaway, Olszewski, & Jenkins, 2007). Is "good-enough," pertinent information? Is "good-enough" for a novice searcher the best choice or the most accurate? Regardless of whether good-enough is pertinent information, or the best, in the modern information environment the good nuts are obscured by all the thousands of nuts available.

If we take away the clues we followed in finding good, edible nuts, and do the same for the novice searcher in the library environment, essentially we have a barren landscape with no cues on how to find the information needed. A system that provides the title, author, publication and abstract after I have entered one or two terms might be helpful, but it isn't enough. Take away the librarian, hide the information structure, strip the landscape, and you have a clueless nut-gatherer.

The organizational structure of the information in today's information seeking environment is typically not accessible to the searcher. Instead of a tool for analysis and discovery of semantic relationships within and among documents, the information retrieval system standard has been more like a simple search mechanism (Lin, Soergal, & Marchionini, 1991). The systems do not provide any clear indications of the relationships among retrieved documents (Korfhage, 1991). Systems should provide associations beyond conventional thesaural relationships to allow alternative paths for accessing documents (Bates, 1986). At the very foundation of looking for information, “the basic problem is to increase the mental contact between the reader and the information store so that the reader can proceed unerringly and swiftly to identify and receive the message he is looking for.”(Doyle, 1961, p. 553). Similar to working with my grandfather, give the searcher enough information to permit her to narrow her focus by recognition (Card, Mackinlay, & Shneiderman, 1999), to help her make an informed decision. Give her, as the primary searcher, the information she needs to understand and place her information need within a landscape. Show her the trees, the squirrels, the ground beneath her feet. Give her those things and you give her additional keys to aid in judgment and decision (White & McCain, 1997) and the keys to help her find what she is looking for, not just what is “good-enough” (Simon, 1996 p. 27).

Information retrieval systems have grown with the power of computing. Automatic indexing, ranking algorithms, and other tools seek to improve the quality of the returns without searcher intervention, understanding or even knowledge of these tools. With automatic Boolean, automatic query expansion, and other system-side tools it might be said that the demand on the user, for what they need to bring to that system, is less and less. If these advances in IR systems provided more clues and means of cognitive support through the search process for the end-user it might be argued that a lesser demand was a “good thing”. However, the bulk of modern systems do not provide these cues for the information forager. What appears to be more the case, particularly with the novice searcher interacting with a retrieval system, is a volley; input some terms, get some returns, input more terms, get more returns. There needs to be a shift from "retrieval" to "display," that moving from the "query answering system to an information organization and display system would" better support the searcher (Korfhage, 1991).

Additional aids to judgment and decision (White & McCain, 1997) in the form of a concept map might support Jesse, a first year undergraduate student, looking for medical information. Automatically extracting pertinent information and displaying it in a manner the expert inherently would (Buzydlowski, 2003), might help Jesse recognize related concepts and the semantic relationships between them by the use of a visualization display (Card, Mackinlay, & Shneiderman, 1999). It might help her choose more precise terminology, have more confidence and satisfaction in those terms, and in turn help her better understand the domain she is searching. "People could manage more powerful searches quickly if an initial submitted term or topic yielded a screen full of term possibilities, related subjects, or classifications for them to choose from (Bates, 1998).

In this research we seek to understand the impact computer-generated concept maps have on 1) the information behavior, 2) perceptions of performance; 3) knowledge, as well as 4) perceptions of usefulness in the information retrieval environment. Does the use of a computer-generated concept map impact how a novice searcher constructs their search query, does it impact the terms they choose and how specific they are, does the searcher learn while they are searching and enrich their knowledge about a topic as they are looking for information? Does the system involve them in the search process such that they would be willing to engage in using that system in the future?

PURPOSE

The purpose of this research, using a between subjects experimental design, is to measure the impact computer-generated concept maps have on the information behavior, perceptions of performance and knowledge, as well as perceptions of usefulness, of the novice, undergraduate student in the information retrieval environment. Participants in this experiment explored one of three displays using *VisualConceptExplorer* (VCE), a real-time information visualization system attached to a medical database and accessible through the World Wide Web. The VCE system generates two different types of concept maps and an alphabetical list containing the 25 most highly co-occurring terms based on a seed topic. Using hypothetical information needs on two medical topics (cholesterol, depression), participants explored the three different display formats. Data was collected using a pre- and post-test instrument, a general background questionnaire and a reaction questionnaire on perceptions of assigned displays.

RQ INTRODUCTION

In looking at the impact of visualizations in the form of a computer generated concept map, the following research questions were considered.

RQ1: Information Behavior

Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher's information behavior?

If visualizations can display otherwise hidden information and through the use of our perceptual processing system amplify cognition, understanding how visualizations in the information retrieval environment impact a novice searcher’s behavior is important. This question seeks to understand how concepts maps might differentially impact the number of terms used by participants, the incorporation of display terminology, where they choose terminology from and the level of specificity of search statements.

RQ2: Self-Reported Perceptions of Performance

Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher’s self-perceptions of performance?

RQ3: Self-Reported Perceptions of Knowledge

Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher’s self-reported knowledge?

Research in the early 1990’s found that the affective feelings of searchers have an impact on overall search behavior (Kuhlthau, 1992). The affective reaction of a novice searcher in an unknown system, with a foreign display format is an important area to investigate. In our modern “Google World” precision and recall are not as important when it comes to focusing on the user. This research and the interest of the experiment is not focused on performance from a system perspective (precision and recall), but rather on the user and their behavior, self-perceptions of performance, learning and system usefulness. Therefore, precision and recall will not be used as metrics. More qualitative means of measuring relevance and user perceptions of information and the information environment rather than on concrete matching of topics will be used (Barry, 1994; Froehlich, 1994; Schamber, 1994; Schamber, Eisenberg, & Nilan, 1994; Saracevic, 1975). Perceived relevance, utility, and satisfaction, which represent overall assessments of system performance from the viewpoint of the user, will be employed (Börlünd & Ingwersen, 1997; Schamber, 1994; Schamber, Eisenberg, & Nilan, 1990).

RQ4: Overall Perceptions of Display Usefulness

Do concept maps, used as an exploratory tool in the information retrieval environment impact participants' overall perceptions of usefulness?

Evaluating perceptions of usefulness, in a new system is important. If a system does not engage a searcher and/or is more cumbersome to use than the information necessary it will not be used (Norman, 2004). This ties into the "law of least effort" (Dervin, 1983; Durrance, 1988; Case, 2002). The law or principle of least effort states that, "an information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it" (Mooers, 1959/1996). There is a relationship to Simon's bounded rationality theory as well. Given the possibilities and the cognitive capabilities, users have to satisfice and go with a "good-enough" answer or decision (Simon, 1996). This research question seeks to understand users overall reaction to the visualization display tools used. This covers areas on search formulation, understanding and sense of the display format as well as current and prospect of future use.

CHAPTER 2: LITERATURE REVIEW

INTRODUCTION

Two persons observing the same behavior or event can have slightly different perceptions of that behavior or incident. Consider eyewitness testimonies to a crime or accident. It is known that each witness' story will likely vary. Perhaps the stories will vary only slightly or they might vary dramatically. However the degree of variation, the composite and overlap will give the police an overall image of what happened (albeit perhaps still fuzzy). Each person's physical viewpoint as well as personal knowledge and experiences will influence how they perceived what happened, and in turn it will influence how they then retell the story.

Likewise, as we will soon see from the literature review, each observer of information seeking behavior and each field considering information behavior has a different perspective from which they perceive the behavior as well as its significance. That perspective in turn, like the eyewitness accounts, influences how the research or the field relates the story of what happens during information seeking. In turn, however, collective mosaics of understandings help to create an overall picture of information behavior.

As an example of the multi-faceted research literature, let us imagine Jesse, a student at a university library. She is a looking for medical information to write a paper for her health class and is typing into the computer. Now let us also imagine a group of people standing behind Jesse with notebooks. They are copiously writing in their notebooks as they observe her actions. Each of these people watches what she does through the different lenses of their discipline. There might be an information system designer, a psychologist, a librarian, and an information scientist. What those observers see and how they interpret and retell the story of Jesse's actions depends upon the individual lenses through which they are looking. The system designer might be eager to see how many times Jesse has to click before she reaches the article she is looking for, while the psychologist might be interested in why she chose a particular citation to write down and what cognitive process brought her to that decision, while the librarian is focused on what database Jesse chose to use and the terms she typed into the computer. This literature review, in essence, is a discussion of the different lenses through which the information behavior research views, in our example, Jesse. The different perspectives are each seeking to understand her behavior in different ways. The ultimate goal of all of the previous research and development is to sketch a picture that can be used to inform the greater whole so that systems can be developed which ultimately help her retrieve better.

A number of facets within the information science and cognitive psychology disciplines will be used to consider the impact of computer-generated concept maps on the novice, undergraduate student searching a computer-based retrieval system. Within the discipline of information science we will draw upon the perspectives offered by the literature in information behavior, information retrieval, visualization and concept map development as well as domain knowledge. Within the discipline of cognitive psychology we will draw upon the perspectives offered by research in visual processing, experts and novices, and learning.

To understand the novice behavior as well as the state of mind and information need of Jesse when she searches, the broader information behavior literature will be applied to place searchers within the context of interacting with an information system from the human perspective. This section also addresses the information retrieval (IR) literature focused tightly around the users' behavior and interaction with the system. The visualization literature addresses the use of tools to help users view large amounts of data visually and the need for better understanding of how these visualizations are perceived, their usefulness in real time IR situations, and their overall impact on the novice. Finally the cognitive psychology literature focuses our gaze upon the novice searcher and their lack of domain knowledge in contrast to an expert, and this literature also provides a lens to help us understand how visual information is processed and how that perception in turn can influence information behavior and learning.

To fully understand the factors influencing the information seeking behavior of the user, a general discussion needs to be drawn together: the information process, models of information behavior, and the context of information needs, and problems behind the stated need. It will also be necessary to consider some aspect of information retrieval and different measures used in IR to understand how the system can better support retrieval based on user behavior. This is where the visualization literature comes into play and is discussed as a means of supporting the information seeker by providing more clues to the literature through the use of visualization tools. The conclusion will briefly review the myriad of lenses through which we see the information behavior and finish with the rationale for the research questions this research explored. These lenses are many and varied each providing a slightly different picture of the information retrieval process and the background setting for this research.

INFORMATION SCIENCE DOMAINS

Information Behavior (Seeking & Retrieval) Perspectives & Models

Information retrieval by its very nature might be termed a paradox. On one side of the problem there is the user, with a need for information, and the behavior involved in locating information which often is a complex iterative process (Bates, 1989; Belkin, 1982; Chen & Dhar, 1991; Dervin, 1977; Ellis, 1989; Kuhlthau, 1993; Marchionini, 1989). Because it is a lack of knowledge regarding the topic that has brought the user to the system, how do they know what terms to use (Belkin, Oddy, & Brooks, 1982a, 1982b)? Most users input their query using very broadly stated concepts, which retrieve more documents only adding to the confusion (Bates, 1986; Ingwersen, 1982). Many early information systems were designed for the skilled information professional who would sit down to type in a well-formulated query using terms the system would recognize. The librarians and information professionals were familiar with the information space and the terms to use in order to retrieve documents effectively. That is to say, matching a query representation to a document representation is the act of matching, through the user's query, the conceptual presentations central to the performance of a information retrieval system (Fowler & Dearholt, 1990).

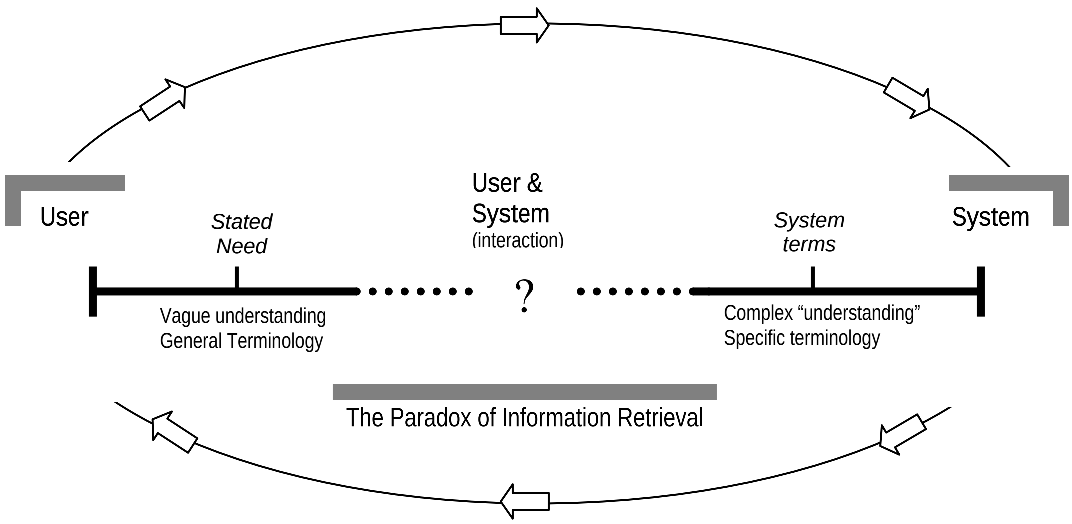


Figure 3. Information Retrieval Paradox

In the 1960's, researchers began to recognize that the user needed better tools to help them be more successful in retrieving desired information. While Doyle's (1961) work surrounded semantic maps, his vision was to increase the "mental contact between the reader and the information store so that the reader can proceed unerringly and swiftly to identify and receive the message he is looking for." Researchers continue to recognize the need in system design to open the information landscape to the user. Bates (1996) suggested that systems should provide information for users beyond "conventional thesaural relations." Marchionini (1995) also asks how the information in the computer systems we use should be organized to "reveal itself" to users. Saracevic (1995) in a discussion of IR evaluation methods also inquires, how do we organize information intellectually, how can the search interaction be constructed so that it is intellectual? Systems driven models and developments are very important and will be addressed in the system section; however, they only give a limited view of the overall information retrieval environment. The growth of a more user-driven holistic model is an important and still-evolving concept. The progression in the literature from focusing on just the information system toward the whole setting of system, people (cognitive), environment and work as well as the various contributing factors is discussed below.

Taylor's (1968) research, by focusing on how and why people come to look for information, helps to place the average user within a framework which explains why novices' queries tend to be "overly general" (Bates, 1998; Case, 2002, p. 69; Jansen, Spink, Bateman, & Saracevic, 1998; Sutcliffe, Ennis, & Hu, 2000): Taylor developed models focusing on user behavior and needs, describing the information seeking process as one where user needs transform throughout the searching process. In his theory, Taylor explains that a users' information need begins as an unexpressed need for information and that need becomes less and less abstract as the user gains more information. One of the key factors for this research was that users typically begin searching for information with "a vague sort of dissatisfaction" which is likely "inexpressible in linguistic terms" (Taylor, 1968). Taylor’s research helps to place the average user within a framework which explains why novices’ queries tend to be “overly general” (Bates, 1998; Case, 2002, p. 69; Jansen, Spink, Batemen, & Saracevic, 1998; Suttcliffe, Ennis, & Watkinson, 2000).

The focus on the stated need (or query formulation) as part of the information retrieval process, branches into a specific research stream within the information seeking literature focusing on query formulation and reformulation. In the very basic sense, the online searching aspect of queries can be broken into these two stages: 1) the initial query formulation and 2) reformulation(s) of the initial query (Efthimiadis, 1996). Query expansion involves removing or adding terms to the original query (Efthimiadis, 2000). There are three primary areas of research under the query expansion umbrella, they include manual, automatic and interactive (Efthimiadis, 1996). It is an important area as our research addresses system-supported query expansion with user control. Hsieh-Yee (1993) as well as Sihvonen and Vakkari (2004) found that more terms were used in reformulation when participants had access to a thesaurus. Other research, involving the primacy effect found that the order of presentation of results impacts query reformulation. Terms presented first were more likely incorporated into the reformulated query than terms and information presented farther down the list of results (Allen, 1994). Interactive query formulation, where the user has control over the system-suggested terms for query expansion, improves search effectiveness (Koenemann & Belkin, 1996; Sihvonen & Vakkari, 2004). In addition, real-time query expansion with user control has also been found to increase the general usage of query expansion and improved quality of initial queries, leading to higher satisfaction (White & Marchionini, 2007). Also, query expansion with user control, as opposed to automatic expansion is preferred by users (Belkin, et al., 2001; Brajnik, Mizzaro, & Tasso, 1996; White & Marchionini, 2007). For an extensive review of query formulation, Efthimiadis' (1996) review in ARIST covers the history and research. Overall, choosing terms to reformulate a query is a difficult task for the novice (Efthimiadis, 1996; Greenberg, 2001a). Our discussion of connecting the user with the information landscape during the information seeking process directly relates to one specific method of supporting query reformulation.

The focus on the user and their behavior as opposed to a focus on their stated need and how that need is reformulated after interaction with a system continues to evolve with the work of researchers who call for better understanding of the problem behind the stated need (or query) as it emanates from the user rather than from the system and a shift away from the system toward the user (Belkin, Michell, & Kuehner, 1980; Dervin, 1992; Dervin & Nilan, 1986; Saracevic & Kantor, 1988a, 1988b; Savolainen, 1993). The focal point for these researchers is on the problem driving the stated need. Dervin (1977), for example, began to see the information seeking process as one of reducing uncertainty and sense-making. The sense-making approach consists of a set of conceptual and theoretical premises as well as a set of related methodologies for assessing how people make sense of their worlds and how thy use information in those worlds during the information seeking process (Dervin & Nilan, 1986). The theory both has impacted how we look at the cognitive strategies used by seekers in problematic situations and has supplied a framework for research on information behavior (Savolainen, 1993).

Belkin & Oddy and Brooks' work also explores the problem behind the stated need but shifts slightly to focus closely on uncertainty. Their work found the motivating factor for information seeking to be a recognition that an anomaly in the knowledge state exists. Faced with this "anomalous state of knowledge" (ASK), a person will address his uncertainty by looking for information (Belkin, Oddy, & Brooks, 1982a, 1982b). The key discussion in this work is that a document in an information retrieval system is a "coherent state of knowledge," while the user's query or stated need is an incoherent state of knowledge (Belkin, Oddy & Brooks, 1982a). For example, a research article on *The effect of conjugated equine estrogen on diabetes incidence* is an expression of the author’s knowledge about a specific subject area. At the other end of the system you currently have Jesse who needs information for a paper on the different uses of horse estrogen, but with a lack of understanding about the domain (an anomalous state of knowledge), she states her need in an inadequate or incoherent manner (Belkin, Oddy, & Brooks, 1982a).

Taylor's and other researchers' assertions that user queries typically are very general and broad because the user stating their need is unfamiliar with the topic area they are searching is exemplified in the patron who comes in and asks the reference librarian for books on English Literature (broadly stated) and after a long extensive reference interview with the librarian, walks out with a book on Falstaff from Shakespeare's Henry IV (more coherent state). The anomalous state of knowledge or gap in the user’s knowledge is an important thread this research uses as a foundation. Novice searchers in a complex subject area like medicine have little or no medical knowledge. Therefore an information system with a system supported query expansion tool in the form of a visualization provides an ideal context in which to explore the impact on the information behavior of the novice searcher.

The focus in the literature shifts again from the concentration on the stated needs of the information seeker and the problem behind the need toward a more holistic approach. Belkin and Vickery (1985) note an information need is hard to study because it exists inside a person’s head and is inferred. To place the user within the wider information behavior framework and, thereby be able to better understand the social, environmental and cognitive context that impacts the seeking behavior and ultimately the search outcome, the entire behavior of the seeking process needs to be considered. To iterate this assertion, Wilson (1999) reviews different information behavior models that have emerged in the literature. He presents an illustration of a nested model of the information seeking and information searching research areas. The importance of this nested model lies in Wilson's reminder to researchers that the study of a particular topic needs to be undertaken in the context of the surrounding fields (Wilson, 1999). To tie our initial image of Jesse in the library at the computer, the views of all of the people observing are important and need to be acknowledged and addressed to get an even clearer picture of the whole.

Herbert Simon in Sciences of the Artificial would argue that human beings, seen as a "behaving system" are essentially quite simple and that the "apparent complexity of our behavior over time is a reflection of the complexity of the environment" (Simon, 1996, p. 80). This claim that human behavior is not really complex, and only appears as such as a reflection of the complex environment in which is it behaving can be intertwined with the need for a more holistic approach in research on information behavior Wilson and others recognized. The need for a more encompassing approach began to be filled in the literature through the later 1980s and 1990s. After the late 1980s we can efficiently divide the literature into cognitive approaches (Ellis,1992), focusing upon the thought processes; and the more holistic approaches which include the cognitive, affective and physical aspects of searching (Sugar, 1995; Hewins, 1990). The holistic approaches focus on identifying the user's characteristics, rather than measuring system performance (Hewins, 1990) and analyzing stated needs (Wilson, 1999). In order to address a wider view of information behavior it is necessary to incorporate the affective dimensions of user problems as well as the cognitive and physical (Kuhlthau, 1993). The focus shifts, then, from the specific information need (stated or perceived) to the whole stage of information behavior, that is, to the broader context of the whole system of which the searcher is one part.

Two key researchers who developed holistic models of the information behavior of the user are Kuhlthau and Marchionini. Marchionini's focus was upon the human-computer aspect while Kuhlthau's focus was more toward the cognitive aspects, particularly the affective constructive process of information seeking, independent of any system. Kuhlthau explored the information-seeking process in three realms: the affective (feelings), the cognitive (thoughts), and the physical (actions) incorporating what MacMullin and Taylor concluded in 1984; that a model representing a user's sense-making process should include those three components. Research on the affective aspects of the information seeking process have framed it as a process of reducing uncertainty and making sense(Dervin, 1977; Kuhlthau, 1993), and research has identified sharp increases in uncertainty and decreases in confidence after searches with novices were initiated (Kuhlthau, 2004). As defined by Belkin and Oddy's ASK model the process of interacting with the system is one of reducing uncertainty (Belkin, 1986; Belkin, Oddy, & Brooks, 1982a, 1982b).

While not focusing on the affective, Marchionini's (1995) holistic approach views information seeking as a directed process, driven by an information problem (p. 7), which humans purposefully engage in order to change their state of knowledge (p. 5). Marchionini and Kuhlthau have started to look at the whole process of information seeking, not just the user or just the system components, but rather the system, the person, the work, and the environment of the information behavior landscape.

Let us also be reminded that during the late 1980's and beginning in the early to mid 1990's the landscape of information retrieval started to change dramatically. It is not just the librarian or information specialist who is using the retrieval system to find information, more information and materials are being made available to the novice searcher and the average university student or library patron has access to these systems from the dorm, from home 24 hours a day, seven days a week. Those who have little to no knowledge of how the information is organized began to use systems initially designed for the experts. Because of the increase of available information and more open access, the information-seeking process became more a "gathering of sources rather than of hunting sources" (Blandy & Libutti, 1995). It isn't just about finding a right answer or a right document; it is how users seek meaning in the myriad of resources rather than how user's seek a right answer (Kuhlthau, 1993).

Recall Figure I, illustrating the paradox of information retrieval with the user on the left with a vague understanding of how to state their information need and the system on the right with an inherent way of representing its own content. This figure will help us recall the different focal points of previous research and see the still developing framework of lenses which have emerged. Previous research focused primarily on different characteristics of the searcher, the searcher's stated need, and the searcher in relationship to their information system interactions. Next there is a developing area of research, not only exploring the behavior and application of cognitive theories to behavior and the searcher, but also focusing upon the outside influencing factors which we might term as Simon's complex environment which he claims is the key to understanding behavior. The factors impacting Jesse's behavior are complex and interrelated.

Focusing more closely on the influencing factors of the environment in conjunction with the user involved in the information seeking process Marchionini and Shneiderman identify five components that affect the information seeking process. They are: the *setting* in which the seeking takes place; the *task domain* or body of knowledge which is composed of entities and relationships; the *search system* itself; the *user* and their mental models; and finally, the *outcomes* which include the products of the search such as articles, abstracts, etc. as well as the experience itself that becomes part of the user's knowledge for dealing with future information problems (Marchionini & Shneiderman, 1988).

Another similar framework for online searching by Fidel and Soergel included: setting, the user, the request, database, search system, searcher, search process, and outcome. They found that these factors are complex and interdependent (Fidel & Soergel, 1983). Some of the characteristics of the searcher/user found to influence the search process included subject background, education level, and prior experience. While Fidel and Soergel focused on intermediaries who conducted searches on behalf of patrons, these factors also apply to novice searchers who conduct their own searches (Fidel & Soergel, 1983; Marchionini, 1989). Another study on information seeking and retrieving identified five components particularly related to the cognitive context and interactions of the search process, they are: 1. the user; 2. the question; 3. the searcher; 4. the search; and 5. items retrieved (Saracevic, Kantor, Chamis, & Trivison, 1988; Saracevic & Kantor, 1988a; Saracevic & Kantor, 1988b). As with Fidel and Soergel, this study separated the searcher from the end user with the originating question (librarian from the patron). For simplification purposes and a clearer picture of the whole setting, we might think of the information seeking environment with the following key components as presented in the literature: the system, the people, the environment and the work/task.

To draw together and apply this part of the literature to a search for information, let us bring Jesse back into the frame. Jesse needed information for a class project but came to the system with a gap in her understanding which led to queries and strategies that tended to be simple and basic (Belkin & Oddy's work on ASK; Dervin's work). By the very nature of having a gap in her knowledge, Jesse cannot state what she doesn't know (Belkin & Oddy, & Brooks, 1982a, 1982b). The problem behind the need is in Jesse's mind and, therefore, is not a simple thought to measure (Belkin & Vickery, 1985). Jesse may state the need in simple terms like English Literature or Plays. The documents on the other side of the system are, however, a "coherent state of knowledge" (Belkin & Oddy, & Brooks, 1982a, 1982b) and are represented with more specific terminology, like *Falstaff from Shakespeare's Henry IV*. It is not only Jesse's state of mind and her construction of meaning which is fraught with uncertainty during the search process, or her stated need that we should examine, but rather the whole context of the process (Wilson, 1999). To do this we should use models of information seeking that exhibit a more holistic approach and provide a framework with which to look at behavior in the context of the whole setting. We need to consider the various elements that impact Jesse's search process including the environment, setting, task domain, search system (Fidel & Soergel, 1983; Marchionini, 1989; Saracevic & Kantor, 1988a; Saracevic & Kantor, 1988b; Hirsh, 1996).

This picture of the seeking process focuses and reveals a clearer image of the user, their need and their state of knowledge. It also suggests that the search system needs to aid the searcher in moving from an anomalous state of knowledge to a coherent state of knowledge during the searching process. This research proposes a method for not only acquainting the novice searcher with the information store through the use of visualizations, but also to measure the impact a visualization display has on information behavior and perceptions of performance, knowledge and use. In the next section we focus specifically on visualizations developed as a means of acquainting the user with the information store and visualization techniques that are components of retrieval systems.

Visualization

We shift to focus on one factor which is known to impact information behavior and is part of the information environment; the system. Our current retrieval systems generally conceal the landscape and organization of their information from the user and fail to reveal important and available clues that would seemingly help Jessie with query formulation as well as help her through the search process (Bates, 1996; Brajnik, Mizzaro, & Tasso, 1996; Korfhage, 1991). One of the foundational goals driving this research and the system it will explore is a means of finding a way to "increase the mental contact between the reader and the information store" (Doyle, 1961). Systems should provide users a means for exploring the relationships between terms, and terms and documents (Brajnik, Mizzaro, & Tasso, 1996). Vennevar Bush's, "As We May Think," called for "association of thought". Lauren Doyle's "semantic roadmaps to literature" cited the need for displaying meaningful associations among documents and related topics as well as applying that to information retrieval. Systems should provide associations beyond conventional relationships (Bates, 1986), and showing those relationships would give "additional aids to judgment and decision" (White & McCain, 1997).

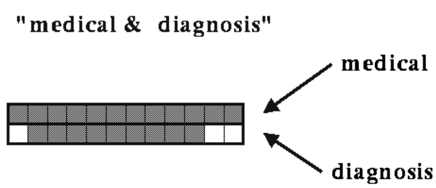
Our early discussion noted that the exploration and techniques to display these associations among documents is an area in information science called visualization. Information Visualization is the "use of computer-supported, interactive, visual representations of abstract data to amplify cognition" (Card, Mackinlay, & Shneiderman, 1999). We might look at visualization techniques as tools which present information in a graphical manner using the brain's perceptual system for processing information (Larkin & Simon as quoted in Card, Mackinley & Shneiderman, 1999, p. 16). A more system-focused definition of information visualization is to provide visual depictions of large information spaces (Hearst in Baeza-Yeates, 1999). When visualization maps help the user browse large general search space like the World Wide Web, they are known as Visual Information Retrieval Interfaces (VIRIs).

The visualization literature can be characterized as centered around two primary concerns, the background design and development of the visualization system and evaluation of system capabilities. There are many articles addressing evaluation of the systems, but typically these articles focus on heuristic evaluation rather than how the visualization impacts the user and their information behavior. Those evaluative articles which do not inform or address user behavior are not discussed in-depth here. Of interest to the current discussion is the portion which addresses the development of visualization techniques and how visualizations impact the user and their information behavior. It will become apparent that this proposed research focuses on information behavior, perceptions of performance, usefulness and learning with a specific type of visualization using concept maps as part of the search process. The visualization display's impact on these areas of the novice is part of an unexplored area in the literature.

Visualization Research

White and McCain (1997) introduce their ARIST chapter on "Visualization of Literatures" with the following statement: "the proper study of information science is the interface between people and literatures". This interface and the current trends, by their terms are, "to combine computerized graphics-visualizations-with computerized documents retrieval" (White & McCain, 1997). Thinking about information retrieval from the perspective of literatures they contain (bodies of work), McCain and White agree with the previous ARIST review by Williams, Sochats, and Morse (1995) that the textual space is too large to be seen in its entirety and drawing on what we know about modern users, those who do not know what they are looking for need the visual simplifications. Visualizing literatures and employing visualization techniques are a somewhat young and emerging area of research. There is still a lot of research needed in order to truly evaluate and understand the usefulness and impact visualization techniques have on the user. As discussed in the review of the information behavior literature, the call was to better support the user in the information retrieval environment; the use of visualization tools in the information seeking and retrieval environment makes logical sense.

Systems have emerged which bring the information retrieval technology and information visualization together. The mid-nineties saw systems like VIBE, Hearst's TileBars and Cat-a-Cone, Cone trees, Pathtrieve, and others appearing (Card, Robertson, & Mackinlay, 1991; Card, et al., 1991; Card, et al., 2001; Hearst, 1995; Hearst & Karadi, 1997; Lin, 1992; Card, Robertson, & Mackinlay, 1999; Veerasamy & Belkin, 1996). Hearst's TileBars displays a graphical representation in addition to the search results. The representation Hearst uses corresponds to the documents' contents with respect to the query terms used by the searcher (Hearst 1995).



The goal (of the visualization) is to simultaneously and compactly indicate:

1. *(i) the relative length of the document,*
2. *(ii) the frequency of the term sets in the document, and*
3. *(iii) the distribution of the term sets with respect to the document and to each other.*
   (Berkely Digital Library Project < http://elib.cs.berkeley.edu/tilebars/about.html#what>

Figure 4. Visualization: TileBars

A missing component of many of the papers emerging during this time is a validation of the maps and a suitable way to measure their impact on the user, and the user's understanding or lack of understanding of the visualization. Testing on the visualizations is rarely undertaken except as part of a usability study (Fowler, Fowler, & Wilson, 1991; Morse & Lewis, 2000), and these evaluations have been slow and isolated

(Chen & Czerwinski, 2000). A myriad of the papers written about visualizations in retrieval settings are discussions of methods for development and system implementation, some with usability studies (Lin, Soergal, & Marchionini, 1991; Lin, White, & Buzydlowski, 2003; Wise, et al.).

Chen and Yu conducted a meta-analysis of empirical information visualization studies in the literature and found that the studies were very diverse and difficult to apply meta-analyses methods. They also found that the combined effect side of visualization is not statistically significant. Their paper sought to raise the awareness that empirical studies of information visualizations are crucial to this area of developing research and the challenge is the design of practical and realistic tasks that can put different visualization features to the test (Chen & Yu, 2000). It would stand that the explanations of the systems and the discussion of the algorithms behind the system should emerge in the literature first and the evaluations and empirical testing literature follow. In 1996, Chris Ahlberg, on a panel discussion asking where visualization technology is going states: "we need a definite task focus in information visualization … which leads to the utilization of methods which focus on evaluation of performance in terms of user tasks such as time to complete a task, number of errors, learning time" (Hascoet-Zizi, et al., 1996). Some of the emerging frameworks in the literature address how to evaluate or develop the visualizations themselves and develop taxonomic guides (Chen, 1998; Graham, Kennedy, & Benyon, 2000; Morse, Lewis, & Olsen, 2000). Some designers have built user evaluations into their work (Lin, 1996; Veerasamy & Belkin, 1996; Lin, 1996; Sutcliffe, Ennis, & Hu, 2000). And others have undertaken the task of larger empirical studies (Buzydlowski, 2003; Pirolli, Card, & Van Der Wege, 2000; Westerman & Cribbin, 2000).

Pirolli, Card, and Van Der Wege (2000) used focus plus context information visualizations to understand how distortion affects information foraging behavior. Their research considered the "attentional spotlight" and how it would be affected by the density of information in a visualization using information scent as an indicator. They found that their hyperbolic browser yielded better overall results than a browser with windows which operated similarly to Internet Explorer. They also found that Hyperbolic users learned more of the tree structure and examined more nodes, faster than the window browser. Additional research on the Hyperbolic display by the same researchers found that strong information scent improves the visual search and that in compressed regions of the hyperbolic tree, crowded targets degrade the visual search (Pirolli, Card, & VanDerWege, 2003).

Other research into visualizations explored the relationships between spatial abilities and visualizations. Allen (1998) looked at the interaction of two dimensional data representations and the spatial ability of general students at a university. The design features of the system he used employed two different types of cues; spatial and perceptual. He found that users with lower levels of spatial abilities performed better than those with higher levels. One of the important questions arising from research like this is how can you support users with all different kinds of abilities with one system?

Buzydlowski (2003) explored the preferences of humanities experts and established a ranking of two different concept map visualizations (Kohonen Self-Organizing display and a Pathfiner Network display) using authors and author co-citation techniques. In that research he explored how well the maps reflected the mental models of experts as well as what type of map was preferred for presentation of data. Buzydlowski specifically addressed in his research the suggestion that a comparison of the different techniques to displaying concept maps was needed (Buzydlowski, 2003; White & McCain, 1998; White & McCain 2000; Lin 1993; Lin 1997). It is important to study how well the algorithms used to generate the displays reflect the understandings and mental models of experts. Experts are more adept at retrieving in an information retrieval situation because of their knowledge and command of domain-specific vocabulary.

In contrast, finding a means to support a novice searcher in the same search environment might be considered even more important. Experts have an understanding of a topical space and the background, experience and knowledge to use appropriate terms when searching. Novice searchers, however, do not have the same ability and we know from the information behavior literature that the very essence of needing information means one doesn't know (Belkin, 1982a; Dervin, 1977; Dervin & Nilan, 1986).

After almost ten years of developing research in the visualization information retrieval area White and McCain's questions in their 1997 ARIST review about visualizing literatures are valid and should be highlighted:

- Is the display an improvement over a simple list?
- Does it provide new capabilities?
- Is it rapidly intelligible?
- Is it helpful in real time?
- Is it tied to an important collection?
- Is it scalable upward to collections greater in size?
- Is it readily available at a reasonable cost?

A similar comment was made in 2003 by Marti Hearst in a CHI tutorial on Visualizations asking if visualizations help, the answer; is that the jury is still out. Koshman (2006), conducted a review of some of the modern visualization research as it pertains to library science. She concluded as well that while previous research has contributed to the overall development of the field, more work is needed to test users in the web environment.

It has been highlighted in this section of the visualization literature review that many articles describe systems with visualization components, and discuss the need for further development and understanding of how these can be better designed or evaluated. Shouldn't visualizations all be structured so that users solve their short term problems, but also take away more than a solution, something which helps them apply the experience of their past work toward future work (Matthew Chalmers, panelist as quoted in Hascoet-Zizi, et al. 1996)? Researchers suggest cues be taken from the information seeking literature and cognitive sciences by designing systems that emulate or support what is known about the cognitive structure and perceptual processing capabilities of the user. Understanding how visualizations, employed as part of an information retrieval tool, impact information behavior, perceptions of performance, knowledge and usefulness of the novice searcher in a real-time setting is a step toward answering that question and key to this proposed research.

Concept Maps and Learning

There are many different types of visualizations, previews and overviews, displays of hierarchical data, etc. For this research the focus is on visualizations which show semantic relationships among concepts in a given domain. Concept maps, even in paper form, are still an emerging educational tool used for organizing and representing knowledge. In the 1960's Novak began to study concept mapping techniques as tools for learning based on the theories of David Ansubel (Novak, 1993). Ansubel stressed the importance of prior knowledge for learning new concepts. His thought was that meaningful learning involves the assimilation of new concepts into existing cognitive structures (Novak & Canas, 2006). This echoes some of Kuhlthau's work borrowing from the constructivist view of learning, that the "experience of individuals involved in the process of constructing new meaning from the information they encounter" (Kuhlthau, 1993, p. 29). Novak and Canas present an adapted continuum of learning. They consider concept maps to be a higher level of meaningful learning in the educational process in contrast to something like receptive instructions which is rote learning. The primary idea is that meaningful learning requires "well organized relevant knowledge structures, and emotional commitment to integrate new knowledge into existing knowledge," whereas rote learning, there is "little to no relevant knowledge and no emotional commitment to relate new knowledge with existing relevant knowledge (Novak & Canas, 2006, p. 4).

There are many different applications for concept maps. Novak's work with concept maps deals with learning and construction of the actual concept maps as part of the process of construction while other research has focused on the manual generation of expert concept maps in diverse fields such as Pulmonary Physiology (McGaghie, McCrimmon, Mitchell, & Thompson, 2004). Others have studied the impact of concept maps on text comprehension and summarization (Chang, Sung, & Chen, 2002). Some preliminary research focusing mainly on disorientation found that learning was best, in the hypertext environment, with the conceptual map versus a spatial map (McDonald & Stevenson, 1999). Concept maps are important in meaningful learning because they serve as a template "to help organize knowledge and to structure it" (Novak, 1993). In a 12-year longitudinal study, he found there are three key factors:

1. Meaningful learning involves the assimilation of new concepts and propositions into existing cognitive structures.
2. Knowledge is organized hierarchically in cognitive structure, and most new learning involves subsumption of concepts and proposition into existing hierarchies; and
3. Knowledge acquired by rote learning will not be assimilated. (Novak, 1993).

Concept maps have been used in textbooks to provide chapter overviews to show relationships among complex concepts and used as a construction tool in the classroom to promote learning. If the knowledge structure of an expert can be represented using a concept map (Cooke, 1994), it is possible to use a visualization technique to display a system-extracted knowledge structure from a set of documents. Concept mapping and concept modules in the online environment are a "powerful approach for developing technology-based learning materials for higher and continuing education" (Gupta, Ramadoss, & Zhang, 2003). Our previous discussion on visualization demonstrated that by using the brains perceptual system visualizations can present and reveal complex and abstract concepts in an intuitive visual manner to both the novice and expert. This was likened to reading a book where the picture conveys part of the story and the text fills in gaps. Most of the research with concept mapping focuses upon the actual construction of the maps as integral to the process of meaningful learning and as a learning strategy (Novak, 1993; Novak, 2006; McCagg, & Dansereau, 1991). However, used as a representational piece of structure in a book chapter, or the knowledge of an expert, this structure might also then be presented as part of a visual information retrieval interface to support novice searchers.

Domain Knowledge and Retrieval

The difference between novices and experts, beyond the very general concept, is an underlying premise we have so far not mentioned. Think about the primary distinctions between the domain novice and the domain expert. In a very straightforward manner it would be a difference in their knowledge, more specifically as it pertains to searching; their domain knowledge. The effects of domain knowledge on information retrieval have been approached from various angles. Recall researchers in the information seeking behavior literature noted that factors impacting user behavior are interrelated and complex (Fidel & Soergel, 1983; Marchionini, 1989). One of the factors known to impact search behavior is domain knowledge. Many studies have focused on the differences between experts and novices and their behavior and performance. Domain knowledge has also been measured by varying characteristics of the participants from number of classes taken, to education level to performance on standardized tests on a specific domain given to participants. This research does not seek to measure the difference in the domain knowledge of novices and experts, therefore domain knowledge will be addressed insofar as to highlight the impact domain knowledge has on search behavior as seen in the literature. We use the domain literature to highlight the general differences between the knowledge and skills of an expert versus the novice searcher used in this experiment.

We know from research that domain experts notice meaningful patterns of information; have acquired a great deal of content knowledge and that knowledge is organized and reflects a deep understanding of their subject matter; and domain experts have specific skills that allow them to assess, process, and understand a problem differently than that of a novice (Bransford, Brown, & Cocking, 1999). Novices and experts differ in how they behave when interacting with an information retrieval system. One of the primary differences is the use of different vocabularies.

Subject experts tend to use their own terminology. Bates (1977) found that in subject searching in the manual environment that overall matching success was surprisingly low and subject familiarity had a detrimental effect (though this result was not statistically significant). The primary goal of her research was to test the effects of familiarity with subject area of searching and the principles of catalog organization. She attributed the low matching success to the precise vocabulary that experts used in expressing their subject needs, which differed from the specificity of the vocabulary found in the catalog. This research was not in the electronic environment and focused on subject heading searches. Sihvonen and Vakkari found that with an easy task, novices incorporated more thesaurus terms whereas experts used more of their own terminology when reformulating queries with the use of a thesaurus (Sihvonen & Vakkari, 2004). Shute and Smith (1993) developed a model for applying certain knowledge-based search tactics to a specific type of knowledge structure, based on the performance of an expert human intermediary and her interactions with 17 information seekers. They found that the primary searcher makes extensive use of her domain knowledge to generate suggestions for refining a topic (Shute & Smith, 1993). Hirsh, while working with children, found that high domain searchers (high domain was defined by performance in science class) were more successful than searchers with low domain knowledge (Hirsh, 1996). This corresponds to studies cited above which investigated adults. Domain knowledge helps participants in formulating and finding vocabulary to express what they are looking for. Domain experts tend to use more precise terminology which may not reflect or match the terminology used within the system. It is possible then, that displaying system terminology to both the expert and the novice can in fact, positively support the information behavior of both groups.

Other research on domain knowledge and information behavior found that search experience is an important component which influences behavior. Hseieh-Lee investigated the effects of search experience and subject knowledge on search tactics in the online environment. Her results showed that search experience did affect the searcher's use of search tactics and also suggested that subject knowledge was a factor only after a certain amount of search experience was gained by the searchers (Hsieh-Yee, 1993). Another study found that domain knowledge has an impact on searching assuming that users have a sufficient command of the system being used (Vakkari, Pennanen, & Serola, 2003). Their research explored how search tactics and search terms changed during the preparation of a research proposal by students. The increase in a user's domain knowledge during the research process generated a growth and change in the vocabulary which was reflected in the increased use of search terms. It was also found that in subject searching in the manual environment, catalog familiarity had a significant beneficial effect on search matching (Bates, 1977).

Important to mention is the development of tactics by Bates (1979a; 1979b). *Search tactics* are moves made to further a search (Bates, 1979a), while *idea tactics* are new ideas or solutions (Bates, 1979b). Many of the researcher's looking at domain knowledge as a variable also developed coding schemes for analyzing moves between query formulation and reformulation (Hsieh-Yee, 1993; Shute & Smith, 1993;

Wildemuth, 2004), these were based on Bates work with tactics (Bates, 1979a; 1979b). In this section we reviewed that domain knowledge impacts vocabulary use, in the next section we will review how search experience in conjunction with domain knowledge also has an impact on search behavior.

Novices start with an anomalous state of knowledge and move toward a more coherent state while searching (Belkin, Oddy, & Brooks, 1982a, 1982b). Wildemuth, using search tactics concentrated on the impact of each student's domain knowledge on search behavior and search outcomes. She found that when domain knowledge was low, there were more search moves per search (2003). As a beginner or newcomer to a city, I might try a different way each day to work until I found the shortest or quickest route to reach my destination. Novices, in varying situations will have a trial and error approach. This can also apply to searching.

Various research has found that domain knowledge impacts information behavior. A novice searcher is at the very early stages of learning about a domain. As domain knowledge and grasp of the information space increases, so does the command of the terms and language used to express "aboutness" in that domain; so also increases the understanding of the landscape, the routes to take and the best means of maneuvering that landscape. The driving force in reviewing the domain knowledge literature reflects back to use of document clues, to opening the landscape of information to the novice. By using the documents in the system and their "resolved state of knowledge" (Belkin, Oddy, & Brooks, 1982a, 1982b) we can extract bibliographic and textual data. Recalling the nut gathering metaphor we can pull out the leaves, the shapes of the trees and see the squirrels. In the form of a concept map, we can present that information to the novice as a visualization component to a retrieval system and allow that user to interact and connect with the landscape, to let them see the trees and gather their own hickory nuts.

Learning

From user information behavior and information retrieval, to visualizations connecting the user to the information store, and concept maps as tools for learning. We have come around to the discussion of how concept maps, historically used as an educational tool, might impact the novice searcher in the information seeking environment. The last area we will touch upon is a brief discussion of meaningful learning and learning transfer.

The domain knowledge discussion helped highlight differences between experts and novices. One avenue this naturally leads to is a discussion of learning. The domain novice can become an expert through experience and learning over time. The learning literature from the psychology and cognitive psychology perspective is vast and almost limitless. The study of the human mind, education and cognitive capabilities has been a topic of research for many years. There are many reviews which comprehensively review this literature in its vastness (see: Barnett & Ceci, 2002; Bransford, Brown, & Cocking, 1999) To focus our discussion for the purposes of this research proposal, we concentrate on learning and learning transfer with a focus within the information retrieval framework.

Our discussion of meaningful learning was mentioned as part of Novak's (1998) work with concept maps. Meaningful learning and rote learning are not simply split, but rather fall on a continuum where meaningful learning can lead to creativity and rote learning, while also in long-term memory, involves "little or no integration of new knowledge with existing knowledge (Novak & Canas, 2006 p. 5). In its most general sense, transfer of learning is "our use of past learning when learning something new and the application of that learning to both similar and new situations" (Haskell, 2001). It would seem as if it were a simple enough definition. However over the past 100 years there has been a plethora of research showing how people fail to transfer learning (Barnett & Ceci, 2002). This can be seen as quite a paradox because "transfer of learning underlies the ability to think, reason, plan and make good decisions" (Haskell, 2001). One means of supporting a user's learning refers to Dewey's theories which can be summarized by the following, "learning by doing." In the context of information retrieval, our modern retrieval systems have taken a great deal of the searching process and hidden it. This might be viewed as a manner of disengaging the searcher from the information store and the seeking process. The computer system "does it" for us as we search. We have seemingly smartened our systems and engaged the user less and less in the process. Bringing in Novak's work on concept maps, by engaging the searcher less in the information seeking process it is possible that there is less "emotional commitment to relate new with existing relevant knowledge" (Novak & Canas, 2006, p.4). This would also lessen the chance for meaningful learning during the search process and not support the movement from an "incoherent state of knowledge to a more "resolved state of knowledge" (Belkin & Oddy, & Brooks, 1982a, 1982b).

The information behavior research that addresses learning will reinforce the need for engagement in the search process. The work of Kuhlthau borrowed from the constructivist view of learning, leaning heavily on the work of Kelly as well as Dewey and Bruner. Dewey's views on learning are typically summarized by the axiom "learning by doing." In summarizing the constructivist views of learning, Kuhlthau states "it is an active, engaging process in which all aspects of experience are called into play … the Personal Construct Theory describes the experience of individuals involved in the process of constructing new meaning from the information they encounter" (Kuhlthau, 1993, p. 29). While Kuhlthau focused primarily on feelings and how those feelings impacted the searching process and construction of meaning, Marchionini addressed learning insofar as to view information seeking as a type of learning as "the goal in both cases is to change knowledge" (Marchionini, 1995, p.8). However, information seeking as learning is different from conventional learning in the degree of necessary retention; one is for a temporary task (information seeking) while traditional learning demands retention (Marchionini, 1995). It is possible that Marchionini's dichotomous statement about learning is too simple and Novak's continuum is more appropriate. Perhaps the information seeking process falls between rote and meaningful learning with the possibility of greater meaningful learning.

Conclusion

The view of the information behavior process as an active, engaging process where new meaning is constructed can be tied into this research. They key concepts to highlight are "active, engaging process". Instead of the ATM model (I feed you a term, you give me results), a system which involves the user for more than basic information exchange might in turn involve and aid the user in more meaningful learning and in turn support their extraction of terms to use, help them place their need within a context and ultimately retrieve better. To further set the scene, a quotation that Michelene Chi (1978, p. 161) also used in a Cognitive Psychology Chapter on Mental Models, will be used:

> "Tell me and I forget
> Teach me and I remember
> Involve me and I learn."
> ~Benjamin Franklin

This quotation addresses many of the basic premises of this research. Let us recall Jesse sitting in front of the computer in the university library searching for information for her paper. To further illustrate the three different lines of Franklin's quotation, we will have each of our discipline-specific observers play a part.

One of the observers, you will recall, is a librarian. One of the jobs as a librarian is bibliographic instruction sessions for students at orientation and in Freshman Comp. As part of new student orientation Jesse was given a tour of the library, told about the different resources and given papers on Boolean searching and the use of the different databases. Now three weeks into the semester with an assignment to write a paper, Jesse no longer has those materials, recalls perhaps how to get to the library databases and she remembers where the copiers and printers are in the library. She was told about different aspects of the library that the librarian thought were important, but she was not really involved in any manner of learning about these resources, it was more a passive conveyance of information which didn't seem at the time to pertain to her. She seems to have forgotten what she was told.

Next we have a teacher, Jesse's health professor. He is in the computer lab working with student and observing Jesse as she searches for information on the computer. During class he had a lecture on current trends of medical crises in the United

States as well as general topics pertinent to student health. She has seen presentations on Diabetes, Depression, Heart Disease and Obesity. She has been taught about the importance of maintaining her own personal health in order to prevent these diseases and she can remember some of the symptoms of diabetes and depression so that she is able to seek medical attention if she or a member of her family presents with those symptoms. She felt a little more commitment to this information because her Aunt has diabetes and it runs in her family. She knows some of the terms to search and so types diabetes into the box with the blinking cursor and retrieves over 100,000 returns. She sighs and starts sifting through the results finding other terms that might be more appropriate to what she is looking for. In this instance she has learned and been able to remember some of the information from her class. Her emotional connection to it was seemingly greater because of her personal health and family health history. On Novak's continuum this might be perceived as more than rote learning with a movement toward more meaningful learning.

Lastly, we have a system developer who works in the computer science department with visualizations. In his work with the library staff he has designed a front-end component to a database that would take Jesse's general diabetes query and give her the terms that the system uses, employing modern computing capabilities to extract the pertinent terms and concepts from the documents themselves. Using different background algorithms he can help her see how the literature is organized so she can see different relationships and perhaps recognize what she is looking for when she sees it. The system can be made to be more interactive allowing Jesse to "play around" which would possibly then involve her more in the process of exploring for information. We are not claiming that this interaction would be meaningful learning at the highest level, but perhaps it is a movement away from rote learning and the use of this visualization tool can support her information seeking more strongly than an ATM retrieval system.

To conclude this portion of the discussion, the librarian can tell Jesse how to search for articles, her professor can teach her about diabetes, but it is possible the system designer, through the use of visualizations can involve her in the topic she is searching. How can we help her develop skills during the process of her information seeking that might help her in the future? If the information seeking process is seen as one of construction of meaning (Kuhlthau, 1993) with the potential outcome of new knowledge (Marchionini & Shneiderman, 1988) then understanding how different tools in the information retrieval process impact the information behavior, the perception of performance, knowledge and usefulness is an appropriate question to study.

SUMMARY AND RATIONALE FOR RESEARCH QUESTIONS

To bring our discussion full round and focus upon the different areas our literature review highlighted, let us bring our attention once again to Jesse, sitting at the computer in the library looking for information.

As a novice searcher, Jesse has a vague understanding of her information need. She bypasses the librarian and has no clue that there are thesauri and other resources that might help her search. By the very nature of needing information, she comes to the retrieval system with a gap in her understanding which leads to queries and strategies that tend to be simple and basic yielding thousands if not hundreds of thousands of results. But it isn’t only her state of mind, or her stated need which should be explored, but rather the whole context of the process Jesse is involved in should be examined. Marchionini and Kuhlthau's models of information seeking exhibit this holistic approach and provide a framework in which to look at information seeking as a construction of meaning (Kuhlthau, 1993) in which various elements impact the search process including environment, setting, task domain, search system (Fidel, 1984; Fidel & Soergel, 1983; Marchionini, 1989; Marchionini & Shneiderman, 1988; Saracevic & Kantor, 1988a; 1988b; Hirsh, 1996). This picture of the seeking process focuses and reveals a clearer image of the user, their need and their state of knowledge.

The system designer observing Jesse in the library might conclude that connecting her with the information landscape is one manner to support her information behavior. "By allowing bibliographic relationships to be seen at a glance… anyone interested in the literature is given additional aids to judgment and decision" (McCain & White, 1997). The brain's perceptual system for processing information is vast in its capabilities. Visualizations can present complex and abstract concepts in an intuitive and more readily understood visual manner than large bodies of text. They can also convey large amounts of information with less cognitive load. In the information retrieval setting, concept maps, historically a learning tool, used in information retrieval might support Jesse's searching behavior by engaging her in the landscape of information and helping her to see and understand the relationships between concepts related to her area of searching.

Understanding how concept maps in the retrieval environment impacts the user is important. We highlighted the differences between a domain expert and a novice searching the same system. Domain experts have specific skills that allow them to assess, process, and understand a problem differently than that of a novice (Bransford, Brown, & Cocking, 1999). An expert notices, organizes, processes and interprets information in their environment differently than a novice that makes them more successful in their searching. It is possible a concept map, used as a tool during the information retrieval process, will help Jesse formulate better queries, see relationships and connections she might not otherwise noticed. It is also possible that a visualization display will engage Jesse in the exploration for information instead of just the searching for a document that is “good enough” (Simon, 1996).

With access to information increasing exponentially, traversing the ins and outs of the digital library and electronic access to information is more circuitous and potentially frustrating for the novice searcher. Despite all of the technological and changing computer capabilities, our standard retrieval systems mostly hide the landscape from the searcher. The system perspective of design has focused on increasing the level of what the system can do automatically for the searcher. This model engages the searcher less and less. It smartens the system, not the user. A system should connect with the user more to help them understand and learn. A system should engage a novice searcher to help them explore the information domain. Use what we know from research to engage the user. Smarten the system & smarten the user.

It is possible to better acquaint the searcher with the information they are searching. Based on calls from the information behavior literature and recent developments in visualization techniques used as components of large databases, by showing the organization of concepts that are related to a generally stated information need, it is possible to support the novice searcher in their information behavior to help them formulate better queries, know more about the topic they are searching, and feel more confident in the search process. This proposed research seeks to understand the following research questions.

RQ1: Information Behavior
Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher's information behavior?

RQ1.1	Is there a change in the **number of terms** used pre-test to post-test by display type?
RQ1.2	Is there a difference between display types in the number of **full, formal MeSH terms** incorporated post-test?
RQ1.3:	Is there a difference in the number of **full and partial MeSH terms** used pre-test to post-test by display type?
RQ 1.4:	Is there a relationship between **display type and the interface level** participants chose terms from on the post-test?
RQ 1.5:	Is there a change in the **level of specificity of search statements** pre-test to post-test by display type?

RQ2: Self-Reported Perceptions of Performance
Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher's perceptions of performance?

RQ2.1:	Is there a difference in participants' **self-perceptions of satisfaction** in the terms they chose pre- and post-test by *display type*?
RQ2.2:	Is there a difference in participants' **self-perceptions of confidence** in the terms they chose pre- and post-test by *display type*?
RQ2.3:	Is there a difference in participants' **self-perceptions of success** in the terms they chose pre- and post-test by *display type*?
RQ2.4:	Is there a difference in participants' **self-perceptions of relevance** based on the terms used pre- and post-test by *display type*?

RQ3: Self-Reported Perceptions of Knowledge
Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher's self-perceptions of personal knowledge?

RQ3.1:	Is there a difference in participants' self-perceptions of **query topic knowledge** on the pre--test by *query topic*?
RQ3.2:	Is there a difference in participants' self-perceptions **of query topic knowledge** post-test by *display type*?
RQ3.3:	Is there a difference in participants' self-perceptions **of corrected knowledge** post-test by *display type?*

RQ4: Overall Perceptions of Display Usefulness
Do concept maps, used as an exploratory tool in the information retrieval environment impact participants' overall perceptions of usefulness?

RQ4.1: Is there a difference in participants' perceptions of system support for **search formulation help** by *display types?*

RQ4.2: Is there a difference in participants' perceptions of overall display **sense and understanding** by *display types?*

RQ4.3: Is there a difference in participants' perceptions of **learning and knowledge** by *display types*?

RQ4.4: Is there a difference in participants' perceptions of **visual appeal** by *display types*?

RQ4.5: Is there a difference in participants' perceptions of **current and future system use** by *display types*?

CHAPTER 3: DESCRIPTION OF THE SYSTEM

INTRODUCTION

VisualConceptExplorer designed by Xia Lin, Ph.D. is a front end visualization component designed for use with PubMed. PubMed is the National Library of Medicine's search engine for retrieving documents from the Medline database. *VisualConceptExplorer(VCE)* generates concept maps which are viewed and explored before document specifics are retrieved through the PubMed search engine. Since the VCE system uses PubMed, and because there are certain components to both systems which are important to recognize and understand in the overall context of how *VisualConceptExplorer* generates concept maps, PubMed will also briefly be discussed however our focus in this section is on VCE. The sections below will explain the different systems in more detail as well as the integral components pertinent to this research. This is done to contrast a standard retrieval interface against visual exploration of concepts that VCE offers users which is the primary focal point of this research.

PUBMED: Standard Retrieval Interface

The PubMed retrieval engine and interface uses the Unified Medical Language System and Medical Subject Headings, MeSH, to facilitate retrieval of documents from the MEDLINE database. MEDLINE is the name of the National Library of Medicine's database of indexed journal citations and abstracts. Most of the publications covered in MEDLINE are scholarly journals with a minute number of newspapers, magazines and newsletters that are considered useful. MEDLINE now covers nearly 4,500 journals published in the United States and more than 70 other countries. All of the citations in MEDLINE are assigned MeSH® Terms and Publication Types from NLM's controlled vocabulary. It encompasses those areas of the life sciences, like behavioral sciences, chemical sciences, and bioengineering. Retrieved January 15, 2004, from http://www.nlm.nih.gov/pubs/factsheets/pubmed.html.

Searching PubMed

When users search PubMed, they are given a box for inputting query terms. Boolean search operators can be used in PubMed, but if no Boolean operators are input, PubMed assumes the AND operator between concepts. Once terms are input and sent to the system, a display screen of results is returned if a match occurs. If a match between the query and the system occurs then document surrogates like author, title and brief abstract are presented in a list format. (Please see the screen shot of a PubMed result screen searching on the MeSH term *diabetes mellitus*.)

Figure 5. PubMed Interface Screenshot

2: Sakaguchi K, Horio H, Kuwabara K, Terao Y. Links

Large chest wall reconstruction using a pedicled osteomuscle composite flap: report of a case.
Surg Today. 2006;36(2):180-3.
PMID: 16440168 [PubMed - in process]

Figure 6. Detailed PubMed Display Entry

UMLS

The UMLS or Unified Medical Language System is a project of the National Library of Medicine. The UMLS is a collection of national and international vocabularies and classifications which provide a mapping structure between different vocabularies. It is a means to identify the information sources most relevant to a user inquiry and “to negotiate the telecommunications and search procedures necessary to retrieve information from these sources”. The UMLS was created to “facilitate the development of computer systems that behave as if they ‘understand’ the meaning of the language of biomedicine and health" (UMLS National Library of Medicine. Retrieved January 13, 2005, from the World Wide Web: http://www.nlm.nih.gov/research/umls/about_umls.html).

MeSH Subject Headings

Another important component to both systems is the medical subject headings. MeSH, Medical Subject Headings, is a controlled vocabulary system (also produced by the National Library of Medicine) used for indexing, cataloging. Terms from that vocabulary can be used for searching biomedical and health-related information and documents (MeSH National Library of Medicine. Retrieved January 13, 2005, from the World Wide Web: http://www.nlm.nih.gov/mesh/intro_preface.html)

To summarize the important aspects of the above section, PubMed is the search engine and interface used to search the MEDLINE database. PubMed uses a controlled vocabulary for medical literature called MeSH, Medical Subject Headings, used for indexing, cataloging, and searching biomedical and health-related information. The UMLS (Unified Medical Language System) is composed of national and international vocabularies and classifications which provide a mapping structure between the different vocabularies. All of these components are a part of the PubMed search engine and integral to the back end components necessary for *VisualConceptExplorer* and the functions it provides.

***VISUAL CONCEPT EXPLORER (VCE)*: Visualization Tool**

In contrast to the standard search engine of PubMed, the visualization system which generates the concept maps in this experiment is called *VisualConceptExplorer*(VCE). VCE is a web-based, interactive, visualization tool that directly links to the PubMed search engine. It was designed for the exploration of medical concept relationships. VCE takes a seed term input by a user and creates a visual concept map of related medical concepts as found in the MEDLINE database. It employs the use of two different types of maps for displaying concept relationships using term co-occurrence counts. VCE is a real-time application used in conjunction with PubMed. For any medical concept, the system can generate two different styles of maps, a Pathfinder Network (PFNET), or Kohonen Self Organizing Map (SOM). The system can also produce an alphabetical list of terms. Typically the process of creating these concept maps from a bibliographic database was done manually and took a considerable amount of time. *VisualConceptExplorer*, the system used in this research, automates that process.

Searching *VisualConceptExplorer*

In contrast to the PubMed information retrieval engine, the visualization system created to generate the concept maps in this experiment is called *VisualConceptExplorer* (VCE). When searching VCE, the system presents the user a box for inputting query terms. If VCE recognizes the word entered as a MeSH term, it will display a map of related concepts. However, if the word is not recognized as a MeSH term, it will display a menu which contains MeSH terms that might be related.

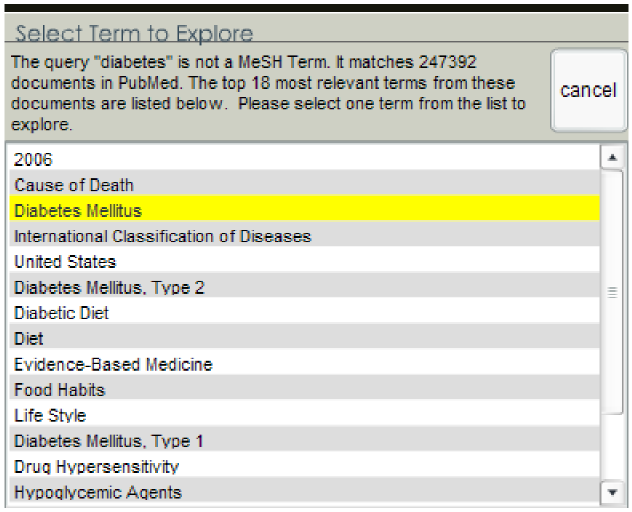


Figure 7. VCE System-Suggested Alternate Terms

Once the search term has been recognized or is selected from the list, VCE displays a concept map in the form of a PathFinder Network or a Kohonen Self Organizing map to show the conceptual relationships to the user.

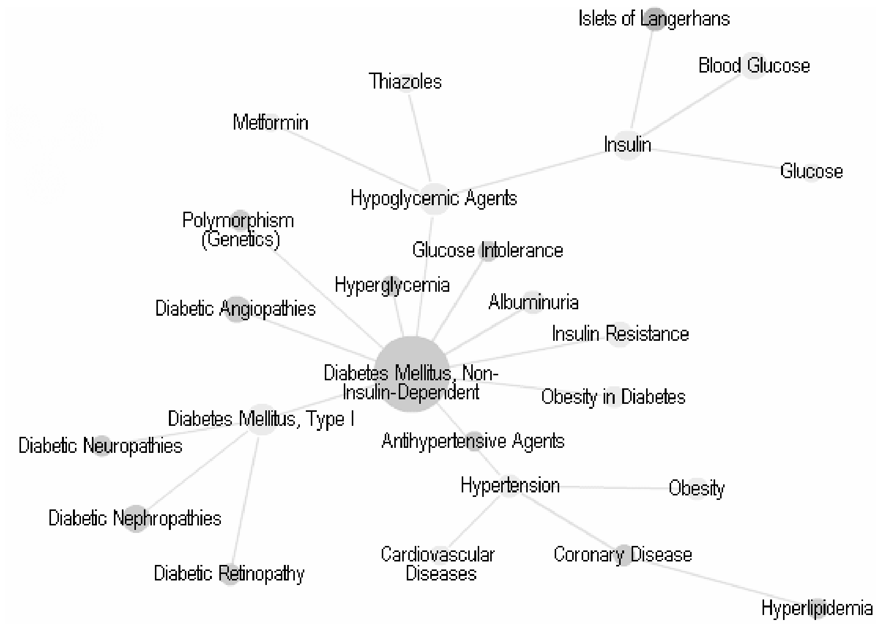


Figure 8. Pathfinder Network Display of the Term Diabetes Mellitus

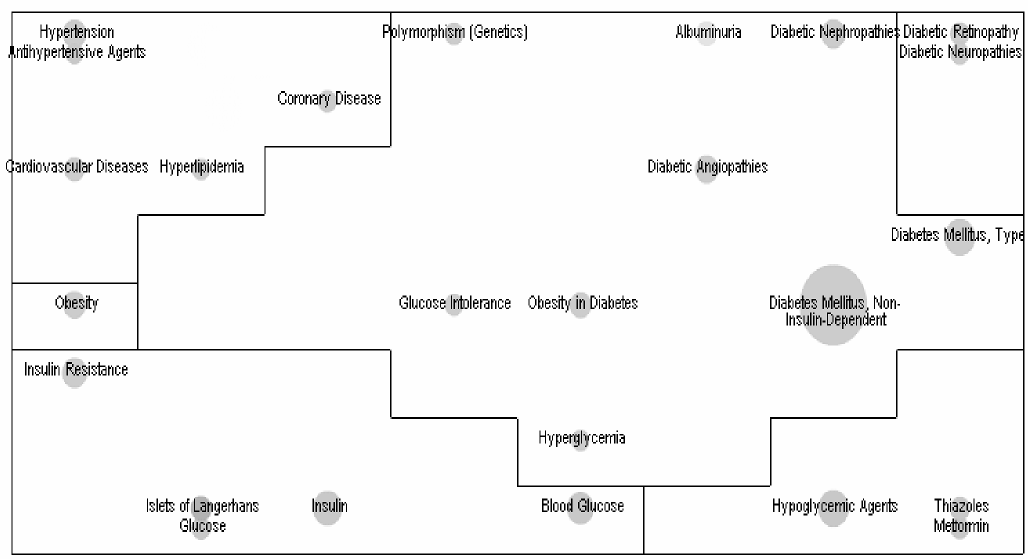


Figure 9. Kohonen Self-Organizing Map of the Term Diabetes Mellitus

Users can click on the nodes (circles) to bring up a map of the terms related to a particular word. For example, from the diabetes mellitus map above, a user might doubleclick on blood glucose. That would cause the system to then display the top 25 terms related to blood glucose in the form of a concept map, thus allowing for exploration of the topical organization of terms related in the Medline database.

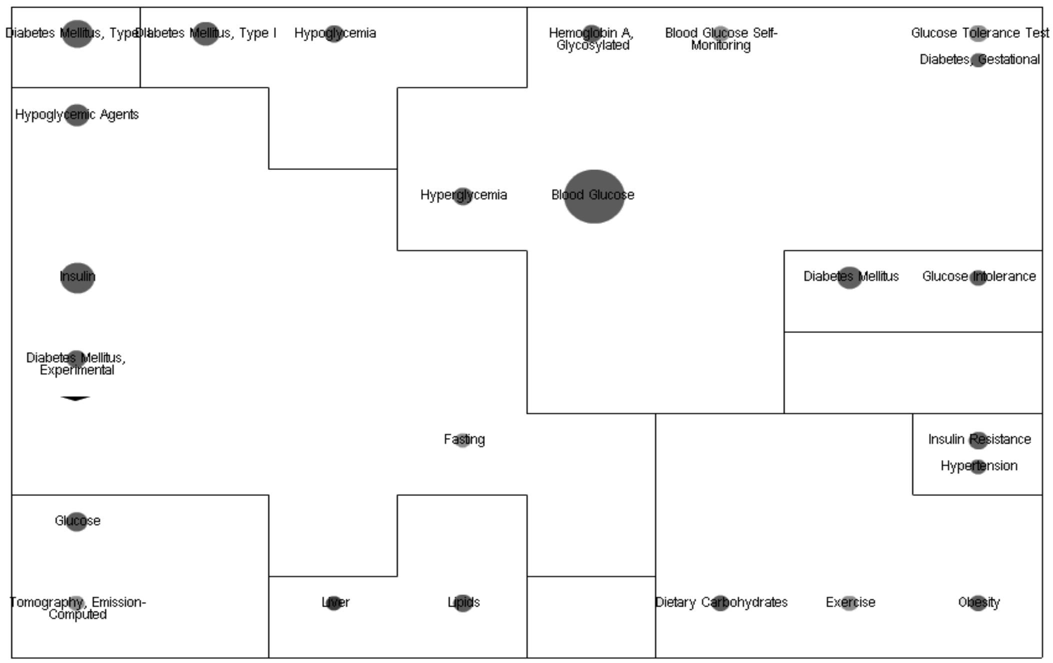


Figure 10. Kohonen SOM of "blood glucose"

Once a user has explored the map they can then use the VCE system to send their query terms to the PubMed interface to retrieve documents.

System Specifics: *VisualConceptExplorer*

It is important to discuss the more specific system details of *VisualConceptExplorer* to explain how it generates the concept maps because of its uniqueness as a visual information retrieval interface. It is also important to discuss the development of the techniques used in generating these maps. *VisualConceptExplorer* displays complex relationships among different medical concepts. It generates instant concept maps using the Unified Medical Language System (UMLS) co-occurrence database. VCE uses term co-occurrence counts in documents to proximate the semantic relationships of the terms (Lin, Xia ConceptLink. Retrieved January 13, 2005, from the World Wide Web: http://project.cis.drexel.edu/conceptlink/) Term co-occurrence counts are derived from the MeSH descriptors discussed above. If two MeSH descriptors often occur in same documents, these two descriptors are then likely semantically related. When all of the pair-wise co-occurrence patterns are taken into consideration, a concept map in the form of a PathFinder Network (PFNET or a Kohonen Self Organizing Map(SOM) can be created to show these conceptual relationships to the user.

There are three major components to *VisualConceptExplorer*: a front end, a backend, and a set of visualization procedures. The front end consists of an interactive interface implemented with FLASH. The backend includes a series of Java servlet applications that process requests from the front end and then redirect those requests to PUBMED or UMLS servers. The results from PUBMED searches are also processed and prepared for use with the visualization procedures by the backend. The visualization procedure used applies several visualization algorithms including a Path Finder Network (PFNET) and Kohonen Self-organizing mapping algorithm. The implementation optimizes the algorithms so the maps can be generated within seconds. Because of the FLASH interface and the system design, this tool can create maps of topic areas in less than 5 seconds. This allows for an experiment with real-time data and real-time users whereas in the past, using bibliographic data to visualize literatures was much more time-intensive and involved. In order to further clarify the difference between the two display formats and to discuss the history of using bibliographic data to generate the visualizations used in *VisualConceptExplorer,* the displays will be illustrated in the following section from the user's perspective.

VisualConceptExplorer: Development

As noted in Information Retrieval by Chris van Rijsbergen (1999), Luhn pioneered an approach using the frequency counts of words in the text of the document to determine those words which were "sufficiently significant to represent or characterize the document in the computer" (van Rijsbergen, 1999, p. 8). Carrying on the work of Luhn, in the 70's Sparck Jones used measures of association between keywords based on the "frequency with which any two keywords occur together in the same document," called co-occurrence (van Rijsbergen, 1999, p. 8). Exploring and displaying these associations among terms in documents is a more recent exploration in information retrieval called visualization.

Term Co-Occurrence

Within any document store there are relationships among those documents that are inherent in the text. There are author associations, research stream associations, and subject associations. Until recently the only means of really knowing those relationships was something an expert in a given field could discern or "perceive." For example, seeing the author name *White* on a paper as well as the name *McCain* would lead someone to the conclusion that both researchers are related in some manner. Someone who is an expert in the field would know that *McCain* frequently co-occurs with *White* as one was the pupil the other the teacher and their research streams have been very similar. But, it would take some following, retrieval, and reading of their papers to see the extent of this relationship. While that association is historical in nature, it is something that the novice searcher would not know or be able to discern without a good amount of field expertise. Just as there are author associations, there are also subject associations which would help a novice searcher understand the interrelations between different concepts in the information space she is searching. Term co-occurrence is one manner to explore document relationships that are implicit. Using what is called "term seeding" these relationships can be made more explicit with visualization techniques and brought into the forefront to be used by the novice searcher with little or no background knowledge. These tools can also simultaneously support the expert searcher.

Given a single word or phrase as a "seed," other terms can be retrieved which most frequently occur in designated fields across a large document collection. In order to determine all the various relationships from all the terms that are returned from the seed term, "the analysis of co-occurrence will give a metric to determine the strength of the association (Buzydlowski, White, & Lin, 2001). This metric measures strength by the number of times two terms occur together in a document collection. Another example from a one of the concept maps used in this experiment, the term *diabetes mellitus* will likely co-occur with many times *blood glucose* but with *podiatry* fewer times and with the concept *knee replacement* almost never. So the more times those terms co-occur the stronger the relationship, the fewer times, the weaker. This point illustrates what is known as co-citation analysis (ACA) which has been developed over a 20 year period

(White, 1990; McCain, 1990; White & McCain, 1998). This same methodology and argument for author co-citation analysis' validity can be applied to repeated terms to form a more general term co-occurrence analysis (Buzydlowski, White, & Lin 2001).

Again, given a single word or phrase as a "seed" other terms can be retrieved which most frequently occur in designated fields across a large collection like the UMLS co-occurrence data from Medline. In *VisualConceptExplorer*, documents which contain the seed terms can then be systematically examined to return the related terms rank ordered by the frequency with which they occur (Buzydlowski, White & Lin 2001).

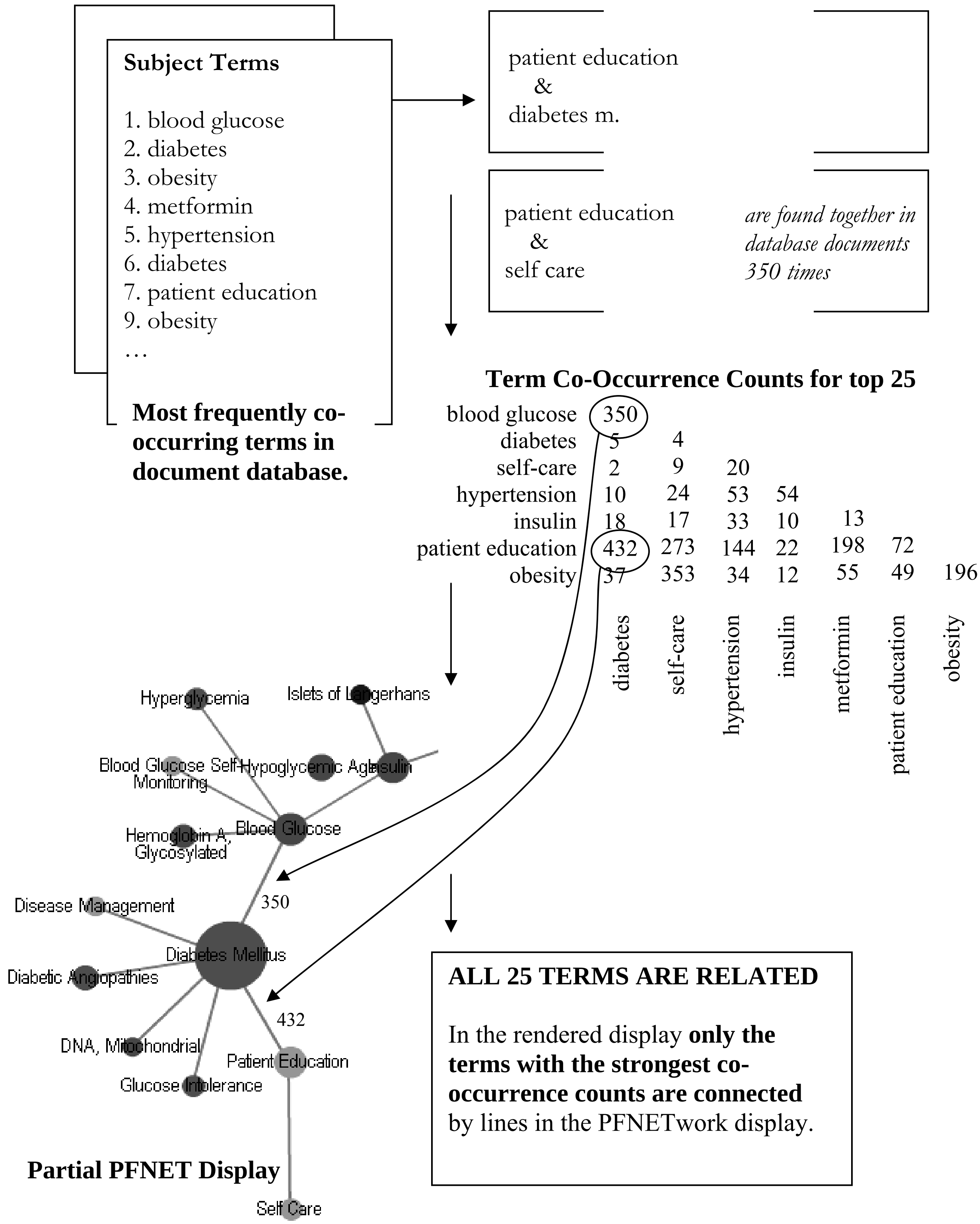


Figure 11. Term Co-Occurrence Matrix and Rendered Diabetes Display

Display techniques

When the co-occurrence numbers are compiled they form a matrix. The row and column headings represent the terms that are of interest while the intersections of the rows and columns are where we find the counts for the number of times the terms co-occurred. (See Figure 11) Once the pair-wise co-occurrences have been derived, the raw co-occurrence frequencies can be examined, but it makes it difficult because of the large number of them. Understanding what those numbers mean is challenging to grasp in its raw format. This is where visualization techniques come into play, and this is where the human ability to process information spatially becomes key.

The historical technique used to present all of the co-occurrences simultaneously is multidimensional scaling (MDS). Pathfinder Networks(PFNET), and Kohonen Self-Organizing Maps(SOM), which will be part of this experiment, have also been applied as methods for visualizing author co-citation data and term co-occurrence data.

Multi-Dimensional Scaling (MDS)

As mentioned above because of limited cognitive capabilities for the raw data generated by author co-citation analysis, this it is necessary to employ a method for visualizing results. Multidimensional scaling is a technique that reduces the dimensionality of the co-occurrence matrix to something that is more visually represented in either two or three dimensions. This methodology has been well-established using author co-citation data (White, 1990; McCain, 1990; White & McCain, 1998). It creates points in a two or three dimensional space which represent authors or terms from a given dataset. This means of data visualization helps to reveal structures in the data that cannot be captured in any other way (Cleveland, 1993). The MDS technique seeks to compute the co-ordinates in the map for author names so as to preserve the order and distance determined by the co-citation counts (Buzydlowski, 2003).

Self-Organizing Maps (SOM)

Another technique for displaying complex relationships applied to co-citation data was developed by Lin (1997). This technique uses self-training neural networks to determine the placement of authors or concepts on a map. The Kohonen's Feature Map algorithm can be used to survey contents of a given document space, to detect semantic relationships, and then generate a map display that shows both contents and semantic relationships (Lin, 1997). This method also produces a map similar to MDS where authors or terms are represented as points on a page. It permits similar authors to be automatically grouped into word or concept ideas. The map displays the data in what are called regions. The size of an area indicates relative importance of that area and the neighboring regions show associations. With the Kohonen map, concept relationships are represented by regions and their neighborhood relationships. In this map, the closer the two terms are, the more often they are co-occurred in the MEDLINE database. *(Please see Figure 12)*

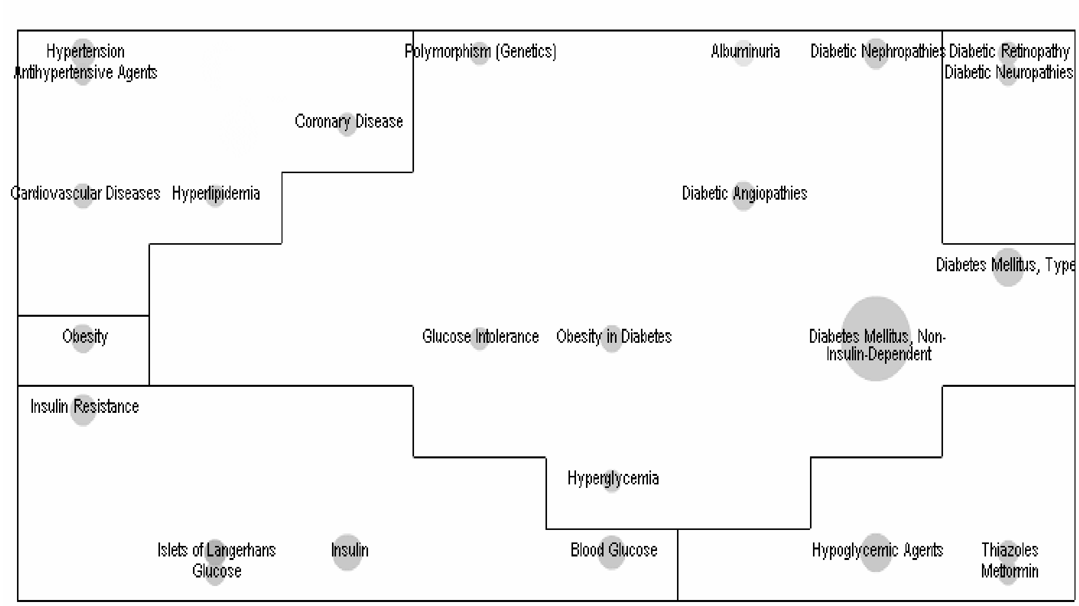


Figure 12. Kohonen SOM of Diabetes Mellitus

Pathfinder Networks (PFNET)

The third mapping technique is called a Pathfinder Network. This technique was developed by Schvaneveldt (1990) growing out of the realization that while network representations are prevalent in theoretical work in cognitive psychology and artificial intelligence, there were few methods for obtaining a network representation from empirical data. Pathfinder Networks are determined by identifying those proximities which have the most efficient connections between the entities. It considers the indirect connections that are provided by paths through other entities. (Schvaneveldt, 1990, p. ix) These maps do not show regions in their display like the SOMs discussed above, but rather as points on a page with lines connecting those points most directly associated with one another.

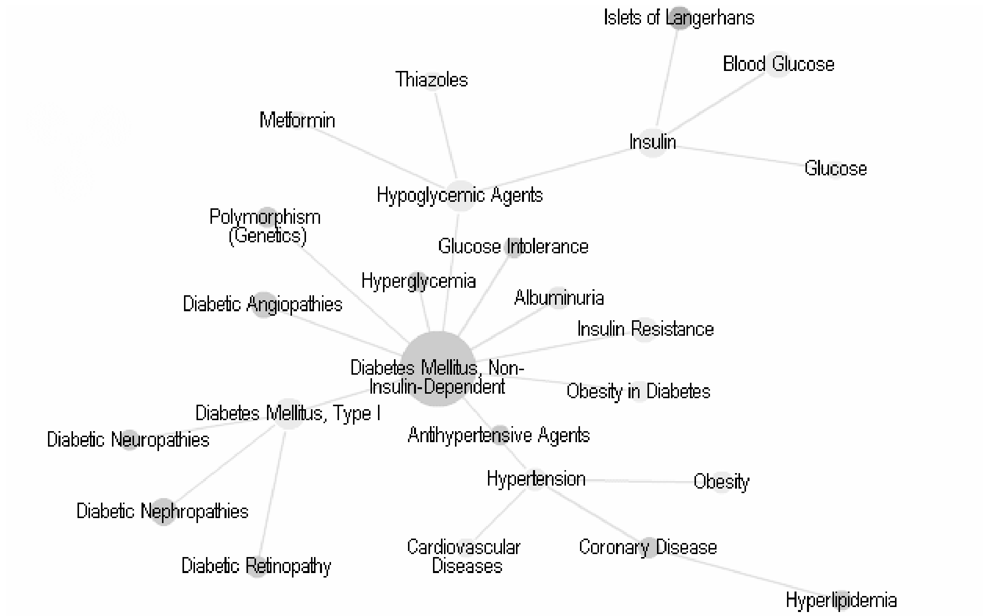


Figure 13. Pathfinder Network of Diabetes Mellitus

These three different methods for visualizing co-occurrence data differ in their display structures and format. It has been suggested that a comparison of the different techniques is needed (White & McCain, 1998, 2000; Lin 1993, 1997). If the three types of displays were placed on a continuum with SOM on one end and PFNET on the other, MDS would fall in the middle (White & McCain, 2000; Buzydlowski, 2003).

SOM---------------------------MDS---------------------------PFNET

To reduce the complexity of the study and compare the most disparate types of maps, this research only compares PFNETS and Kohonen SOMS. This is the same rationale as Buzydlowski (2003). Also, maps which can be rendered in real-time using an important system was necessary in order to compare these displays. Currently there is no real-time system which can display an MDS map. In Buzydlowski's (2003) research that task was to explore the preferences of the expert users and try to establish a ranking of the map types. This research attempts to understand the impact concept maps have on novice searchers information behavior and perceptions of performance, knowledge and use with a real-time, system and search task.

Interpretation of the Different Display Formats

As noted above the VisualConceptExplorer (VCE) system used for this experiment generates two forms of concept maps based on term co-occurrence (Kohonen SOM and Pathfinder Network). The system also generates a basic LIST of terms. Each display presents the same 25 most-related terms to the topics of depression and cholesterol. However, each of the displays organizes the concepts in a different manner.

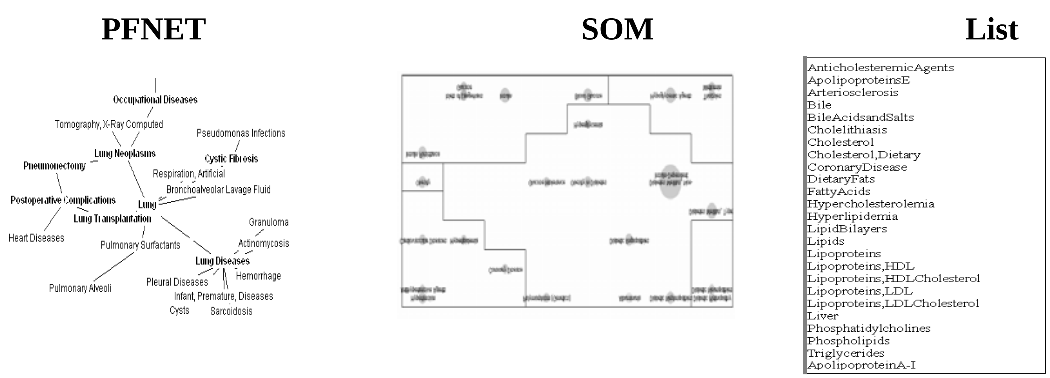


Figure 14. Display Types at a Glance (PFNET, SOM, List)

In the above illustration we get a quick understanding for the differences between the display types. For example, we note the PFNET uses line to connect concepts while the SOM uses box-like regions or “concept areas” (Lin, 1992). The LIST group does not use lines or regions to connect concepts and does not show any indication of relationship by proximity or line connection.

The primary difference between the PFNET and SOM is how each indicates term relatedness. For the PFNET, line segments directly connect related concepts while a Kohonen SOM uses contiguity (Buzydlowski, 2003). The size of an area or concept area corresponds to the frequencies of occurrence of the words and the neighboring relationships of areas are an indication of frequency of co-occurrence of the concepts represented by the areas (Lin, 1992). In a PFNET, two concepts are linked if their terms share a line, while a SOM allows for multiple points of contact (either by sharing a side or by being contained in the same regions) (Buzydlowski, 2003).

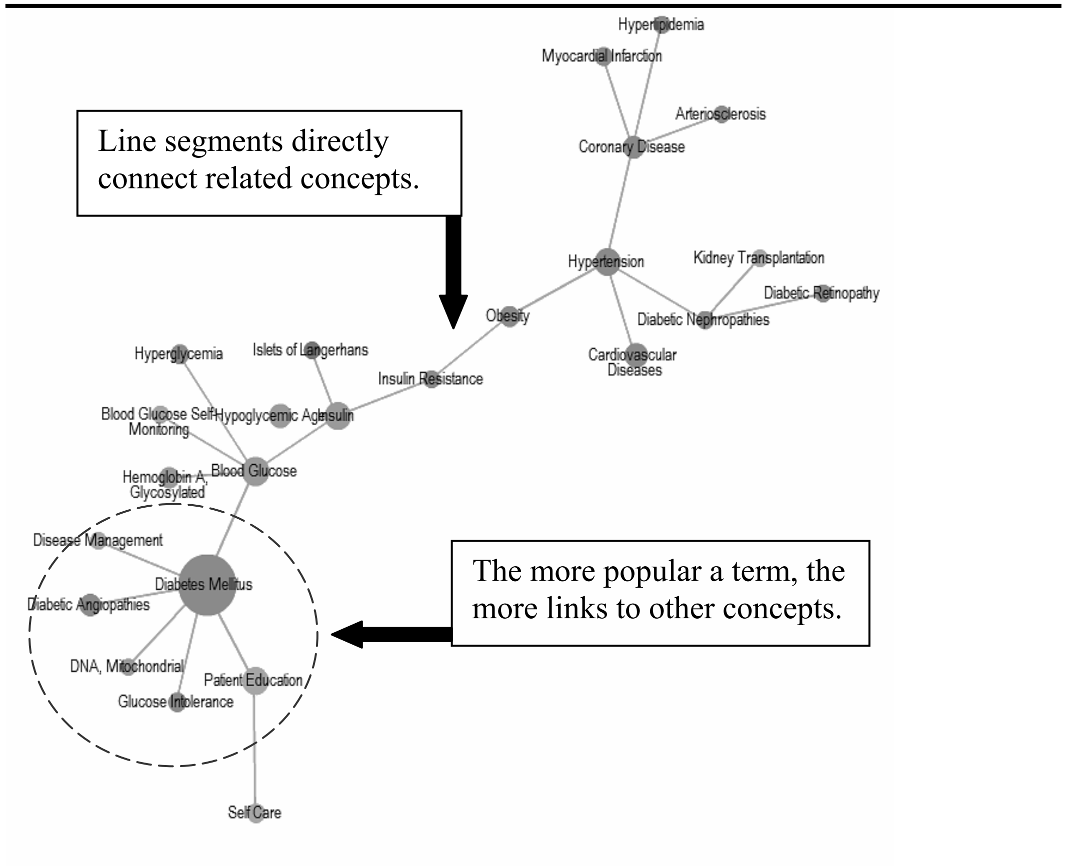


Figure 15. Interpretation- Diabetes PFNET

A SOM displays terms with relative frequency differently, if a term occurs with frequency (the more popular a term is), the more space mapped out on the display. This differs from a Pathfinder network in that the more popular a term is, the more links to other concepts explicitly displayed through directly connected lines. Another key difference in the interpretation of the relationships is the direct versus indirect relationships shown by the PFNET versus the SOM. The PFNET displays show relatedness of concepts by the direct lines which seem to lead the reader of the map while the SOM with its more open regions and implicit associations leaves more room open for interpretation. The intuitive nature and interpretation of these displays by domain novices was important to this experiment. This importance develops from some basic questions about visualizations asked almost 10 years ago, “Is the display an improvement over a simple list; is it rapidly intelligible; is it helpful in real time?” (White & McCain, 1997). Therefore interpretation of the displays was left up to the participants in the experiment.

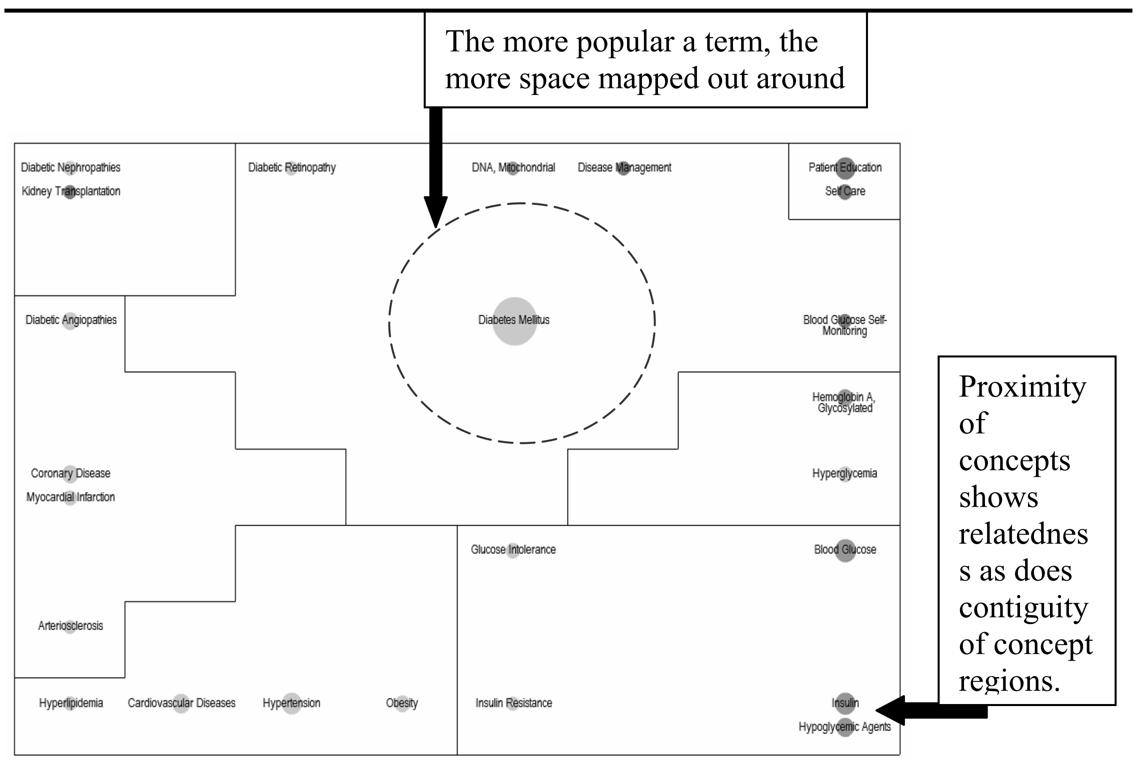


Figure 16. Interpretation Diabetes SOM

Perceptual Processing and Interpretation

This discussion of the different display types naturally leads to a discussion of perceptual processing, or the brain's ability to process visual components. The study of how we perceive visual components has had many contributing theories over time. One theory or principle that has had a lasting impact is to work of Gestalt psychologists in the 1940's. These psychologists in Germany noted that the brain has self-organizing tendencies and that the whole of objects are organized into patterns or groups instead of many different individual parts (Ware, 2000).

The Gestalt principles of visual perception address 6 areas and attempt to describe how people organize visual elements as a unified whole: 1) figure/ground (elements are separated based on contrast); 2) similarity (similar elements are seen as a group); 3) proximity/contiguity (elements that are close together are seen as a group); 4) continuity (viewers expect elements to extend along a continuous line); 5) closure (tendency to see complete figures); and 6) area (two overlapping areas, the larger is seen as background, the smaller object a figure) (Köhler, 1969).

To briefly discuss this in relationship to a specific display type, let us consider the visual layout of the PFNET used in this research. Terms in the PFNET displays typically flow along lines of connected concepts. The principle of continuation suggests that the eye is compelled to move along one object and continue to another object. To state the same principle another way, when viewing a group of visual entities, we are more likely to create visual entities out of elements that are smooth and continuous (Ware, 2000). Secondly, terms in the PFNET display are typically clustered around central nodes or visual objects. Because of the closeness of the terms in the PFNET display the theory of proximity also comes into play. Elements placed close together tend to be perceived as a group and these groups in turn suggest relatedness.

As mentioned above, PFNET concepts are directly connected to one another with lines. It has been argued that connectedness is a more powerful grouping principle than proximity, size, or shape, and it is a more fundamental organizing principle when it comes to visual perception (Palmer & Rock, 1994). When the mind attempts to "attend to a single dot, our attention spreads instead across the entire group in which it falls" (Driver & Baylis as quoted in Scholl, 2001).

Using these principles as a guide to developing effect displays is important; however, this is not a focus of this research. The Gestalt Principles of Perception were brought into the discussion of the visualization displays as one manner for possible discussion of user behavior.

This next section addresses the experimental design, participants, methods and data collection instruments used in this experiment.

CHAPTER 4: METHODS

DEFINITION OF TERMS

Display type refers to the independent variable which is a form of visualization tool. There are three different display types used in this experiment: LIST, SOM, PFNET.

Full, formal MeSH terminology refers to the complete term/term phrases used by subjects on the post-test that are the full, formal MeSH terminology as found on the displays used in the experiment. This refers only to full, formal medical subject headings, for example, "lipoproteins, LDL cholesterol" is a full, formal MeSH term whereas "LDL" is not.

Interface Level refers to the number of displays drilled into by participants by clicking on terms in the map. For example, by using the mouse, a participant on the first interface level of depression can click on the term "anxiety". This action pulls another map of the term anxiety which replaces the first one. This next display contains the 25 most highly co-occurring terms related to the "seed" term anxiety. This would be considered a second-level interface because it replaced the first map (depression) participants saw.

Kohonen Self-Organizing Map (SOM) is a visualization technique which uses self-training neural networks to determine the placement of authors or concepts on a map. The Kohonen's Feature Map algorithm can be used to survey contents of a given document space, to detect semantic relationships, and then generate a map display that shows both contents and semantic relationships (Lin, 1997). This method also produces a map similar to MDS where authors or terms are represented as points on a page. It permits similar authors to be automatically grouped into word or concept ideas.

Natural Language refers to terms and term phrases used by participants that are not found in MeSH. For example the term “cheerios” or “disadvantages,” used in this experiment by participants is not considered MeSH terminology.

Partial MeSH terminology refers to terms which might be used by participants that are MeSH terminology, but not full, formal MeSH. For example while “Antidepressive Agents, Tricyclic” would be considered a full, formal MeSH term, “antidepressive,” “tricyclic” or “agents” used alone or in combination would be counted as partial MeSH terminology.

PathFinderNetwork (PFNET) is a visualization technique developed by Schvaneveldt (1990) and growing out of the realization that 1) while network representations are prevalent in theoretical work in cognitive psychology and artificial intelligence, 2) there were few methods for obtaining a network representation from empirical data. Pathfinder Networks are determined by identifying those proximities which have the most efficient connections between the entities. It considers the indirect connections that are provided by paths through other entities. (Schvaneveldt 1990, p. ix)

Query scenario or query topic refers to the medical topics used in this experiment. Two primary medical topics were used; cholesterol and depression. These scenarios set the topical and situational information need for the experiment.

Visual Concept Explorer (VCE) is a web-based, interactive, visualization tool that directly links to the PubMed search engine. It was designed for the exploration of medical concept relationships. VCE takes a seed term input by a user and creates a visual concept map of related medical concepts as found in the MEDLINE database. It employs the use of two different types of maps for displaying concept relationships using term co-occurrence counts. It also generates a standard alphabetic list of terms.

METHOD OVERVIEW

Using a between-subjects experimental design, novice undergraduate students were observed in the information retrieval environment in order to measure the impact of computer generated concept maps on 1) the information behavior, 2) self-perceptions of performance and knowledge, as well as 3) perceptions of usefulness of the displays.

Participants in this experiment explored one of three displays using *VisualConceptExplorer* (VCE), a real-time information visualization system attached to a medical database and accessible through the World Wide Web. The VCE system generates two different types of concept maps and an alphabetical LIST containing the 25 most highly co-occurring terms based on a seed topic. Using hypothetical information needs on two medical topics (cholesterol, depression), participants explored the three different display formats. Data was collected using a pre- and post-test instrument, a general background questionnaire and a reaction questionnaire on perceptions of their assigned displays.

System: *VisualConceptExplorer*

The visualization system used in this experiment, *VisualConceptExplorer* (VCE), was developed by Dr. Xia Lin of Drexel University's i-School. As discussed in more detail in the system section of this dissertation, VCE uses term co-occurrence counts and generates concept maps in conjunction with various National Library of Medicine tools. VCE generates two forms of concept maps: Kohonen SOM and Pathfinder Network. Both of these concept maps as well as a third alphabetical LIST use MeSH subject headings to display related concepts. There are other components of VCE, for example a real-time searching function; however, this experiment limited the use of the system to just the displays in order to eliminate potential confounding variables, such as system design issues, and the learning curve of the retrieval aspect of the *VisualConceptExplorer* system. Confounding interaction among features can be a recurring problem in usability studies and scaling back the experiment until only the visualization remains is one technique to control for those issues (Morse, Lewis, & Olsen, 2000). For this research we want to understand the impact the visual presentation has on the behavior and perceptions of the participant. Therefore, the interpretation of these displays by domain novices was not explained to the participants.

Participants

The participants in this experiment were undergraduate students at the University of Maine at Augusta who volunteered to take part. For this experiment, the use of novice searchers, or educated non-specialists, refers to students enrolled at the undergraduate university level. The primary intent of this research is to evaluate how the use of concept maps impacts the novice searcher's information behavior and perceptions of performance and usefulness. This population was a convenience sample from the available university student population. None of the students in this experiment had any prior interaction with the *VisualConceptExplorer* system. Solicitation of volunteers for this experiment followed the policies of the University of Maine at Augusta as mandated by its Institutional Review Board. Subjects in this study were not paid and received no other compensation for participating. Permission to conduct experiment was granted by both the University of Maine at Augusta and Drexel University Institutional Review Boards.

Search Task Scenarios

For the display exploration sessions, two information scenarios were developed using search tasks that might require undergraduates to us a library database system. This allowed each visualization tool to be presented to participants with the results of an initial probe. The queries for this experiment were developed using the frequent topic searches in MedlinePlus, as well as the broad topic categories in both MedlinePlus and WebMD, two predominant Web-based health information sources for the general public. Compiling the topics from both those sites created a list of 119 potential topics.

Seventeen topics overlapped in subject matter, those were kept. Next, the set of seventeen was reduced by eliminating concepts that were too broad in scope and would overlap in concept coverage. For example, topic areas like “mental health” or “cancer” were eliminated because they were too expansive and would include as sub-topics disorders like Alzheimer’s disease or prostrate cancer. The remaining topics were more specific medical concepts such as diabetes and depression. This brought the topic pool from seventeen to nine medical concepts. Each concept was searched in PubMed for the number of returns containing that concept. The outliers (minimum and maximum) were then eliminated leaving seven topics. There were three topics with roughly the same number of returns. Two topics (depression and cholesterol) were chosen at random to be used for query scenarios in this experiment.

General queries were developed from the chosen medical topics for hypothetical scenarios. The query scenario phrasing was purposefully kept broad and open so that the users would explore the medical topic and not focus on one specific area as a result of the query itself.

For your health class you have been asked to give a 30 minute presentation on diabetes. You need to find out as much about the topic as possible and locate articles that contain the information needed.

Figure 17. Sample Query

In order to obtain an estimate of the length of time each exploration might take participants, two librarians were asked to test the three query scenarios and evaluate them for clarity, system presentation issues or any other unforeseen problems. The final queries were then tested in a small pilot study on the intended subject audience to make sure there were no large variances in amount of time or other ambiguities found. Pilot study information as well as changes in experiment due to pilot study are discussed in Appendix J in more detail. (*Also see Appendix F & G: Task Scenarios*)

Experimental Design: Between Subjects

A mixed between-subjects design was used in order to: 1) isolate the learning effect, 2) minimize cognitive overload, and 3) accurately measure the impact of the individual visualization tools. This design consisted of two within subject variables (search topic and pre-post), each with two levels, and one between subjects variable (display type), with three levels (LIST, SOM, PFNET). Participants were randomly assigned to the different levels of the independent variable.

Independent Variable

The independent variable manipulated for this experiment was display type. All participants were given the same two query scenarios on cholesterol and depression and all other aspects of the experiment remained constant. There are three different levels to the independent variable of display type. (*IV = independent variable and subscript T1,T2,T3, = tool*.)

Table 1. List of Independent Variables and Participant Groups by Display Type

Independent Variable: Display type		
Display	Name	Type
IVT_1	Tool: LIST of Terms	Query Scenario & terms (list)
IVT_2	Tool: Kohonen SOM	Query Scenario & terms (regions)
IVT_3	Tool :Pathfinder Network	Query Scenario & terms (links)

Participants were randomly assigned to one of the three display types (LIST, SOM, Pathfinder). Assigning each of the three display types randomly to subjects before the experiment began ensures the same number of participants in each of the three display types (LIST, SOM, Pathfinder). For example, Subject A, randomly assigned to Group T2, was given two search scenarios, each covering a different medical topic; cholesterol and depression. For each of these topics, Subject A will also have a SOM display. The query scenarios will stay consistent among all the subjects in this experiment. Each query will be employed by each participant and since the type of task might influence participants (Beaulieu, 1997; White & Marchionini, 2007), query scenario topic is treated as an independent variable.

Table 2: List of Query Scenario Variations and Order of Assignment to Participants

Participant	*Display Type*	*Query Order(alternating)*
Participant A	IV_{T1} : Pathfinder Network	Cholesterol → Depression
Participant B	IV_{T2} : Kohonen SOM	Depression → Cholesterol
Participant C	IV_{T3} : LIST of Terms	Cholesterol → Depression
Participant A_1	IV_{T1} : Pathfinder Network	Depression → Cholesterol
Participant B_2	IV_{T2} : Kohonen SOM	Cholesterol → Depression
Participant C_3	IV_{T3} : LIST of Terms	Depression → Cholesterol

Dependent Variables

The primary dependent variables measured were:

- *information behavior*,
- *self-perceptions of performance*,
- *self-perceptions of reported knowledge*, and
- overall *perceptions of usefulness*

Information behavior refers to changes in participants' actions after being exposed to the independent variable of display type (LIST, SOM, PFNET). This question focuses upon terms used for query formulation and reformulation, as well as search statements made by participants. Changes in search terms and search statements were analyzed for pre- and post-exposure to the display. Information behavior was examined by looking at the terms and term phrases participants used, as well as by looking at written search statements pre- and post-viewing of the display type. The dependent variable of information behavior was evaluated by examining the following:

- number **of terms** used pre-test to post-test
- number of **full, formal MeSH terms** incorporated post-test
- number of **full and partial MeSH terms** used pre-test to post-test
- relationship between **display type and the interface level** participants chose terms from on the post-test
- **level of specificity of search statements** pre-test to post-test

The next dependent variable, s*elf-perceptions of performance* refers to how a user feels(confidence, satisfaction) about the terms they have chosen and how a participant believes those terms will perform in a search process(success and relevance). This dependent variable was measured by asking participants to circle the most appropriate selection on a scale of 1(not at all) to 5 (very) for the below statements:

- I am satisfied with the terms I used above.
- I am confident these terms will help me locate resources for the assigned task.

- I would be successful locating information for this project using the terms above.
- I feel the terms I used above are relevant to the search topic.

The third dependent variable focused on *perceptions of knowledge*. Participants' perception of knowledge refers to self-reported ratings of how much a subject feels they know about the medical topic they are being asked to explore (cholesterol, depression). Participants were asked to rate three statements using a scale of 1 to 5 about their knowledge on the query topic used for the experiment.

- How would you rank the amount of knowledge you possess on this topic?
- I feel I know more about the topic than I did before searching.
- My post-search knowledge has corrected what I knew before searching.

The last dependent variable explored was general participant reaction to the displays they explored. These questions addressed the following:

Search Formulation

- The system helped me formulate my search.
- After using the system I decided to change what I was looking for.
- Using the system made coming up with search terms easier.

Sense and Understanding

- The system was easy to understand.
- The system made no sense at all.
- I understood the medical topics better using the system.

Knowledge

- I learned about the medical topic with the system.
- I know more about Medical Subject Headings.

Visual Appeal

- The system was visually appealing.

System Use

- I would have preferred not using the system.
- I would use the system again if I had the option.

Data Collection Instruments

Three different instruments for collecting data were developed using prior research in the information behavior literature, including retrieval, as well as the cognitive psychology literature.

Data was collected using the following:

1. general information questionnaire (ex. *age, gender, computer experience*)
2. tool reaction questionnaire *(ex. aesthetics, helpfulness, layout*)
3. pre/post-test on query **cholesterol/depression** (ex. *search statements, terms for searching, knowledge statements, and self-perceptions of satisfaction; confidence; success; relevance and knowledge*)

Information behavior and Domain Knowledge, Understanding

The instruments for this proposed research were developed based on a variety of prior research in order to elicit the best possible varied means of measuring information behavior. In the very basic sense behavior will be measured by terms and term phrases written down by participants' pre- and post display interaction (Borland & Ingwersen, 1997; Greenberg, 2001; Kuhlthau, 1993; Saracevic & Kantor, 1987b). Information was also collected on search statements and knowledge statements. Bloom's Taxonomy, which ranges from simple recall or recognition to the increasingly more complex learning; evaluation, was important to consider (Bloom, 1956). Within the area of learning another discussion addressing different types of learning transfer from the cognitive psychology literature is also important. The questionnaires developed for this experiment were partially developed using prior research on information seeking behavior from Kuhlthau (addressed in the information seeking discussion), Novak's work on learning, and Chi's work with self-explanations.

Self-explanations refer to an activity in the context of learning, wherein explanations to oneself are generated. The goal of a self-explanation is to make sense of what one is reading or learning and to move beyond simple memorization. Within a self-explanation there may be a unit of inference which can be identified as a piece that states something beyond what the sentence or text explicitly stated. The questionnaires were used as a direct method for eliciting knowledge and information need statements about the search domain both before and after exploring the displays (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, DeLeeuw, Chiu, & LaVancher, 1994; Chi, Feltovich, & Glaser, 1981; Kuhlthau 1993). Search statements were used to gauge specificity of stated search needs before and after seeing the interface and knowledge statements were a qualitative manner of gauging what participants felt they learned during the exploration of the displays.

Precision, Recall, and Relevance

Precision and recall are the primary measures historically used to capture aspects of the retrieved documents, and used to measure the effectiveness over a set of queries processed in batch mode. They are typically the primary measurements used to evaluate a system's algorithm based on retrieval performance. Therefore when attempting to explore retrieval effectiveness from the user's perspective, measures which quantify how informative the retrieval process is might be more appropriate (Baeza-Yates & Ribeiro-

Neto, 1999; Saracevic, 1995). In our modern "Google World" precision and recall are not as important when it comes to focusing on the user. Because this research and the interest of the experiment is not on performance from a system perspective, but rather on the user and their behavior, perceived performance, learning and system usefulness, precision and recall will not be used as metrics. More qualitative means of measuring relevance will be used. Perceived relevance, utility, and satisfaction, which represent overall assessments of system performance from the viewpoint of the user, will be employed (Börlünd & Ingwersen, 1997; Schamber, 1994; Schamber, Eisenberg & Nilan, 1990).

Procedures

Institutional Review Board (IRB) documents, including information about informed consent, were distributed to the participants at the beginning of the experiment and a standard experiment protocol was read (*See Appendix C: Experimental Instructions*). After IRB documents were signed and collected, participants were asked to fill out a general information questionnaire. Upon completion of the general information questionnaire, subjects were given a sample search topic, an introduction and tutorial on how to navigate system, and allowed to explore their assigned display for up to five minutes. Participants were also allowed to ask questions during this time. After initial exploration was completed, the experiment and data collection began.

Exploration Sessions

Each participant in the experiment explored the assigned display on two different medical topics (cholesterol, depression). Each of these sessions had a pre- and post-test instrument. Upon completion of both sessions a general system reaction questionnaire was distributed. For a summary of steps in the experimental session, see, Figure 18.

Steps

1. Introduction to study
2. General Info Questionnaire
3. Introduction to Displays
4. Query scenario 1(depression, cholesterol)
    a. Pre-tool Test
    b. Tool (List, Regions, Links)
    c. Post-Tool Test
5. Query scenario 2(cholesterol, depression)
    a. Pre-tool Test
    b. Tool (List, Regions, Links)
    c. Post-Tool Test
6. Visualization Reaction Questionnaire
7. Completion-Thank you

Figure 18. Outlined Steps Through the Experiment

RESEARCH QUESTIONS

The purpose of this research, using a between subjects experimental design, was to measure the impact computer generated concept maps have on the information behavior, perceptions of performance and knowledge, as well as perceptions of usefulness, of the novice, undergraduate student in the information retrieval environment.

RQ1: Information Behavior

Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher's information behavior?

RQ1.1 Is there a change in the **number of terms** used pre--test to post-test by display type?

RQ1.2 Is there a difference between display types in the number of **full, formal MeSH terms** incorporated post-test?

RQ1.3: Is there a difference in the number of **full and partial MeSH terms** used pre-test to post-test by display type?

RQ 1.4: Is there a relationship between **display type and the interface level** participants chose terms from on the post-test?

RQ 1.5: Is there a change in the **level of specificity of search statements** pre-test to post-test by display type?

RQ2: Self-Reported Perceptions of Performance

Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher's perceptions of performance?

RQ2.1: Is there a difference in participants' self-perceptions of **satisfaction** in the terms they chose pre- and post-test by *display type*?

RQ2.2: Is there a difference in participants' self-perceptions of **confidence** in the terms they chose pre- and post-test by *display type*?

RQ2.3: Is there a difference in participants' self-perceptions of **success** in the terms they chose pre- and post-test by *display type*?

RQ2.4: Is there a difference in participants' self-perceptions of **relevance** based on the terms used pre- and post-test by *display type*?

RQ3: Self-Reported Perceptions of Knowledge

Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher's self-perceptions of personal knowledge?

RQ3.1: Is there a difference in participants' self-perceptions of **query topic knowledge** on the pre-test by *query topic*?

RQ3.2: Is there a difference in participants' self-perceptions **of query topic knowledge** post-test by *display type*?

RQ3.3: Is there a difference in participants' self-perceptions **of corrected knowledge** post-test by *display type?*

RQ4: Overall Perceptions of Display Usefulness

Do concept maps, used as an exploratory tool in the information retrieval environment impact participants' overall perceptions of usefulness?

RQ4.1: Is there a difference in participants' perceptions of system support for **search formulation help** by *display types?*

RQ4.2: Is there a difference in participants' perceptions of overall display **sense and understanding** by *display types?*

RQ4.3: Is there a difference in participants' perceptions of **learning and knowledge** by *display types*?

RQ4.4: Is there a difference in participants' perceptions of **visual appeal** by *display types*?

RQ4.5: Is there a difference in participants' perceptions of current and future **system use** by *display types*?

CHAPTER 5: RESULTS

ORGANIZATION OF RESULTS

This chapter reports the data and methods of analysis arranged by research questions. It is organized in the following manner:

- Introduction
- Overview of the sample population
- RQ1: Information behavior
- RQ2: Self-reported perceptions of performance
- RQ3: Self-reported perceptions of knowledge
- RQ4: Overall perceptions of display usefulness

INTRODUCTION

This study examined the impact of computer generated concept maps on the information behavior, perceptions of performance, knowledge, and overall perceptions of usefulness of the undergraduate student in the information retrieval environment. There were two primary research questions driving this experiment with two secondary questions. The first research question (RQ1) examined the impact display type had on the novice searcher's information behavior (*what terms they chose, how many terms, full MeSH terms, partial MeSH, what level of the display interface terms were taken from* etc.). The next research question (RQ2) investigated the impact the different display types had on participants self-reported perceptions of performance (*satisfaction, confidence, success, relevance).* The third research question (RQ3) addressed self-reported perceptions of knowledge and understanding while the last research question explored the impact display type had on participants' overall reactions to the displays (*ease of use, future use, search formulation, visual appeal, etc.).*

Data was collected over the Fall 2006 semester at the University of Maine at Augusta using three data collection instruments:

- general information questionnaire (ex. *age, gender, computer experience*)
- display type reaction *(ex. aesthetics, helpfulness, layout*)
- pre/post-test on query **cholesterol/depression** (ex. *search statements, terms for searching, knowledge statements, and self-perceptions of satisfaction; confidence; success; relevance and knowledge*)

Each participant in the experiment was given a hypothetical information need on two different medical topics (cholesterol, depression). Using these medical topics, each participant explored one of three interactive visual information retrieval displays, answered a pre- and post-test instrument and then completed a final questionnaire on their perceptions of the displays.

Different types of inferential statistical tests were used to examine the research questions. When appropriate, factorial ANOVAs, mixed between-within ANOVAs, and CHI-SQUARE tests of independence were conducted.

SUMMARY OF SIGNIFICANT RESULTS

RQ1 Information Behavior		
RQ1.1	Terms used pre to post-test	Significant **general effect** pre- to post-test across all groups. *(Rep. Measures ANOVA)* ($F = 10.51$, $df = 1$, $p < .01$)
RQ1.3	Full & partial MeSH terms pre- to post-test	Significant **general effect** pre- to post-test across all groups. *(Rep. Measures ANOVA)* [$F(1, 57) = 21.515$, $p < .001$]
RQ1.4	Display type and interface level terms chosen from	For both cholesterol and depression there is a **relationship between display type and interface level term.** *(CHI-SQUARE)CH* X^2 (4, N = 398) = 19.298, $p < .01$; & DE X^2 (4, N = 361) = 14.864, $p < .01$
RQ2: Self-Perceptions of Performance		
RQ2.1	Feelings of satisfaction	Significant **general effect** for the within subject factor of pre- to post-test across all groups. Significant interaction effect of Pre-post by query topic. *(Rep. Measures ANOVA)*
RQ2.2	Feelings of confidence	Significant **general** pre- to post-test across all groups. Significant interaction effect of Pre-post by query topic. *(Rep. Measures ANOVA)*
RQ2.3	Feelings of success	Significant **general effect** pre- to post-test across all groups. (*Rep. Measures ANOVA)*
RQ2.4	Feelings of relevance	Significant **general effect** pre- to post-test across all groups. *(Rep. Measures ANOVA)*
RQ3: Self-Reported Knowledge		
RQ3.1	Query topic knowledge, pre-test	Significant difference in **participants' knowledge by query topic.** (Cholesterol/Depression) *(ANOVA)*
RQ4: Overall Perceptions of Display Usefulness		
RQ4.2	Sense and Understanding	Significant difference **between groups on system understanding, and "making sense of the syste**m". *(ANOVA)*
RQ4.4	Visual Appeal	Significant difference between groups on visual appeal. *(ANOVA)*
RQ4.5	Future System Use	Significant difference between groups on future use of the system. *(ANOVA)*

SUMMARY OF NON-SIGNIFICANT RESULTS

RQ1 Information Behavior		
RQ1.2	Full formal MeSH terms	No difference between groups. *(ANOVA)*
RQ1.5	Level of specificity of search statements	No significant difference in the level of specificity of search statements from pre- to post-test. *(Rep. Measures ANOVA)*
RQ3: Self-Reported Knowledge		
RQ3.2	Query topic knowledge post-test	No significant difference in the level of post-test knowledge between display types, nor by query topic. *(ANOVA)*
RQ3.3	Post-Search knowledge correction	No significant difference in post-test correction of knowledge between display types, nor by query topic. *(ANOVA)*
RQ4: Overall Perceptions of Display Usefulness		
RQ4.1	Search Formulation	No significant differences. *(ANOVA)*
RQ4.3	Learning & knowledge	No significant differences. *(ANOVA)*

Overview of Sample Population

The University of Maine at Augusta is a Regional Baccalaureate institution in the Capitol city of Augusta in the State of Maine. Currently UMA offers 17 baccalaureate degrees and 16 associate degrees. According to the most recent data, approximately 73% of UMA students are part-time, 74% are female; and over 97% are Maine residents.

Sixty students (approximately 1% of the student population) at the University of Maine at Augusta were sampled during the fall 2006 semester. This was a convenience sample from the available university student population who volunteered to participate in this research. (See Table 3.)

Table 3: Distribution of Participants by Gender and Age Range

		Age Range				Total
		under 21	21-34	35-49	50-64	
Gender	Female	7	8	15	5	35 58.3%
	Male	1	12	12	0	25 41.7%
Total		8 3.3%	20 3.3%	27 45%	5 8.3%	60 100%

RQ1: INFORMATION BEHAVIOR

Do concept maps, used as an exploratory tool in the information retrieval environment, impact the novice searcher's information behavior?

The first dependent variable examined was behavior. Information behavior refers to changes in participants' actions after being exposed to the independent variable of display type (LIST, SOM, PFNET. Information behavior was examined by looking at the terms and term phrases participants used, as well as by looking at written search statements pre- and post-viewing of the display type. It was evaluated by examining the following dependent variables:

- number **of terms** used pre-test to post-test
- number of **full, formal MeSH terms** incorporated post-test
- number of **full and partial MeSH terms** used pre-test to post-test
- relationship between **display type and the interface level** participants chose terms from on the post-test
- **level of specificity of search statements** pre-test to post-test

RQ1.1: Information Behavior -Terminology

Number of Search Terminology Used Pre- and Post-Test

Participants were asked both pre-test and post-test, "What terms or term phrases would you use to search for information on the above topic?" Each subject wrote out the terms they would use to search on the medical topics (cholesterol, depression).

RQ1.1: *Is there a change in the number of terms used pre-test to post-test by display type?*
HYP 1.1: *There is a change in the number of terms used pre- and post-test by display group.*

Mixed between-within ANOVA with one between subject factor (display type) and two within subject factors (pre-post and query topic) was used to test the effect of display type on the dependent variable (number of terms used by participants). The means and standard deviations are shown in Table 4, following.

Table 4. Descriptive Statistics for the Number of Terms Used Pre- and Post-Tests

	Display Type	Mean	Std. Deviation	N
CH-PRE Number of terms used.	LIST	7.00	2.340	20
	SOM	6.50	1.960	20
	PFNET	5.90	2.426	20
	Total	6.47	2.258	60
CH-POST Number of terms used.	LIST	7.65	2.007	20
	SOM	7.70	2.155	20
	PFNET	6.50	2.259	20
	Total	7.28	2.179	60
DE-PRE- Number of terms used.	LIST	6.40	2.604	20
	SOM	6.90	2.469	20
	PFNET	6.40	2.437	20
	Total	6.57	2.473	60
DE-POST Number of terms used.	LIST	7.45	1.761	20
	SOM	7.05	2.395	20
	PFNET	7.20	2.687	20
	Total	7.23	2.280	60

As shown in Table 5, we did not find difference in how arrangement of terms on the different displays impacts the number of terms used by participants. These results show that while there is a significant increase in the number of terms used from pre- to post-test after subjects are exposed to the independent variable of display type [$F(1,57) = 10.51$, df = 1, p = .002], this increase in use of terms occurs across the groups with no significant difference between them.

Table 5. ANOVA Within Subject Results for the Number of Terms Used by Participants during Pre- and Post-Tests

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Pre-Post	33.004	1	33.004	10.516	.002**
Pre-Post *by* Display Type	.358	2	.179	.057	.945
Query Topic	.037	1	.037	.013	.909
Query Topic *by* Display Type	10.675	2	5.338	1.872	.163
Pre-Post *by* Query Topic	.338	1	.338	.197	.659
Pre-Post *by* Query Topic by Display Type	6.175	2	3.088	1.801	.174

(**p < .01)

No one display type outperformed the other with the number of terms used post-test. Given feedback from a system which includes targeted search topic terminology from an outside source, participants incorporated terminology from the displays on the post-test regardless of arrangement.

RQ1.2: Information Behavior –Full, Formal MeSH Terminology

Our previous analysis explored the relationship between display type and the overall number of terms used by participants' pre- to post-test (RQ 1.1). This was one basic way of looking at impact of display type. This next research question explores the incorporation of full MeSH terminology. Full MeSH terminology refers to the complete term/term phrases used by subjects on the post-test that are the exact formal MeSH terminology also found on the displays used in the experiment. For this analysis we are only looking at the use of full, formal MeSH terms, for example, "lipoproteins, LDL cholesterol" is a full, formal MeSH term whereas "LDL" is not. Each display contained the same terminology arranged in different visual manners. Evaluating the number of full MeSH terms participants used on the post-test is another way to explore the impact the display types had on query reformulation and overall information behavior.

Terms and term phrases were first evaluated in a binary manner against the formal MeSH terminology. This means that the total number of formal MeSH term/term phrases used by each subject out of total number of term/term phrases was determined. Ratios of full MeSH terms to total terms used post-test were calculated for each participant.

For example:

Subject #1 had 5 post-test term/term phrases total:

1. **Alzheimer disease**
2. treatments
3. effects
4. **mood disorders**
5. types of depression

2 full MeSH terms
3 partial MeSH terms
5 total term/term phrases

$$\frac{\textbf{full MeSH terms}}{\textbf{total terms}} = \frac{\mathbf{2}}{\mathbf{5}} = \mathbf{.4}$$

Figure 19. Sample Coding of Full, Formal MeSH Participant Terms

RQ1.2: Search terminology: number of full, formal MeSH terms incorporated on post-test response

RQ1.2 *Is there a difference between display types in the number of full, formal MeSH terms incorporated post-test?*

HYP1.2 *There is a difference between display types in the number of full, formal MeSH terms incorporated post-test.*

A mixed between-within subjects ANOVA was conducted to explore the impact of display type on incorporation of display terminology (Tabachnick & Fiddell, 2007). Subjects were randomly placed into three groups for the independent variable (Group 1: LIST; Group 2; Kohonen SOM; Group 3: Pathfinder Network). There was no statistically significant difference for the main effect of query topic [$F(1, 57) = .388$, p = .536]. For the interaction effect of display type by query topic, there was also no significant effect [$F(2, 57) = .695$, p = .503].

Table 6. ANOVA Results for Full, Formal MeSH Terms Used by Participants during Post-Tests

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Query Topic	.010	1	.010	.388	.536
Query Topic *by* Display Type	.034	2	.017	.695	.503
Error (Query Topic)	*1.403*	*57*	*.025*		

Table 7. Descriptive Statistics for Full, Formal MeSH Terms Used by Participants during Post-Tests

	Display type	Mean	Std. Deviation	N
CH Total full MeSH incorporated out of total # used.	LIST	.7510	.24862	20
	SOM	.7140	.25574	20
	PFNET	.7940	.21917	20
	Total	.7530	.23984	60
DE Total full MeSH incorporated out of total # used.	LIST	.6975	.26807	20
	SOM	.6865	.30339	20
	PFNET	.8215	.18678	20
	Total	.7352	.26044	60

Based on the above analyses it would appear that display type has no effect on the number of full MeSH terms incorporated on the post-test response between groups. If we look at the mean for each of the groups in Table 7, we find that on average the Pathfinder Network group (M = .7940) appears to incorporate more full MeSH terms than either the Kohonen SOM(M = .7140) or LIST (M = .7510) groups for the query of cholesterol. For the query of depression (PFNET, M = .8215; SOM, M = .6865; LIST, M = .6975;), the same stands true. While this is not statistically significant it is meaningful as we will see a pattern develop throughout the results where the PFNET display type appears to come out a bit ahead of both the LIST and SOM groups.

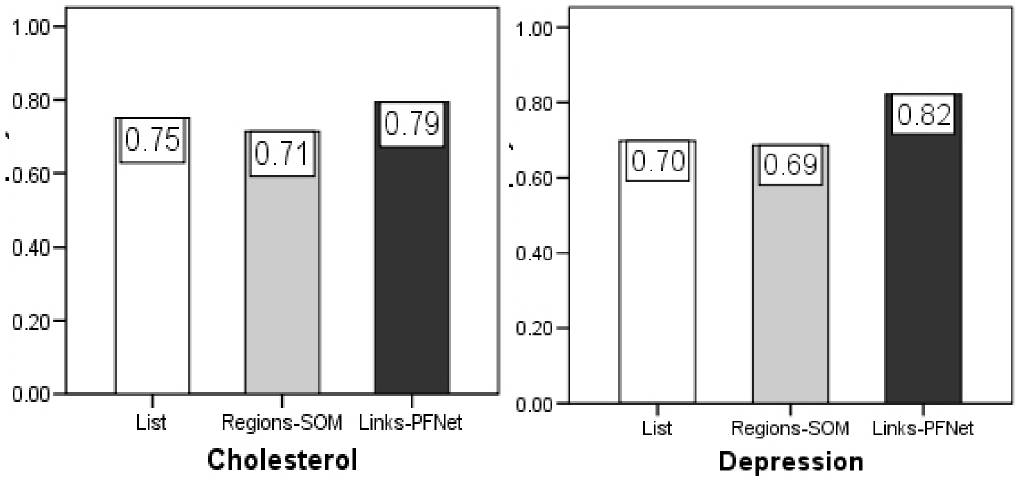


Figure 20. Comparison of Formal MeSH Term Use Broken out by Display Type

Both of the previous analyses looked at information behavior from a very basic level of counting number of term/term phrases used. In the next section, the analyses attempt gain a finer picture of what is happening pre- to post-test by looking at incorporation of both full and partial MeSH terminology.

RQ1.3: Information Behavior-Full and Partial MeSH Terms

Counting the number of terms used pre- and post-test (RQ 1.1) and counting the number of full, formal MeSH terms incorporated post-test (RQ 1.2) are both methods of evaluating display impact on the information behavior of the participants. However, that analysis doesn't measure or really tell us if there is a shift in term use from pre-test to post-test. In order to clarify the picture of what is happening with the use of participant terminology; this section will analyze the number of full and partial MeSH terminology used by participants not only post-test but also pre-test. This is different from the previous analysis which looked only at full, formal MeSH terms incorporated on the post-test.

To further define this method of examination, the full MeSH term analysis looked at only the incorporation of the full, formal MeSH terminology like “Antidepressive Agents, Tricyclic”. It is likely that participants would use some full MeSH terminology; however it would seem even more likely that participants would incorporate bits and pieces of full MeSH terminology. For example instead of writing out “Antidepressive Agents, Tricyclic,” a participant might only choose the term “Antidepressive”. It is likely that these partial Mesh terms would be used both pre-test and post-test. It is also important to note that the nature of terms in our popular language like depression, are also considered MeSH terms. Therefore using full and partial medical subjects heading as a standard against which to measure participant search terms would help us determine change in the nature of those terms after display interaction.

The terms and term phrases used by participants were counted as either partial MeSH, or full MeSH terminology (as in the analysis on full MeSH term incorporation RQ1.2 above). Term/term phrases that were either full or partial MeSH were counted together.

Subject #1 had 5 post-test term/term phrases total:

1. Alzheimer disease
2. **treatments**
3. **effects**
4. mood disorders
5. types of **depression**

2 full MeSH terms
3 partial MeSH terms
5 total term/term phrases

$$\frac{\textbf{full/partial MeSH}}{\textbf{total terms}} = \frac{\mathbf{5}}{\mathbf{5}} = \mathbf{1}$$

Figure 21: Explanation of Full, Formal MeSH and Partial MeSH Terminology

For example "Alzheimer disease" is a full MeSH phrase, and "types of depression" is only partial MeSH; both of these terms would be counted in this analysis. Term/term phrases were coded appropriately and for each participant a ratio was calculated of number of full and partial MeSH terms/term phrases out of total number of terms/term phrases used.

RQ1.3: *Is there a difference in the number of full and partial MeSH terms used **pre-test to post-test by display type**?*
HYP1.3: *There is a difference in the number of full and partial MeSH terms used **pre-test to post-test by display type**.*

Mixed between-within ANOVA with one between subject factor (display type) and two within subject factors (pre-post; query topic) was conducted to test the impact of display type on the number of full and partial MeSH terms used by participants pre- and post-test. The means and standard deviations are shown in Table 8, following.

Table 8. Descriptive Statistics for Full & Partial MeSH Terms Used by Participants during Pre- and Post-Tests

	Display Type	Mean	Std. Deviation	N
CHOLESTEROL- PRE- Full/partial MeSH terms used.	LIST	.8860	.14372	20
	SOM	.9050	.13081	20
	PFNET	.8180	.19490	20
	Total	.8697	.16067	60
CHOLESTEROL-POST- Full/partial MeSH terms used.	LIST	.9500	.13845	20
	SOM	.9890	.03401	20
	PFNET	.9940	.02683	20
	Total	.9777	.08468	60
DEPRESSION-PRE- Full/partial MeSH terms used.	LIST	.8935	.13666	20
	SOM	.8785	.14687	20
	PFNET	.8695	.13968	20
	Total	.8805	.13908	60
DEPRESSION-POST- Full/partial MeSH terms used.	LIST	.9325	.13202	20
	SOM	.9020	.22703	20
	PFNET	.9820	.05988	20
	Total	.9388	.15643	60

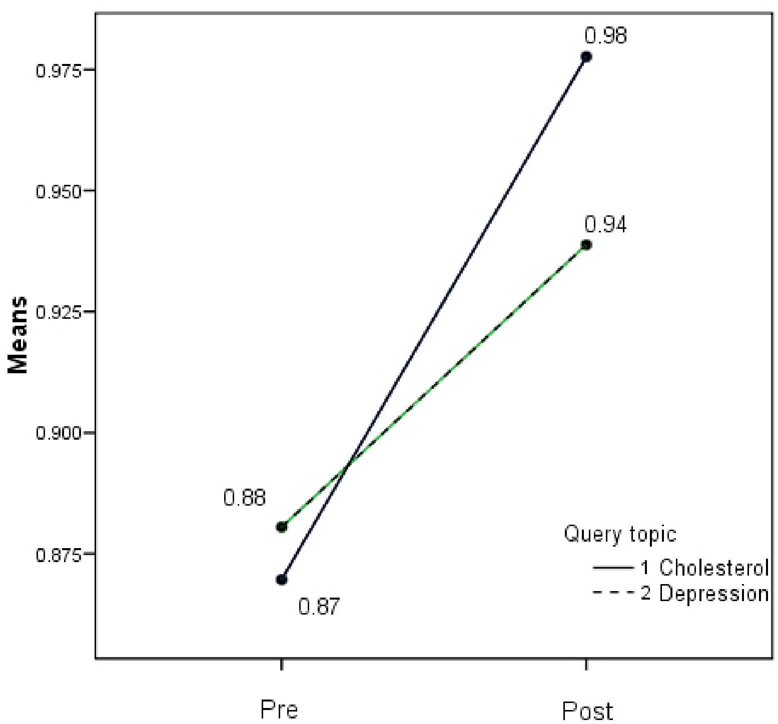


Figure 22. Use of Full & Partial MeSH Terminology Broken Out by Query Topic

There was a significant general effect of pre- to post-test across all groups on the full/partial use of MeSH terms [$F(1, 57) = 21.515$, $p < .001$]. These results show that while there is a significant difference in the use of full and partial MeSH terms pre- to post-test after subjects are exposed to the independent variable of display type, this increase in use of those terms occurs across all the groups with no significant difference between them. We were unsuccessful in establishing a difference in how term arrangement might differentially impact the use of full and partial MeSH terms into query reformulation by participants.

Table 9. ANOVA Within Subject Results for Full & Partial MeSH Terminology Used by Participants

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Pre-Post	.415	1	.415	21.515	.001***	.274
Pre-Post *by* Display type	.112	2	.056	2.903	.063	
Error (Pre-Post)	*1.099*	*57*	*.019*			
Query Topic	.012	1	.012	.994	.323	
Query Topic *by* Display type	.061	2	.030	2.577	.085	
Error (Query Topic)	*.674*	*57*	*.012*			
Pre-Post *by* Query Topic	.037	1	.037	2.979	.090	
Pre-Post *by* Query Topic *by* Display Type	.005	2	.002	.185	.832	
Error (Pre-Post by Query Topic)	*.708*	*57*	*.012*			

(***$p < .001$)

The means show that the PFNET group, on average, seems to use more full and partial MeSH terms for both the cholesterol and depression queries (See Figure 23, following.). This difference in the means is not a statistically significant difference. However, in the larger picture of all the data analysis, it merits mention as part of an emerging trend. The PFNET display appears to come out slightly ahead while the SOM and LIST group remain second.

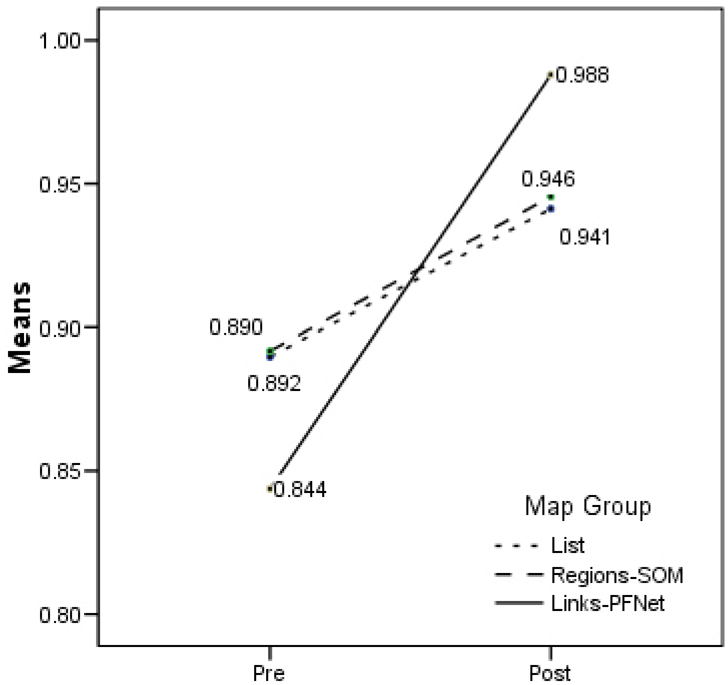


Figure 23. Means for Full & Partial MeSH Terms Used by Display Type

RQ1.4: Information Behavior –Interface Level of Post-test terminology

In the previous analyses we found that display type had no impact on number of terms used or the number of full MeSH terminology used post-test. We also found no statistically significant difference between groups in the full/partial MeSH terminology used post-test by participants. Understanding what level of the interface participants chose terminology from for query reformulation would help us further understand the impact arrangement of display terms on user behavior. It is possible that browsing behavior and interface level participants choose their terminology for the post-test might be impacted by display arrangement. Therefore getting a better picture of interface level terms were chosen from is worthy of investigation.

To preface the analysis and discussion of interface level where terms were chosen from, it is important to discuss the interactive nature of the displays used in this experiment. Each display generated by the VCE system contains 25 of the most highly co-occurring terms related to a “seed” term. For example, the primary display for depression, shown below, has 25 terms on its display. This display is not static in nature.

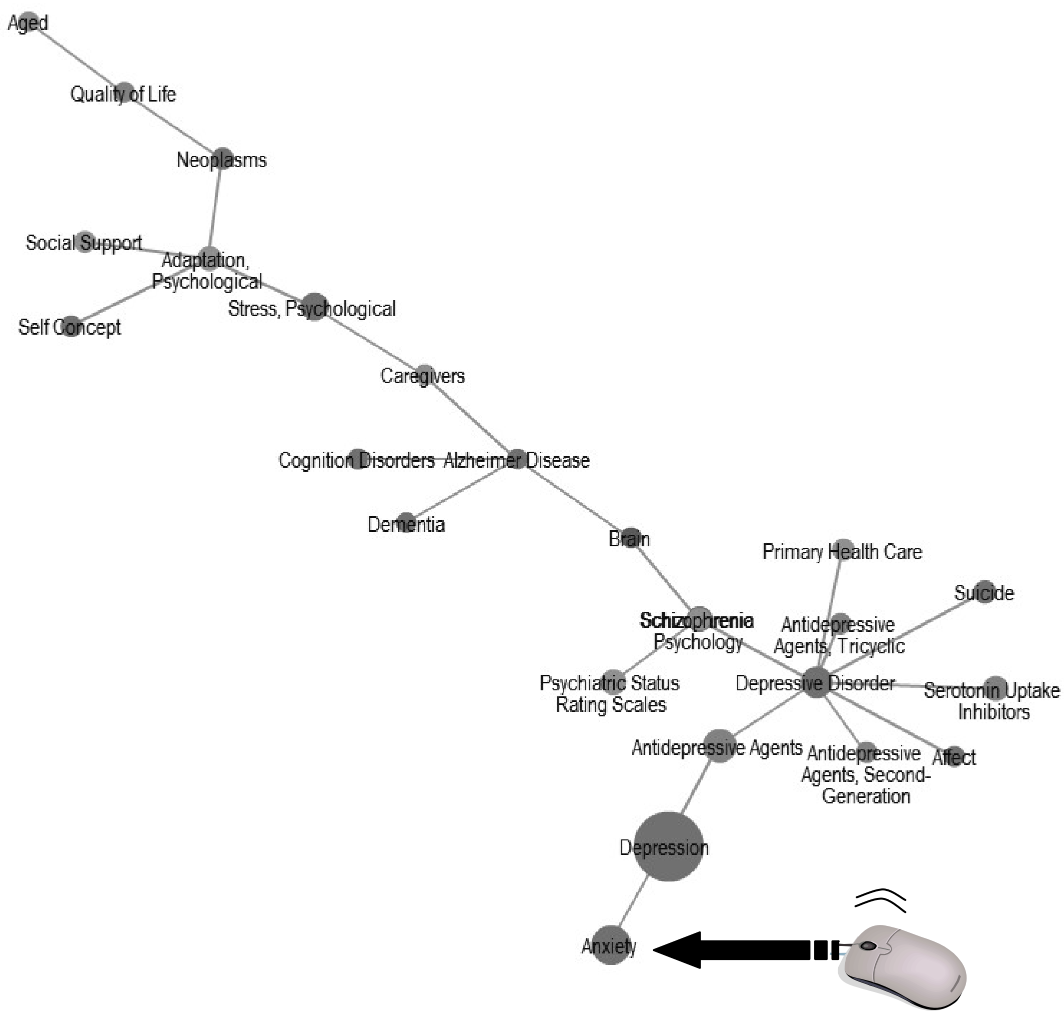


Figure 24. Depression PFNET Illustrating Interactivity of Displays

For example, by using the mouse and clicking on the term “anxiety,” another map replaces the first one. This is another display containing the 25 most highly co-occurring terms related to the “seed” term; anxiety.

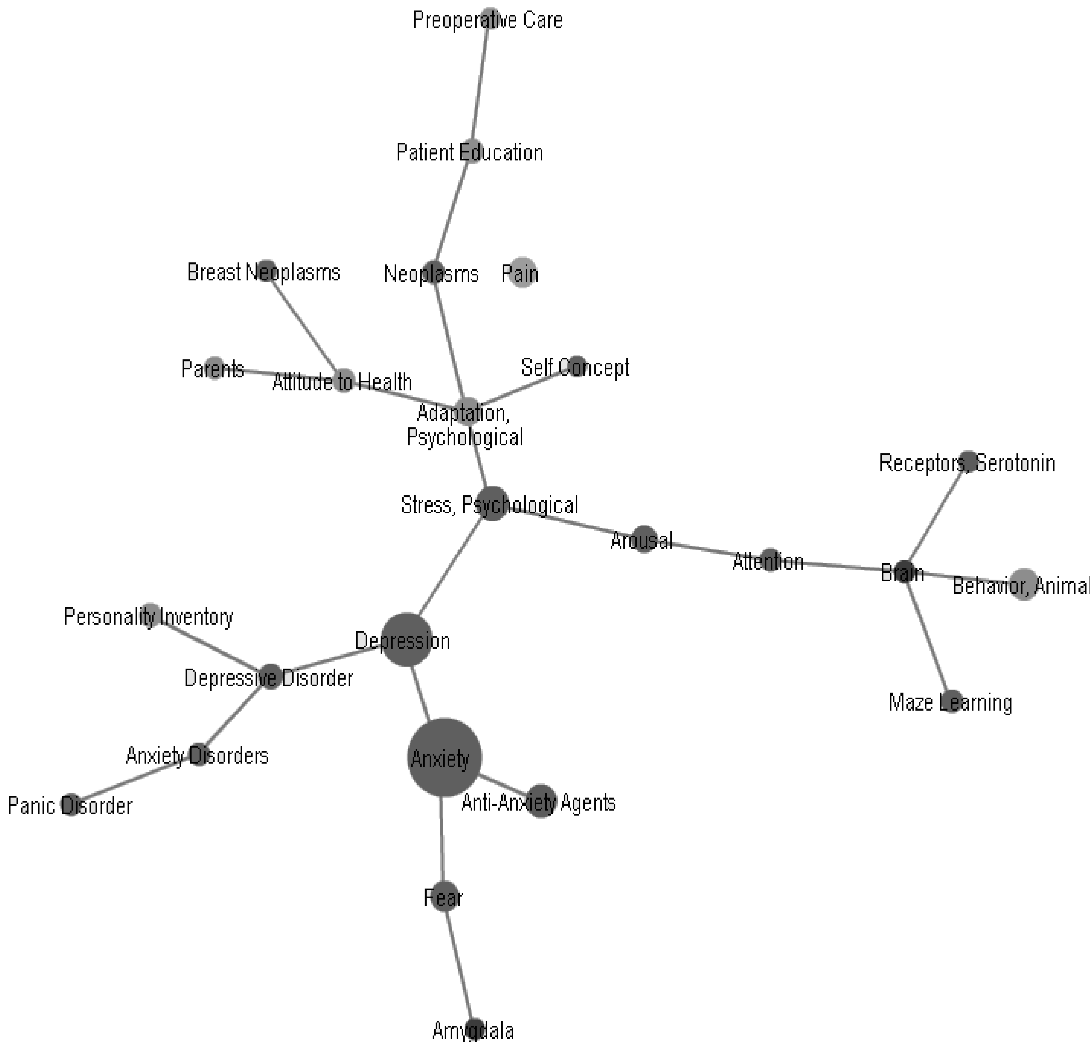


Figure 25. Anxiety PFNET Illustrating Interactivity of Displays

As with the previous depression PFNET, participants had the same option to click on any of the 25 terms to continue deeper into the interface display. As noted in the following Figure 26, after the second interface level, the number of term possibilities increases exponentially.

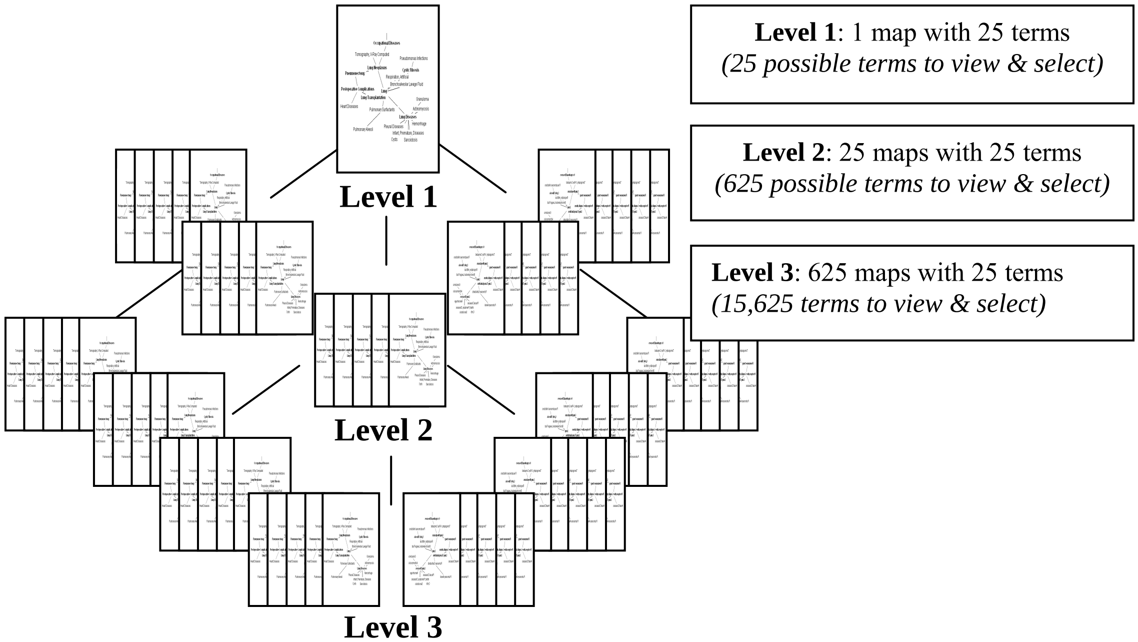


Figure 26. Possible Term/Term Phrases to Choose From on Display Interface Levels

For this analysis, the display terminology incorporated from the maps was labeled according to the map-level the term was taken from. A term was compared against all term/term phrases found on the first level of the display. If a term was found on that first display it was given a designation of one. If a term was not found on the primary interface level it was then compared against all term/term phrases found in level two and was then marked as such. Any terminology that was found in MeSH (meaning it was a display term), but not found within the list of terms for level one or level two was coded as a level three term. Because of the exponential possibilities after levels one and two, terms were not counted beyond the third display-level.

RQ 1.4: *Is there a relationship between display type and the interface level participants took terms from on the post-test?*
HYP 1.4: *There is a relationship between display type and the interface level participants took terms from on the post-test.*

Chi-Square analyses were conducted to determine if there was a relationship between display type (3 levels) and the interface level (3 levels) for post-test terms chosen by subjects. This analysis was conducted for both the cholesterol and depression terminology used by participants. The relationship between these two variables was significant for the cholesterol group, X^2 (4, N = 398) = 19.298, p = .001, and was also significant for the depression group, X^2 (4, N = 361) = 14.864, p = .005. In Table 10, looking across all groups below we find that 67% of the cholesterol terms were chosen from the first level of the display interface while only 22% came from the second level and 9.5% percent from the third level.

Table 10. Post-Test Chi Square Counts for Cholesterol Terms by Interface Level

Interface level of Term used Post-test		Display type			
		LIST	SOM	PFNET	Total
Level 1	Count	83	84	102	269
	Expected Count	93.9	91.2	83.8	269.0
	% of Total	20.9%	21.1%	25.6%	67.6%
Level 2	Count	37	36	18	91
	Expected Count	31.8	30.9	28.4	91.0
	% of Total	9.3%	9.0%	4.5%	22.9%
Level 3	Count	19	15	4	38
	Expected Count	13.3	12.9	11.8	38.0
	% of Total	4.8%	3.8%	1.0%	9.5%
Total	Count	139	135	124	398
	Expected Count	139.0	135.0	124.0	398.0
	% of Total	34.9%	33.9%	31.2%	100.0%

We also find across all groups the topic of depression has a similar allocation of terms to that of cholesterol with 64% of the depression terms chosen by participants from the first level of the display interface, 23% from the second level and 12.2% percent from the third level (see Table 11).

Table 11. Post-Test Chi Square Counts for Depression Terms by Interface Level

Interface level of Term used Post-test		Display type			
		LIST	SOM	PFNET	Total
Level 1	Count	75	64	92	231
	Expected Count	74.9	72.9	83.2	231.0
	% of Total	20.8%	17.7%	25.5%	64.0%
Level 2	Count	24	29	33	86
	Expected Count	27.9	27.2	31.0	86.0
	% of Total	6.6%	8.0%	9.1%	23.8%
Level 3	Count	18	21	5	44
	Expected Count	14.3	13.9	15.8	44.0
	% of Total	5.0%	5.8%	1.4%	12.2%
Total	Count	117	114	130	361
	Expected Count	117.0	114.0	130.0	361.0
	% of Total	32.4%	31.6%	36.0%	100.0%

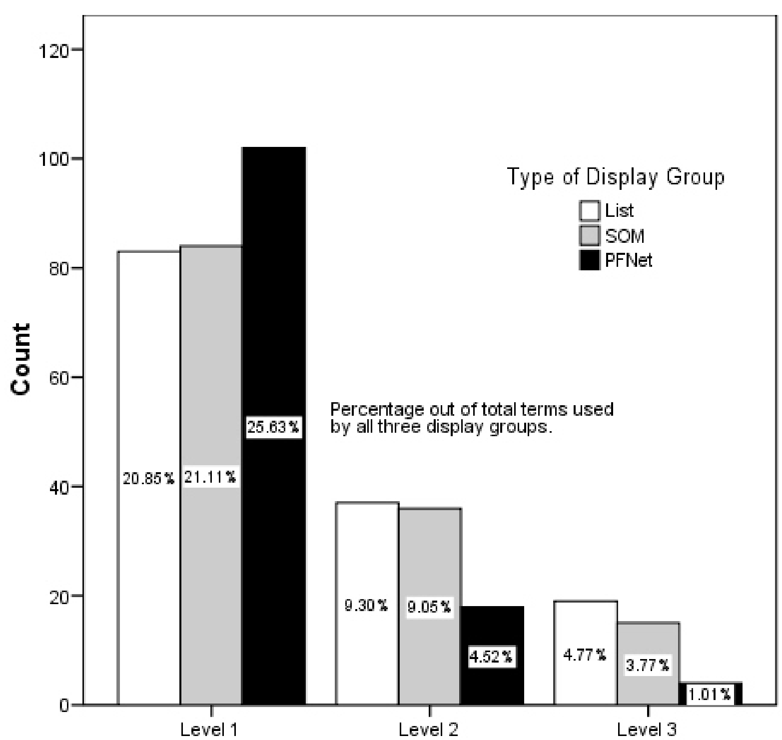


Figure 27. Percentage of Post-Test Cholesterol Terms by Display Interface Level

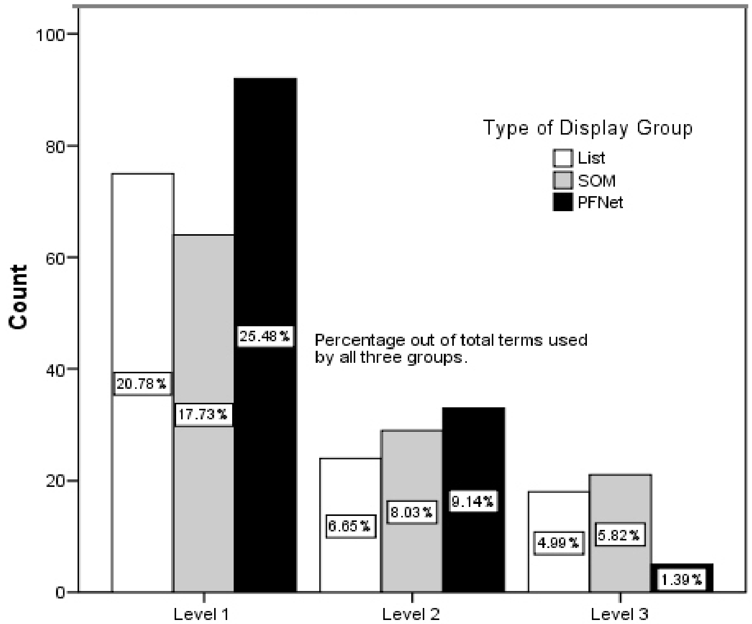


Figure 28. Percentage of Post-Test Depression Terms by Display Interface Level

For the Chi-square analyses on interface level of post-test terms by display type for both cholesterol and depression, we find that the level of the display terms were taken from and display type are not independent. We fail to reject the null hypothesis that interface level terminology was chosen from is independent from display type.

Based on the counts in CH Table 10 for the topic of cholesterol we note that the PFNET group was more likely to take their terms from the first level of the display than the LIST and SOM groups and less likely to choose terms from level two and three of the display. From DE Table 11, we note that for depression the PFNET group was more likely to take terms from the first interface level and less likely to use terms from the third level than the LIST and SOM groups. This is an indication that arrangement of the concepts for the different displays did have an impact on the terms participants chose to use on the post-test instrument. In making this statement it is important to note that this is not an indication of level of map exploration. For this particular research exercise that was outside the scope of data collection.

RQ1.5: Information Behavior –Search Statement Specificity

On both the pre- and post-test, participants were asked "please state what you are looking for". These search statements were used as an indication of search topic formulation. It was expected that the pre-test statements would be more vague and general in nature while the post-test statements would be more specific. All of the pre- and post-test statements for both cholesterol and depression were evaluated by three academic librarians for level of specificity. Librarians were given the statements in random order and asked to evaluate each of the 240 statements on a five-point scale of narrow to broad. Results were compiled and average score of specificity was computed for each statement.

RQ 1.5: *Is there a change in the level of specificity of search statements pre-test to post-test by display type?*

HYP 1.5: *There is a difference between display types in the specificity of search statements pre-test and post-test*

Mixed between-within ANOVA with one between subject factor (display type) and two within subject factors (pre-post and query topic) was used to test the effect of display type on the dependent variable (search statement specificity). As shown in Table 13, no differences in the level of specificity by display type were found.

Table 12: Descriptive Statistics for Search Statement Specificity

	Display Type	Mean	Std. Deviation	N
CH-PRE Level of Specificity.	LIST	3.65	1.137	20
	SOM	3.25	1.164	20
	PFNET	3.75	1.020	20
	Total	3.55	1.111	60
CH-POST Level of Specificity.	LIST	3.60	1.046	20
	SOM	3.65	1.268	20
	PFNET	3.75	1.209	20
	Total	3.67	1.160	60
DE-PRE- Level of Specificity.	LIST	3.55	1.504	20
	SOM	3.45	1.356	20
	PFNET	3.90	1.071	20
	Total	3.63	1.314	60
DE-POST Level of Specificity.	LIST	4.05	.945	20
	SOM	3.50	1.277	20
	PFNET	3.55	1.317	20
	Total	3.70	1.197	60

These results also show that there is no significant difference in the level of specificity of statements from pre- to post-test after subjects were exposed to the independent variable of display type [$F(1, 57) = .764$, $p = .386$). Based on these results it would appear that the display has no impact on early development of search statements. In fact looking at the means for the cholesterol search statements, participants almost stayed constant. With the depression search statements, the opposite seemed to happen as the PFNET group means show that the shift was to become less specific in their statements.

Table 13: ANOVA Within Subject Results for Search Statement Specificity by Display Type

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Pre-Post	.504	1	.504	.764	.386
Pre-Post *by* Group	2.133	2	1.067	1.616	.208
Error (Pre-Post)	*37.613*	*57*	*.660*		
Query Topic	.204	1	.204	.144	.705
Query Topic *by* Group	.433	2	.217	.153	.858
Error (Query Topic)	*80.613*	*57*	*1.414*		
Pre-Post *by* Query Topic	.038	1	.038	.056	.815
Pre-Post *by* Query Topic *by* Group	2.700	2	1.350	1.998	.145
Error (Pre-Post by Query Topic)	*38.513*	*57*	*.676*		

This failure to establish a significant difference between groups might be attributed to the artificial query statements. Because it was important for participants to explore the maps, the queries were purposefully kept general in nature which in turn might have led to the search statements staying similar in nature from pre- to post-test.

RQ1: Review of Information Behavior

In this section of the results we explored the dependent variable of information behavior by looking at the search terms and search statements elicited from participants. Using inferential statistics we explored the number of terms used pre- and post-test, the number of full MeSH terms incorporated into the post-test response, the number of full and partial MeSH terms, the interface level terms were taken from, and the specificity of participants' search statements. We found significant general effects pre- to post-test across all display types when it came to number of terms used and the incorporation of full/partial MeSH terms. We also found a statistically significant relationship between display type and the interface level terminology was chosen from. The CHI-SQUARE analysis showed that overall the PFNET group was more likely to use terminology from the first interface level and much less likely to use terminology from the lower interface levels compared to the LIST and SOM groups. Lastly, we also pointed out a potentially meaningful trend in the information behavior data which suggests the PFNET display is slightly ahead of the SOM and LIST groups when we compare the post-test means.

The next research questions shift the focus away from behavior toward the impact the displays have on of the affective feelings of the user. These questions explore the perceptions of satisfaction, confidence, success and relevance based on the terms participants chose.

RQ2: SELF-REPORTED PERCEPTIONS OF PERFORMANCE

Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher's self-perceptions of performance?

The second dependent variable examined was self-perceptions of performance. Participants' perception of performance refers to self-reported feelings about the terms they decided to use for searching on the medical topic. Participants were asked to circle the most appropriate selection on a scale of 1(not at all) to 5 (very) for the below statements both pre- and post-test:

- I am satisfied with the terms I used above.
- I am confidence these terms will help me locate resources for the assigned task.

- I would be successful locating information for this project using the terms above.
- I feel the terms I used above are relevant to the search topic.

Participant self-perceptions of performance were evaluated using a mixed between-within ANOVA with one between subject factor (display type) and two within subject factors (pre-post and query topic).

RQ2.1: Self-Reported Perceptions of Satisfaction

RQ2.1a: *Is there a difference in participants' self-perceptions of satisfaction in the terms they chose pre- and post-test* ***by display type?***

HYP 2.1a: *There is a difference in self-perceptions of satisfaction in the terms they chose pre- and post-test* ***by display type.***

A mixed between-within ANOVA was conducted to compare scores of satisfaction in terminology chosen by participants on the pre- and post-test. The means and standard deviations are presented in Table 14, following.

Table 14. Descriptive Statistics for Participant Self-Perceptions of Satisfaction

SATISFACTION	Display type	Mean	Std. Deviation	N
Cholesterol PRE: I am satisfied with the terms I used.	LIST	3.10	1.119	20
	SOM	3.35	1.040	20
	PFNET	3.05	1.146	20
	Total	3.17	1.092	60
Cholesterol POST: I am satisfied with the terms I used.	LIST	4.10	.912	20
	SOM	4.10	.968	20
	PFNET	4.10	.912	20
	Total	4.10	.915	60
Depression PRE: I am satisfied with the terms I used.	LIST	3.45	1.146	20
	SOM	3.85	.875	20
	PFNET	3.75	1.118	20
	Total	3.68	1.049	60
Depression POST: I am satisfied with the terms I used.	LIST	3.90	.852	20
	SOM	4.25	.550	20
	PFNET	4.45	.759	20
	Total	4.20	.755	60

There was a significant effect for the within subject factor of pre- to post-test across all groups [$F(1,57) = 37.995$, $p < .001$]. Based on the above analysis we find that across all display types there is a significant increase in satisfaction scores pre- to post-test; however, there is no difference between the display types in level of satisfaction[$F(1,57) = .699$, $p = .501$]. This means that no one display type instilled more feelings of satisfaction in users based on the terms they chose. While we did not find a difference in how the arrangement of terms might impact satisfaction, there is a statistically significant difference on the second within subject variable of medical topic.

RQ2.1b: *Is there a difference in participants' self-perceptions of satisfaction in the terms they chose pre- and post-test* ***by query topic****.*
HYP 2.1b: *There is a difference in participants' self-perceptions of satisfaction in the terms they chose pre- and post-test* ***by query topic****.*

We do find there is a significant within subjects effect of query topic on satisfaction [F(1,57) = 8.785, p < .005] . There is a difference in scores of satisfaction by medical topic.

Table 15. ANOVA Within Subject Results for Participant Self-Perceptions of Satisfaction

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Pre-Post	31.537	1	31.537	37.995	.001***
Pre-Post *by* Group	.900	2	.450	.542	.584
Error (Pre-Post)	*47.312*	*57*	*.830*		
Query Topic	5.704	1	5.704	8.785	.004**
Query Topic *by* Group	2.033	2	1.017	1.566	.218
Error (Query Topic)	*37.012*	*57*	*.649*		
Pre-Post *by* Query Topic	2.604	1	2.604	10.593	.002**
Pre-Post *by* Query Topic by Group	.133	2	.067	.271	.763
Error (Pre-Post by Query Topic)	*14.013*	*57*	*.246*		

(**p < .01, ***p < .001)

Comparisons reveal that mean satisfaction scores for the topic of depression were significantly higher pre-test than satisfaction scores for the cholesterol query. We can see the difference in a more visual format in Figure 29, following.

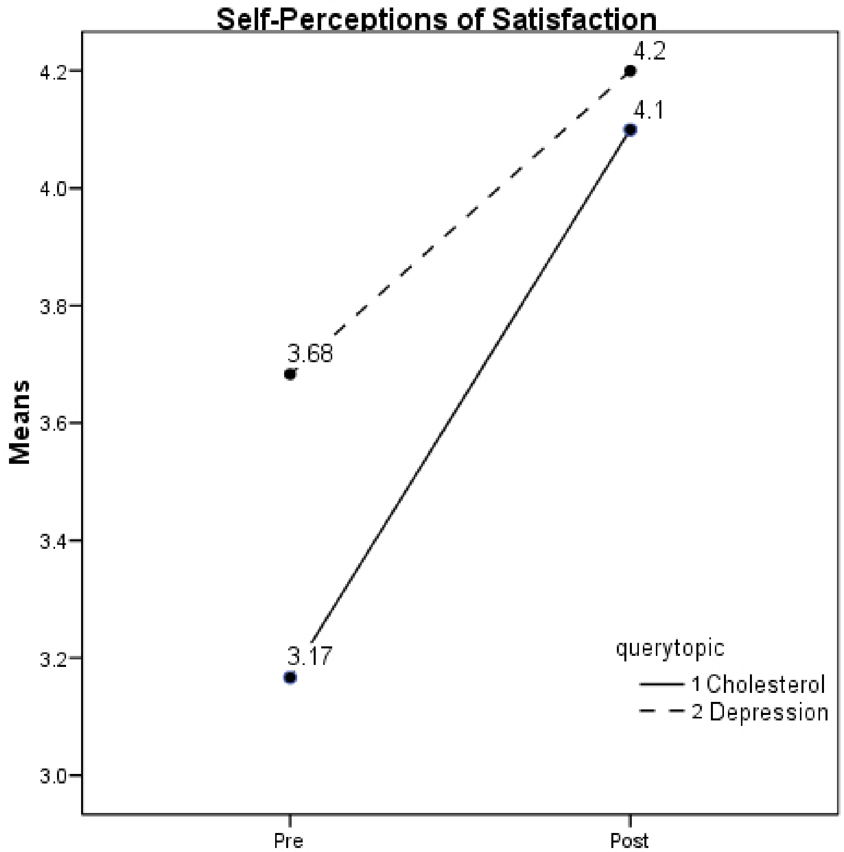


Figure 29. Pre- to Post-Test Means on Self-Perceptions of Satisfaction for Cholesterol & Depression

Further analysis finds there was a significant interaction effect of Pre-post by query topic [F(1,57) = 10.593, p = .002]. This indicates that query topic had different effects on people's ratings of satisfaction in the terms they chose to use. We know that all participants scores of satisfaction improved after seeing the displays; however, we find that with the topic of cholesterol, after seeing the displays, scores of satisfaction in chosen terms significantly shifted to almost match those of depression. The cholesterol scores post-test come within 0.1 of those for depression. Our next self-perception analysis looks at confidence.

RQ2.2: Self-Reported Perceptions of Confidence

RQ2.2a: *Is there a difference in participants' self-perceptions of confidence in the terms they chose pre- and post-test* ***by display type?***

HYP 2.2a: *There is a difference in participants' self-perceptions of confidence in the terms they chose pre- and post-test* ***by display type****.*

Participant self-perceptions of confidence in the terms they chose was evaluated using a mixed between-within ANOVA with one between subject factor (display type) and two within subject factors (pre-post and query topic). The means and standard deviations are presented in Table 16, following.

Table 16. Descriptive Statistics for Participant Self-Perceptions of Confidence

	Display type	Mean	Std. Deviation	N
Cholesterol PRE: I am confident these terms will help me.	LIST	3.25	1.118	20
	SOM	3.65	.813	20
	PFNET	3.65	.933	20
	Total	3.52	.965	60
Cholesterol POST: I am confident these terms will help me.	LIST	4.10	.788	20
	SOM	4.25	.851	20
	PFNET	4.30	.801	20
	Total	4.22	.804	60
Depression PRE: I am confident these terms will help me.	LIST	3.75	.851	20
	SOM	3.90	.852	20
	PFNET	4.00	.918	20
	Total	3.88	.865	60
Depression POST: I am confident these terms will help me.	LIST	4.05	.686	20
	SOM	4.30	.657	20
	PFNET	4.55	.759	20
	Total	4.30	.720	60

There was a significant main effect for the within subjects variable of pre- to post-test across all groups [F(1,57) = 27.773, p < .001]. Based on the above analysis we find that across all display types there is a significant increase in confidence scores pre- to post-test. We can see in Table 17, that all means increased significantly from pre- to post-test. There is no difference on the between subjects variable of display types [F(2,57) = .1.761, p = .181]. This means we did not establish a difference in how arrangement of terms might differentially impact feelings of confidence in user-chosen terms.

Table 17. ANOVA Within Subject Results for Participant Self-Perceptions of Confidence

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Pre-Post	18.704	1	18.704	30.196	.001***
Error (Pre-Post)	*36.546*	*59*	*.619*		
Query Topic	3.037	1	3.037	7.402	.009**
Error (Query Topic)	*24.212*	*59*	*.410*		
Pre-Post *by* Query Topic	1.204	1	1.204	4.168	.046*
Error (Pre-Post by Query Topic)	*17.046*	*59*	*.289*		

(*p < .05, **p < .01, ***p < .001)

RQ2.2b: *Is there a difference in participants' self-perceptions of confidence in the terms they chose pre- and post-test* ***by query topic***.

HYP 2.2b: *There is a difference in participants' self-perceptions of confidence in the terms they chose pre- and post-test* ***by query topic.***

Again, when we look at the other within subject factor of medical topic we do find there is a significant difference for query topic on confidence [F(1,57) = 8.096, p = .006] . As also found in the previous research question, comparisons reveal that mean satisfaction scores for the topic of depression pre-test were higher than the mean satisfaction scores for the cholesterol query.

As also with scores in confidence, further analysis found there was a significant interaction effect of Pre-Post by query topic [F(1,57) = 4.168, p = .046]. This indicates that query topic had different effects on people's ratings of confidence in the terms they chose to use. We know that all participants' scores of confidence improved after seeing the displays; however, we find that with the topic of cholesterol, after seeing the displays, scores of confidence in chosen terms significantly shifted to almost match the scores on the topic of depression. The cholesterol scores post-test come within 0.1 of those for depression. Figure 30 below illustrates this difference through the means for the query topics pre- and post-test.

Figure 30. Pre- and Post-Test Means on Self-Perceptions of Confidence for Cholesterol and Depression

RQ2.3: Self-Reported Perceptions of Success

RQ2.3a: *Is there a difference in participants' self-perceptions of success in the terms they chose pre- and post-test* ***by display type?***
HYP 2.3a: *There is a difference in participants' self-perceptions of success in the terms they chose pre- and post-test* ***by display type****.*

Self-perceptions of success in the terms they chose was evaluated using a mixed between-within ANOVA with one between subject factor (display type) and two within subject factors (pre-post and query topic). The means and standard deviations are presented in Table 18, following.

Table 18. Descriptive Statistics on Participant Self-Perceptions of Success

SUCCESS	Display type	Mean	Std. Deviation	N
Cholesterol PRE: I would be successful locating info using these terms	LIST	3.45	.999	20
	SOM	3.75	.716	20
	PFNET	3.80	.951	20
	Total	3.67	.896	60
Cholesterol POST: I would be successful locating info using these terms	LIST	4.10	.788	20
	SOM	4.25	.851	20
	PFNET	4.35	.988	20
	Total	4.23	.871	60
Depression PRE: I would be successful locating info using these terms	LIST	3.70	1.031	20
	SOM	3.90	.718	20
	PFNET	3.85	1.089	20
	Total	3.82	.948	60
Depression POST: I would be successful locating info using these terms	LIST	4.15	.671	20
	SOM	4.35	.745	20
	PFNET	4.55	.759	20
	Total	4.35	.732	60

There was a significant main effect for success pre- to post-test across all groups [F(1,57) = 23.446, p < .001]. Based on the above analysis we find that across all display types there is a significant increase in feelings of success pre- to post-test. We did not find a difference between the display types in level of success pre- to post-test [F(2,57) = 1.128, p = .331]. This means that no one display type instilled feelings of success in users chosen terms over another group.

Table 19. ANOVA Within Subject Results on Participant Self-Perceptions of Success

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Pre-Post	18.150	1	18.150	23.446	.001***
Pre-Post by Group	.225	2	.113	.145	.865
Error (Pre-Post)	*44.125*	*57*	*.774*		
Query Topic	1.067	1	1.067	2.139	.149
Query Topic by Display Type	.008	2	.004	.008	.992
Error (Query Topic)	*28.425*	*57*	*.499*		
Pre-Post by Query Topic	.017	1	.017	.093	.761
Pre-Post by Query Topic by Display Type	.308	2	.154	.864	.427
Error (Pre-Post by Query Topic)	*10.175*	*57*	*.179*		

(***p < .001)

RQ2.3b: *Is there a difference in participants' self-perceptions of* ***success*** *in the terms they chose pre and post-test* ***by query topic****.*

HYP 2.3b: *There is a difference in participants' perceptions of* ***success*** *in the terms they chose pre and post-test* ***by query topic.***

Unlike the two previous analyses on participant perception of satisfaction and confidence there is no significant difference in feelings of success between the query topics of cholesterol and depression, [F(1,57) = 2.139, p = .149]. In this analysis the depression topic did not instill more feelings of success with the terms participants chose than the cholesterol topic. Next we move on to look at the last perception of performance measured, relevance.

RQ2.4: Self-Reported Perceptions of Relevance

RQ2.4a: *Is there a difference in participants' self-perceptions of relevance based on the terms used pre- and post-test by* ***display type****?*
HYP 2.4a: *There is a difference in participants' self-perceptions of relevance based on the terms used pre- and post-test by* ***display type****.*

Participants' self-perceptions of relevance for the terms they chose was evaluated using a mixed between-within ANOVA with one between subject factor (display type) and two within subject factors (pre-post and query topic). The means and standard deviations are presented in Table 20, following.

Table 20. Descriptive Statistics for Participant Self-Perceptions of Relevance

RELEVANCE	Display type	Mean	Std. Deviation	N
Cholesterol PRE: I feel the terms I used above are relevant to the search topic.	LIST	3.90	.912	20
	SOM	4.15	.813	20
	PFNET	3.95	.999	20
	Total	4.00	.902	60
Cholesterol POST: I feel the terms I used above are relevant to the search topic.	LIST	4.30	.733	20
	SOM	4.30	.801	20
	PFNET	4.35	1.040	20
	Total	4.32	.854	60
Depression PRE: I feel the terms I used above are relevant to the search topic.	LIST	4.20	.768	20
	SOM	4.05	.887	20
	PFNET	4.20	.768	20
	Total	4.15	.799	60
Depression POST: I feel the terms I used above are relevant to the search topic.	LIST	4.30	.733	20
	SOM	4.35	.745	20
	PFNET	4.65	.671	20
	Total	4.43	.722	60

There was a significant main effect for pre- to post-test increase in relevance across all groups [$F(1,57) = 10.753$, $p = .002$]. Based on the above, we find that across all display types there is a significant increase in feelings of relevance in terms selected pre- to post-test. However, again, we did not find a difference between the LIST, SOM or PFNET display types in relevance scores [$F(2,57) = .176$, $p = .981$]. No one display type instilled feelings of relevance in users chosen terms over another display type.

Table 21.ANOVA Within Subject Results for Participant Self-Perceptions of Relevance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Pre-Post	5.400	1	5.400	10.753	.002**
Pre-Post *by* Display Type	.475	2	.237	.473	.626
Error (Pre-Post)	*28.625*	*57*	*.502*		
Query Topic	1.067	1	1.067	2.209	.143
Query Topic *by* Display Type	.908	2	.454	.941	.396
Error (Query Topic)	*27.525*	*57*	*.483*		
Pre-Post *by* Query Topic	.017	1	.017	.060	.808
Pre-Post *by* Query Topic *by* Display Type	.558	2	.279	.999	.375
Error (Pre-Post by Query Topic)	*15.925*	*57*	*.279*		

(**p < .01)

RQ2.4b: *Is there a difference in participants' self-perceptions of relevance in the terms they chose pre- and post-test* ***by query topic?***

HYP 2.4b: *There is a difference in participants' self-perceptions of relevance in the terms they chose pre- and post-test* ***by query topic.***

As with the previous analysis of impact by query topic on success, there was also no significant difference for self-perceptions of relevance by query topic [F(1,57) = .2.209, p > .05]. There was no difference in self-reported relevance scores in chosen terms by medical topic.

RQ2: Review of Self-Reported Perceptions of Performance

In this section of the results we explored the dependent variable of self-perceptions of performance based on the terms they chose to use both pre- and post-test. This dependent variable was measured by asking participants to circle the most appropriate selection on a scale of 1(not at all) to 5 (very) for the below statements:

- I am satisfied with the terms I used above.

- I am confidence these terms will help me locate resources for the assigned task.
- I would be successful locating information for this project using the terms above.
- I feel the terms I used above are relevant to the search topic.

Using inferential statistics we explored participants feelings of satisfaction, confidence, success and relevance based on their term selection. We found that for each dependent variable there was a statistically significant increase in score pre- to post-test across ALL display types. On average each display type increased in feelings of satisfaction, confidence, success and relevance after seeing the independent variable of display type. For this set of analyses there was no difference by display type. Regardless of how the terms were arranged participants on average felt more confident, successful, satisfied and felt their terms were more relevant. What we did find in the self-perceptions of performance analysis (RQ2) was that two of these dependent variables; satisfaction and confidence, differed on the within subjects factor of query topic. Participants across all display types had more feelings of satisfaction and confidence in their chosen terms with the depression query topic than for the cholesterol topic. This was not true for feelings success or relevance. There was no significant difference between the query topics of cholesterol or depression for success or relevance.

The above results lend themselves to a theory to be expanded upon throughout the rest of the discussions. This difference in feelings of satisfaction and confidence by query topic might be attributed to the amount of information participants possess about depression versus cholesterol.

Initial data analysis showed that overall participants appeared to have more terms to describe the concept of depression than they did the concept of cholesterol. If we hypothesize that because participants seemed to have more vocabulary to describe depression, they might also possess more subject knowledge about depression then it would follow that participants might be more satisfied with their depression terms and more confident using them than they would with their cholesterol terminology. This theory will be further discussed in the next section on self-perceptions of knowledge and also woven into the larger discussion and conclusions.

RQ3: SELF-REPORTED PERCEPTIONS OF KNOWLEDGE

Do concept maps, used as an exploratory tool in the information retrieval environment impact the novice searcher's self-perceptions of knowledge?

The third dependent variable explored focused on participants self-perceptions of knowledge. Participants' self-perceptions of knowledge refers to perceptions of personal knowledge about the medical topic they are being asked to explore (cholesterol, depression). Participants were asked to answer three questions using a scale of 1 to 5 about their knowledge on the query topic used for the experiment.

- How would you rank the amount of knowledge you possess on this topic? (pre-test only) (little personal knowledge – considerable personal knowledge)
- I feel I know more about the topic than I did before searching. (post-test only) (not at all – very)
- My post-search knowledge has corrected what I knew before searching. (post-test only) (not at all – very)

RQ3.1: Self-Perceptions of Query Topic Knowledge Pre-Test

RQ3.1: *Is there a difference in participants' self-perceptions of query topic knowledge on the pre-test* ***by query topic****?*
HYP 3.1: *There is a difference in participants' self-perceptions of query topic knowledge on the pretest* ***by query topic.***

Participant self-perceptions of personal knowledge on the query topics of depression and cholesterol was evaluated using a mixed between-within ANOVA with one between subject factor (display type) and 1 within subject factor (query topic). The means and standard deviations are presented in Table 22, following.

Table 22. Descriptive Statistics for Pre-Test Participant Self-Perceptions of Knowledge

Prior Topic Knowledge	Display type	Mean	Std. Deviation	N
Cholesterol PRE: How would you rank amount of knowledge you possess on topic.	LIST	2.55	1.099	20
	SOM	2.60	1.188	20
	PFNET	2.50	.889	20
	Total	2.55	1.048	60
Depression PRE: How would you rank amount of knowledge you possess on topic.	LIST	2.75	.716	20
	SOM	3.00	1.124	20
	PFNET	3.05	1.317	20
	Total	2.93	1.071	60

There was no difference between the display types in level of pre-test knowledge which is as we would expect [$F(2,57) = .202$, $p = .818$]. We did find that the difference in participants' knowledge by query topic is significant [$F(1,57) = 4.297$, $p = .043$]. Participants on the pre-test instrument felt that they had more personal knowledge on the topic of depression than they did on the topic of cholesterol.

Table 23. ANOVA Within Subject Results for Participant Self-Perceptions of Pre-Test Knowledge

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Query Topic	4.408	1	4.408	4.297	.043*
Query Topic by Display Type	.617	2	.308	.301	.742
Error (Query Topic)	*58.475*	*57*	*1.026*		

(*p < .05)

RQ3.2: Self-Perceptions of Query Topic Knowledge Post-Test

RQ3.2: *Is there a difference in participant self-perception of query topic knowledge post-test* ***by display type?***

HYP 3.2: *There is a difference in participant self-perception of query topic knowledge post-test* ***by display type.***

Self-perceptions of personal knowledge about the query topics of cholesterol and depression post-test were evaluated using a mixed between-within ANOVA with one between subject factor (display type) and 1 within subject factor (query topic). The means and standard deviations are presented in Table 24, following.

Table 24. Descriptive Statistics for Participant Self-Perceptions of Post-Test Knowledge

MORE KNOWLEDGE	Display type	Mean	Std. Deviation	N
Cholesterol POST: I feel I know more about the topic than I did before searching.	LIST	3.00	1.124	20
	SOM	3.35	.933	20
	PFNET	3.20	1.281	20
	Total	3.18	1.112	60
Depression POST: I feel I know more about the topic than I did before searching.	LIST	2.90	1.071	20
	SOM	3.00	1.124	20
	PFNET	2.90	1.294	20
	Total	2.93	1.148	60

We did not find a difference in the level of post-test knowledge between the display types [F(2,57) = .259, p = .773]. We also fail to establish a difference in knowledge post-test by query topic [F(1,57) = 2.867, p = .096]. Participants on the post-test instrument did not feel they knew more about the medical topic than they did before searching. While this result is not significant, based on the previous analyses it is important to highlight that there is a slight difference in the means by query topic. Looking at Table 24, we find that participants overall felt they knew more about cholesterol post-test (*M* = 3.18) as they did about depression (*M* = 2.93).

Table 25. ANOVA Within Subject Results for Participant Self-Perceptions of Post-Test Knowledge

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Query Topic	1.875	1	1.875	2.867	.096
Query Topic by Display Type	.350	2	.175	.268	.766
Error (Query Topic)	*37.275*	*57*	*.654*		

If someone already possessed certain amounts of knowledge about depression it would then follow that they did not feel they necessarily knew more about that topic after searching. Whereas if they did not feel they possessed knowledge on cholesterol pre-test the means for knowing more after searching would be higher. This also supports the theory that participants had more personal knowledge about depression pre-test than they had personal knowledge about cholesterol.

RQ3.3: Self-Reported Perceptions of Corrected Knowledge

RQ3.3a: *Is there a difference in participant self-perceptions of corrected knowledge post-test* ***by display type?***

HYP 3.3a: *There is a difference in participant self-perceptions of corrected knowledge post-test* ***by display type.***

Participants in this experiment were asked to rate on a scale of 1 to 5: "My post-search knowledge has corrected what I knew before searching." Participant self-perceptions of corrected personal knowledge post-test on the query topics of depression and cholesterol was evaluated using a mixed between-within ANOVA with one between subject factor (display type) and 1 within subject factor (query topic). The means and standard deviations are presented in Table 26, following.

Table 26. Descriptive Statistics for Participant Self-Perceptions of Corrected Knowledge

	Display type	Mean	Std. Deviation	N
CHPost: My post-searching knowledge has corrected what I knew before searching.	LIST	2.80	1.240	20
	SOM	3.00	1.298	20
	PFNET	3.25	1.446	20
	Total	3.02	1.321	60
DEPost: My post-searching knowledge has corrected what I knew before searching.	LIST	2.70	1.174	20
	SOM	2.65	1.089	20
	PFNET	3.00	1.414	20
	Total	2.78	1.223	60

We were unable to establish a difference between the display types in amount of post-test corrected knowledge [$F(2,57) = .552$, $p = .579$].

RQ3.3b: *Is there a difference in participants' self-perceptions of corrected knowledge post-test by* ***query topic****?*

HYP 3.3b: There is a difference in participants' self-perceptions of corrected knowledge post-test by ***query topic****.*

With this analysis we also find that the difference in participants' self perceived knowledge about cholesterol and depression is not significantly different [F(1,57) = 3.17, p = .059]. Again, while not a statistically significant result, it is meaningful to point out that participants felt their post-search knowledge had corrected what they knew with cholesterol (M = 3.02) MORE than with depression (M = 2.78).

Table 27. ANOVA Within Subject Results for Participant Self-Perceptions of Corrected Knowledge

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Query Topic	1.633	1	1.633	3.717	.059
Query Topic *by* Display Type	.317	2	.158	.360	.699
Error (Query Topic)	*25.050*	*57*	*.439*		

This is additional support for the theory that participants had more personal knowledge about depression pre-test. If a participant already possesses an adequate amount of knowledge about depression, it would then follow that the same participant would feel that their post-search knowledge was less corrected on depression. The opposite then might be true on the cholesterol topic. Participants reported they had less knowledge about cholesterol pre-test and they on average have a higher means post-test.

RQ3: REVIEW of Self-Reported Perceptions of Knowledge

In this section of the results we explored the dependent variable of personal knowledge as perceived by participants. This was measured by asking participants to circle the most appropriate selection on a scale of 1(not at all) to 5 (very) for the below statements:

- How would you rank the amount of knowledge you possess on this topic? *
- I feel I know more about the topic than I did before searching.
- My post-search knowledge has corrected what I knew before searching. *(Note. Suggestive but not significant $p = .059$)*

*(*Significant $p < .05$ between query topic of depression and cholesterol)*

Using inferential statistics we explored perceived knowledge based on responses to the above. We found that there was no statistically significant difference between the display types for any of the questions. However, we did find that on two of the above questions there was a significant difference between the topics of cholesterol and depression. Subjects felt that their amount of pre-test knowledge on the topic of depression was greater than their pre-test knowledge of cholesterol. Falling into line with that, subjects also felt that their post-search knowledge was corrected more for cholesterol than for depression after exploring the displays. On the question of knowing more about the topic post-exploration, we noted that while there was not a statistically significant difference, the means suggest participants felt they learned more about cholesterol than they did depression.

These findings also support the theory introduced earlier about the impact prior depression knowledge had on the dependent variables. We hypothesized that the difference in feelings of satisfaction and confidence by query topic might be attributed to the amount of subject knowledge participants felt they possessed about depression versus cholesterol. In this section of analysis we did find that participants rated their pre-test subject knowledge on depression higher than they rated their pre-test cholesterol knowledge. And, participants also rated the amount of corrected knowledge post-test greater for cholesterol than depression further strengthening this hypothesis.

RQ4: OVERALL PERCEPTIONS OF DISPLAY USEFULNESS

Do concept maps, used as an exploratory tool in the information retrieval environment impact participants' overall perceptions of usefulness?

The last dependent variable explored was overall reaction to the different display types. This refers to self-reported reactions on questions about search formulation, knowledge, sense and understanding, and use. Participants were asked to answer a series of questions and to mark their responses on a scale of "strongly agree" to "strongly disagree".

RQ4.1: Perceptions of Search Formulation Help

RQ4.1: *Is there a difference in participant perceptions of system support for* ***search formulation help*** *by display types?*

HYP 4.1: *There is a difference in participant perceptions of system support for* ***search formulation help*** *by display types.*

Search formulation questions

The system helped me formulate my search.
After using the system I decided to change what I was looking for.
Using the system made coming up with search terms easier.

A one-way between groups analysis of variance was conducted to explore the impact of display type on participant perceptions of how the system helped them formulate their search. Subjects were randomly assigned to three groups; LIST, SOM, PFNET. The means and standard deviations are presented in Table 28, following.

Table 28. Descriptive Statistics for Participant Perceptions of System Support for Search Formulation

Question	Display Type	N	Mean	Std. Deviation
The system helped me to formulate my search.	LIST	20	3.15	.875
	SOM	20	3.10	.852
	PFNET	20	3.60	.598
	Total	60	3.28	.804
After using the system I decided to change what I was looking for.	LIST	19	2.47	1.124
	SOM	20	3.10	.718
	PFNET	20	2.85	1.040
	Total	59	2.81	.991
Using the system made coming up with search terms easier.	LIST	20	3.35	.933
	SOM	20	3.60	.503
	PFNET	20	3.70	.571
	Total	60	3.55	.699

We did not find a difference in how arrangement of terms on the different displays might impact the novice searcher on perceptions of system help on search formulation.

Table 29. ANOVA Results Participant Perceptions of System Support for Search Formulation

Question		Sum of Squares	df	Mean Square	F	Sig.
The system helped me to formulate my search.	Between Groups	3.033	2	1.517	2.459	.095
	Within Groups	35.150	57	.617		
	Total	38.183	59			
After using the system I decided to change what I was looking for.	Between Groups	3.862	2	1.931	2.037	.140
	Within Groups	53.087	56	.948		
	Total	56.949	58			
Using the system made coming up with search terms easier.	Between Groups	1.300	2	.650	1.345	.269
	Within Groups	27.550	57	.483		
	Total	28.850	59			

The next questions address participants' perceptions of overall sense and understanding.

RQ4.2: Perceptions of Overall Display Sense and Understanding

RQ4.2: *Is there a difference in participant perceptions of overall **display sense and understanding** by display types?*

HYP 4.2: *There is a difference in participant perceptions of overall **display sense and understanding** by display types.*

Sense and understanding questions

- The system was easy to understand.
- I understood the medical topics better using the system.
- The system made no sense at all.

A one-way between groups analysis of variance was conducted to explore the impact of display type on participant perceptions of understanding the displays, the medical topics and whether the systems made any sense. Subjects were randomly assigned to three groups; LIST, SOM, PFNET. The means and standard deviations are presented in Table 30 below.

Table 30. Descriptive Statistics for Participant Self-Perceptions of Understanding and Sense

Question	Display	N	Mean	Std. Deviation
The system was easy to understand.	LIST	20	3.15	.813
	SOM	20	2.50	.827
	PFNET	20	3.55	.605
	Total	60	3.07	.861
I understood the medical topics better using the system.	LIST	20	2.85	.933
	SOM	20	2.90	.788
	PFNET	20	3.25	.910
	Total	60	3.00	.883
The system made no sense at all.	LIST	20	1.70	1.031
	SOM	20	2.25	.910
	PFNET	20	1.30	.733
	Total	60	1.75	.968

Table 31.ANOVA Results on Participant Self-Perceptions of Understanding and Sense

Question	Display	Sum of Squares	df	Mean Square	F	Sig.
The system was easy to understand.	Between Groups	11.233	2	5.617	9.851	.001***
	Within Groups	32.500	57	.570		
	Total	43.733	59			
I understood the medical topics better using the system	Between Groups	1.900	2	.950	1.228	.301
	Within Groups	44.100	57	.774		
	Total	46.000	59			
The system made no sense at all.	Between Groups	9.100	2	4.550	5.620	.006**
	Within Groups	46.150	57	.810		
	Total	55.250	59			

(**$p < .01$, ***$p < .001$)

There was a statistically significant difference at the $p < .001$ level in participants perceptions of how easy the system was to understand [$F(2, 57) = 9.851$, $p < .001$]. The effect size, calculated using eta squared, was .25 a large effect (Cohen, 1988). Post-hoc comparisons using the Tukey HSD test indicate that the mean score for the Region-SOM group ($M = 2.50$, $SD = .827$) was significantly different from both the LIST group($M = 3.15$, $SD = .813$) and the Link-PFNET group($M = 3.55$, $SD = .605$). It would appear the SOM participants rated that display the lowest of the three to understand. If we ranked these displays based on the above means, the PFNET group rated the system easier to understand than the participants in the LIST group, and participants in the LIST group rated their display easier to understand than participants in the SOM group. It appears the SOM display was the least easy to understand.

There was no significant difference between the groups on participants self-reported understanding of the medical topics using the difference displays, but there was a statistically significant difference between groups on the question: “The system made no sense at all” [F(2, 57) = 5.620, p = .006]. The effect size, calculated using eta squared, was .16 a large effect (Cohen 1998). Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the Region-SOM group (*M* = 2.25, *SD* = .910) was significantly different from the Link-PFNET group(*M* = 1.30, *SD* = .733).

Table 32. Multiple Comparisons on Self-Perceptions of Understanding and Sense by Display Type

Question	Display Type	Display Type	Mean Difference (I-J)	Std. Error	Sig.
The system was easy to understand.	SOM	LIST	-.650(*)	.239	.023*
		PFNET	-1.050(*)	.239	.001***
The system made no sense at all.	SOM	LIST	.550	.285	.139
		PFNET	.950(*)	.285	.004**

(*p < .05, **p < .01, ***p < .001)

Participants in the SOM group agreed more with the statement, “the system made no sense at all” than the participants in the Link-PFNET group. This suggests that the novice searchers in this experiment were able to make sense of the Link-PFNET display while the participants in the Region-SOM group were not able to make as much sense of that display. The next question asked participants about their overall perceptions of learning about the medical topic and knowledge about Medical Subject headings.

RQ4.3: Perceptions of Overall Learning and Knowledge

RQ4.3: *Is there a difference in participant perceptions of* ***learning and knowledge*** *by display types?*
HYP 4.3: *There is a difference in participant perceptions of* ***learning and knowledge*** *by display types.*

Learning and knowledge questions

I learned about the medical topic with the system.
I know more about Medical Subject Headings.

A one-way between groups analysis of variance was conducted to explore the impact of display type on participants' perceptions of learning about the medical topic and knowledge of subject headings. Subjects were randomly assigned to three groups; LIST, SOM, PFNET. The means and standard deviations are presented in Table 33, following.

Table 33. Descriptive Statistics for Overall Participant Self-Perceptions of Learning and Knowledge

		N	Mean	Std. Deviation
I learned about the medical topic with the system.	LIST	19	2.74	.872
	SOM	20	2.75	.851
	PFNET	20	3.20	.768
	Total	59	2.90	.845
I know more about Medical Subject Headings.	LIST	18	2.89	1.023
	SOM	20	3.05	.759
	PFNET	20	3.50	.688
	Total	58	3.16	.854

There were no significant differences at the $p < .05$ level between groups on questions about learning and knowledge. It is worth mentioning that there was a difference between the means for the LIST ($M = 2.89$, $SD = 1.023$) and the PFNET group

(M = 3.50, SD = .688) [F(2, 55) = 2.82, p = .068]. This is not a statistically significant difference, but once again we find the PFNET group standing out from the LIST group. The means suggest participants felt they knew more about Medical Subject headings after using the PFNET display than the SOM and LIST groups.

Table 34. ANOVA Results for Participant Overall Self-Perceptions of Learning and Knowledge

		Sum of Squares	df	Mean Square	F	Sig.
I learned about the medical topic with the system.	Between Groups	2.756	2	1.378	1.997	.145
	Within Groups	38.634	56	.690		
	Total	41.390	58			
I know more about Medical Subject Headings.	Between Groups	3.876	2	1.938	2.825	.068
	Within Groups	37.728	55	.686		
	Total	41.603	57			

RQ4.4: Perceptions of Overall Visual Appeal

RQ4.4: *Is there a difference in participant perceptions of* ***visual appeal*** *by display types?*

HYP 4.4: *There is a difference in participant perceptions of* ***visual appeal*** *by display types.*

Visual appeal question:

- The system was visually appealing.

A one-way analysis of variance was conducted to explore the impact of display type on reactions of the display's visual appeal. Subjects were randomly divided into three groups (LIST, SOM, PFNET). There was a statistically significant difference at the p < .05 level in means on the visual appeal of the system [F(2, 56) = 3.351, p = .042]. The effect size, calculated using eta squared, was .10 which is considered a large effect (Cohen, 1988). Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the PFNET group (M = 3.20, SD = .834) was significantly different from the LIST group (M = 2.45, SD = .887).

Table 35. Descriptive Statistics on Participant Perceptions of Visual Appeal

The system was visually appealing.	N	Mean	Std. Deviation	Std. Error
LIST	20	2.45	.887	.198
SOM	19	2.74	1.046	.240
PFNET	20	3.20	.834	.186
Total	59	2.80	.961	.125

Table 36. ANOVA Results on Participant Perceptions of Visual Appeal

The system was visually appealing.	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5.725	2	2.863	3.351	.042*
Within Groups	47.834	56	.854		
Total	53.559	58			

(*p < .05)

Table 37. Multiple Comparisons on Participant Perceptions of Visual Appeal by Display Type

Question	Display Type	Display Type	Mean Difference (I-J)	Std. Error	Sig.
The system was visually appealing	PFNET	LIST	.750(*)	.292	.034*
		SOM	.463	.296	.269

(*p < .05)

Participants overall found the PFNET display more visually appealing than the LIST group found their display. Looking at the means, participants in the Region-SOM group fall between the PFNET and LIST on visual appeal. The last questions participants responded addressed preference of use of the system and future use.

RQ4.2: Perceptions of Overall Current and Future System Use

RQ4.5: *Is there a difference in participant perceptions of* ***current and future system use*** *by* display type?

HYP 4.5: *There is a difference in participant perceptions of* ***current and future system use*** *by* display type.

Use questions

I would have preferred not using the system.
I would use the system again if I had the option.

A one-way analysis of variances was conducted to explore the impact of display type on preference of use and whether participants would use the system again. Subjects were randomly divided into three groups (LIST, SOM, PFNET). The means and standard deviations can be found in Table 38, following.

Table 38. Descriptive Statistics on Participant Perceptions of Overall System Use & Future Use

		N	Mean	Std. Deviation	Std. Error
I would have preferred not using the system.	LIST	20	1.90	1.021	.228
	SOM	19	1.89	.937	.215
	PFNET	20	1.60	.940	.210
	Total	59	1.80	.961	.125
I would use the system again if I had the option.	LIST	19	2.79	1.084	.249
	SOM	20	2.80	.894	.200
	PFNET	20	3.55	.759	.170
	Total	59	3.05	.972	.127

We did not establish a significant difference between the groups on the statement: “I would have preferred not using the system.” The means on this response are very close as shown in Table 38 above suggesting that participants overall disagreed with this statement and preferred use of the systems regardless of display.

However when it comes to using the system in the future, there was a statistically significant difference between display types on future use [$F(2, 56) = 4.462$, $p = .016$]. The effect size, calculated using eta squared, was .13 which is considered a large effect (Cohen, 1988). Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the PFNET group (*M* = 3.55, *SD* = .759) was significantly different from both the SOM group (*M* = 2.80, *SD* = .894) and the LIST group (*M* = 2.79, *SD* = 1.084). It would appear that when it comes to future use, the Link-PFNET participants would be more likely to use the system again than both the Region-SOM and LIST group participants.

Table 39. ANOVA Results for Participant Perceptions of Overall System Use & Future Use

		Sum of Squares	df	Mean Square	F	Sig.
I would have preferred not using the system.	Between Groups	1.170	2	.585	.625	.539
	Within Groups	52.389	56	.936		
	Total	53.559	58			
I would use the system again if I had the option.	Between Groups	7.540	2	3.770	4.462	.016*
	Within Groups	47.308	56	.845		
	Total	54.847	58			

(*$p < .05$)

Table 40. Multiple Comparisons on Participant Perceptions of Overall System Use by Display Type

Dependent Variable	(I) Display Type	(J) Display Type	Mean Difference (I-J)	Std. Error	Sig.
I would use the system again if I had the option.	PFNET	LIST	.761(*)	.294	.033*
		SOM	.750(*)	.291	.033*

(*p < .05)

RQ4 Display Reactions Review

In this analysis we explored participants' overall reactions to the displays in the areas of

Search Formulation

The system helped me formulate my search.
After using the system I decided to change what I was looking for.
Using the system made coming up with search terms easier.

Sense and Understanding

The system was easy to understand. (p < .001)
The system made no sense at all. (p = .006)
I understood the medical topics better using the system.

Knowledge

I learned about the medical topic with the system.
I know more about Medical Subject Headings. (p = .068)

Visual Appeal

The system was visually appealing. (p = .042)

System Use

I would have preferred not using the system.
I would use the system again if I had the option. (p = .016)

We did not find a difference in how the displays impact perceptions of search formulation; however, on statements pertaining to understanding and making sense of the displays there was a difference between the groups with the Link-PFNET group coming out ahead of the Region-SOM and LIST groups. We also failed to establish a statistically significant difference between display typess on knowledge about the medical topics.

When it comes to visual appeal and future use of the system, participants in the PFNET group rated that display more visually appealing than the LIST group rated their display and given the option, PFNET participants appear to be more likely to use the system again in the future than both the Region-SOM and LIST groups.

RQ 1-4 Review

Overall, we found that, for the impact of display type on the dependent variable of information behavior; 1) there was a significant increase in the number of terms used to reformulate queries across all display types (LIST, SOM, PFNET); 2) there was also a significant increase in the use of partial and full MeSH terminology across all display types; and 3) there was a significant relationship between the display type and the interface level from which PFNET participants chose terms. We were unable to establish a difference between display types on 1) full MeSH terminology used in query reformulation and; 2) and change in specificity of search statements pre- to post-test.

On the dependent variable of performance perceptions, we found that 1) there was a significant increase in feelings of satisfaction, confidence, success and relevance in reformulated queries across all display types after system interaction; and 2) feelings of confidence and satisfaction on the topic of depression were reliably higher pre-test than feelings of confidence and satisfaction for cholesterol. There were no significant differences between display types on any of the performance perceptions questions.

For perceptions of knowledge, we found that; 1) participants reliably rated their pre-test knowledge on the topic of depression higher than that of cholesterol; and 2) the means suggested that post-test participants felt their cholesterol knowledge was more corrected than their depression knowledge after exploring the displays. Our analysis of overall perceptions on system usefulness found, 1) there was no difference between display types on search formulation; 2) participants reliably rated the PFNET display easier to understand and the SOM display as making no sense at all; 3) PFNET participants reliably rated that display more visually appealing; and 4) PFNET participants also would be more likely to use that display again if given the option than both the LIST and SOM participants.

CHAPTER 6: DISCUSSION OF RESULTS

RESULTS: DISCUSSION

Overall findings suggest that all displays were useful to the participants in this experiment and that the PFNET display was particularly useful for the novice searcher. Five main findings resulted from this research: 1) for all display types (LIST, SOM, PFNET) there is an increase in the number of participant search terms and in the incorporation of MeSH terminology from the visualizations following exposure to those displays; 2) there is a relationship between the display type and the interface level from which PFNET participants chose terms; 3) searchers' feelings of confidence, satisfaction, success and relevance increased across all groups after system interaction; however, pre-test feelings of confidence and satisfaction seem to be dependent upon the participant's self-reported prior knowledge of the search topic; 4) feelings of confidence and satisfaction on the topic participants reported less pre-test knowledge on (cholesterol) shifted to match post-test ratings of confidence and satisfaction on the topic they had more pre-test knowledge on (depression); and 5) participants rated the PFNET system more visually appealing, easier to understand and more likely to be used in the future if given the option.

In examining undergraduate students in the information retrieval environment for the impact of computer generated concept maps on their information behavior, perceptions of performance, and knowledge and perceptions of usefulness, two primary research questions were considered: 1) what is the impact of display type on the novice searcher's information behavior; and 2) what is the impact of different display types on the user's self-perceptions of performance, knowledge and the overall use of the system. Each of these main questions, along with related questions, will be used to organize the presentation of results and findings.

Display Interpretation

To set the landscape for the ensuing discussion, recall the PFNET uses lines to connect concepts while the SOM uses box-like regions or concept areas. The LIST group presents concepts alphabetically. See Figure 13 in the Methods Section for diagram of basic differences between displays. The primary difference between the PFNET and SOM is how each indicates term relatedness. PFNET uses lines to directly connect related concepts. Kohonen SOM uses closeness in space with the size of a region or concept area corresponding to the frequencies of occurrence of the words and the neighboring relationships of areas as an indication of frequency of co-occurrence of the concepts represented by the areas (Lin, 1992). In a PFNET, two concepts are linked if their terms share a line, while a SOM allows for multiple points of contact either by sharing a side or by being contained in the same regions (Buzydlowski, 2003).

In the PFNET display, prominent concepts are typically surrounded by other concepts directly linked to them creating a clustering effect. With the SOM, the more prominent a concept or the more it occurs with relative frequency; the more prominent, the larger the area "staked out" around it, and of names that are not directly related, "the distance between them is indicative of relatedness" (Buzydlowski, 2003). Lastly, a key difference in the display and interpretation of the relationships is the direct versus indirect associations shown by the PFNET versus the SOM. The PFNET display shows relatedness of concepts by the direct lines between map concepts which seem to lead the reader of the map from term to term while the SOM with its more open regions and implicit associations, allowing for multiple points of contact, leaves more room open for interpretation (Buzydlowski, 2003).

Constructive criticisms of visualizations have questioned some of the techniques asking whether or not the displays are readily intelligible, helpful in real time and an improvement over a simple list (White &McCain, 1997). After seeing their assigned displays, participants in this study, regardless of type of display, used more terminology and also used more partial and full MeSH terminology on the post-test. This general effect, from having the 25 most highly co-occurring terms present regardless of display format, impacted the way participants reformulated search terms. This finding is consistent with Hsieh-Yee (1993) as well as Sihvonen and Vakkari's (2004) findings that more terms were used in reformulation when participants had access to a thesaurus. The result in this research might simply have occurred because participants stored the terms they saw on the displays in working memory. Other research, involving the primacy effect found that the order of presentation of results impacts query reformulation. Terms presented first were more likely incorporated into the reformulated query than terms and information presented farther down the list of results (Allen, 1994). The implications of our results support the incorporation of system feedback for query term expansion. Interactive query formulation, where the user has control over the system-suggested terms for query expansion, improves search effectiveness (Koenemann & Belkin, 1996; Sihvonen & Vakkari, 2004). In addition, real-time query expansion with user control has also been found to increase the general usage of query expansion and improved quality of initial queries, leading to higher satisfaction (White & Marchionini 2007). And, query expansion with user control, as opposed to automatic expansion is preferred by users (Belkin, et al. 2001; Brajnik, Mizzaro, & Tasso, 1996; White & Marchionini, 2007). Those results in conjunction with the findings from this research further support the implementation of system-supported real-time query expansion.

Other research has questioned at what stage query expansion should be implemented and where concepts for query expansion should be drawn from (Efthimiadis 2000). Fairly sophisticated searchers found that having query expansion early in the search task was more useful when "the searcher needs may be most uncertain" (White & Marchionini, 2007). A small pilot study by Brajnik, Mizzaro and Taso (2002) suggests that novices more readily accept term suggestions, even when they do not have a good understanding of their meaning. This research supports query expansion with novices at the initial stage of searching; however, further investigation into the appropriate stage for implementing query expansion tools should be conducted and compared. Overall, providing real-time query expansion support at an early stage of searching for users and having additional terms present to support reformulation is more important than arrangement and format of the presentation of those terms.

The dependent variable of information behavior was explored by determining the interface level participants chose post-test terminology from. Overall, the bulk of terms incorporated post-test across all display types (LIST, SOM, PFNET) were from the primary display level. For the cholesterol query, 67% of the terms chosen by participants were from the first interface level of the display and for depression 64% from the first level. This research found a relationship between the display type and the interface level from which participants chose terminology for reformulation. A chi-square test of independence showed that the interface level the term was taken from was NOT independent of display type for both cholesterol ($p = .001$) and depression ($p = .005$). Overall, the PFNET group appeared to take more terminology from the primary interface level and less from the deeper levels of the interface. This is in contrast to the SOM group who were less likely to take their terminology from level one than the PFNET group and more likely to choose terms from deeper levels of the interface. This finding suggests that term arrangement on the display did impact search behavior.

We found with information behavior that across all groups participants used more terms after seeing the displays and those terms used post-test were more medical in nature by the use of more partial and full MeSH terminology regardless of display format. However as noted above, PFNET participants were more likely to choose their terminology from the first level of the interface while the SOM participants took more terms from deeper levels of the interface. This suggests that arrangement of terms on the PFNET display did have an impact on where they chose their terms for reformulating their query. It is our argument that PFNET participants see a more coherent structure and therefore recognize and choose more terminology from the primary interface level as opposed to the SOM display which we will argue is more disjointed and leads participants to randomly click on terms causing them to move farther away from the primary map on the query topic. The following discussion is a blend of what we can say based on the data and what we can speculate beyond that from the literature on visual perception and human processing of visual information.

The Gestalt principles of visual perception address 6 areas and attempt to describe how people organize visual elements as a unified whole: 1) figure/ground (elements are separated based on contrast); 2) similarity (similar elements are seen as a group); 3) proximity/contiguity (elements that are close together are seen as a group); 4) continuity (viewers expect elements to extend along a continuous line); 5) closure (tendency to see complete figures); and 6) area (two overlapping areas, the larger is seen as background, the smaller object a figure) (Koffka, 1935).

In relation to the visual layout of the concept maps used in this research, terms in the PFNET displays typically flow along lines of connected concepts. The principle of continuation suggests that the eye is compelled to move along one object and continue to another object. To state the same principle another way, when viewing a group of visual entities, we are more likely to create visual entities out of elements that are smooth and continuous (Ware, 2000). Secondly, because of the closeness of the terms in the PFNET display the theory of proximity also comes into play. Elements placed close together tend to be perceived as a group and these groups in turn suggest relatedness.

As we know PFNET concepts are directly connected to one another with lines. It has been argued that connectedness is a more powerful grouping principle than proximity, size, or shape, and it is a more fundamental organizing principle when it comes to visual perception (Palmer & Rock, 1994). When the mind attempts to "attend to a single dot, our attention spreads instead across the entire group in which it falls" (Driver and Baylis as quoted in Scholl, 2001). Further supporting the above discussion on the rich visual cues the PFNET presents its reader is Tullis' work on display density. Tullis found four basic characteristics which impact how well users can extract information from alpha-numeric displays: 1)overall density; 2) local density; 3) grouping; and 4) layout complexity (Tullis, 1997). Focusing on the third principle, the extent to which characters on the display form well defined perceptual groups, we find that the PFNET display supports better extraction of information than the SOM display by the placement of concepts into well-formed perceptual groups. Because of the richness of visual processing cues as addressed above, it would be reasonable to expect that for the PFNET display focus is held in a tighter area than on the SOM display, and that more focused exploration of the PFNET display along with the rich visual cues help readers infer the semantic relationships between concepts which in turn leads participants to choose terminology from that primary display.

In contrast to the PFNET, the SOM display does not cluster terms around a primary concept. SOM concepts are placed within regions and proximity of concepts denotes relatedness; moreover, there are no lines of connection which suggest continuity within the display. The regions in the SOM display provide frames of reference for the novice searcher by dividing the display into different regions and grouping related concepts within those regions; however, some displays with closed regions segment the display and make it more difficult to compare related information (Ware, 2000). Another factor complicating interpretation of the SOM display, related concepts are placed together in the regions the map creates; however, the more popular a concept, the mapping algorithm generates more space around that concept. What might appear on the display is a concept which ends up appearing farther away from a related term in the rendered display. This is in opposition to the grouping principle discussed above. With regions which might make it more difficult to compare related information and fewer direct cues for understanding the display, the SOM appears to be a more ambiguous display to the novice searcher than the PFNET display. To further extend the Gestalt discussion of visual processing this discussion will briefly step into the realm of textual processing.

Going beyond the work of Campbell (1995) who argued that Gestalt Principles also play a role in processing of text, Riley and Parker (1988) expanded Campbell's discussion of analogs between visual and textual processing. They developed meta-principles which further bridge the similarities between the two domains. Borrowing on the Gestalt principle of continuity (addressed in this research in earlier discussion), Riley and Parker (1988) tie Grice's maxim of relation in the verbal domain together to form their first meta-principle. The maxim of relation refers to readers' expectation that discourse elements are related and will interpret them as related unless there is a compelling reason not to (Riley & Parker, 1998). This first meta-principle tying Gestalt principle of continuity and Grice's maxim of relation together is called cohesion. Cohesion says that a perceiver will interpret a stimulus in the way that requires least effort in relating things.

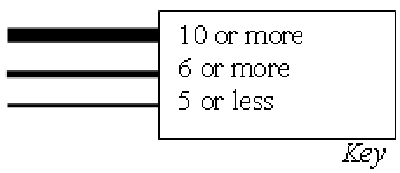
Re-connecting this meta-principle to map displays used in this experiment, we know the SOM display does not connect concepts through direct lines and therefore no immediate suggestion of visual continuity is apparent in the display. It is possible a reader will interpret the stimulus in a way that requires least effort (Riley & Parker, 1998) and in-turn follow a natural English-language reading pattern of left to right, top to bottom. The brain as a "powerful pattern-finding engine" looks for patterns to make sense of the visual information whether that pattern is meant to be there or not (Ware, 2000).

Let us diverge a moment to discuss Mooers' Law (1959). The Law of Least Effort has been discussed in various research and information seeking literature (Dervin, 1983;

Durrance, 1988; Case, 2002; Zipf, 1949). The law or principle of least effort states that, "an information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it" (Mooers, 1959/ 1996, p.1). There is a relationship to Simon's bounded rationality theory as well. Given the possibilities and the cognitive capabilities, users have to satisfice and go with a "good-enough" answer or decision (Simon, 1996). Both of these ideas also tie into optimal foraging theory discussed in the introduction of this dissertation, there are trade-offs in looking for information that are analogous to those of hunter-gatherers… a different context but a similar cost-benefit analysis (Pirolli & Card, 1998). To reconnect with our display discussion, if the SOM display is more ambiguous to the novice searcher, as we claim, and it requires more cognitive work to interpret and understand the display, then it would also stand that a novice searcher instead of putting in the effort to interpret the display format might click through the interface levels to find terms they would recognize, be satisfied with and in turn opt to use.

The Gestalt principles of visual processing and the meta-principles of visual and textual processing can be further illustrated by the display formats shown on the following pages. The top ten terms used by participants post-test and mapped onto their respective displays show some of the perceptual and textual processing differences discussed above.

Term	*Frequency*
cholesterol	13
dietary fats	8
coronary disease	8
lipids	7
triglycerides	8
liver	8
lipoproteins	8
fatty acids	5
LDL	4
HDL	4



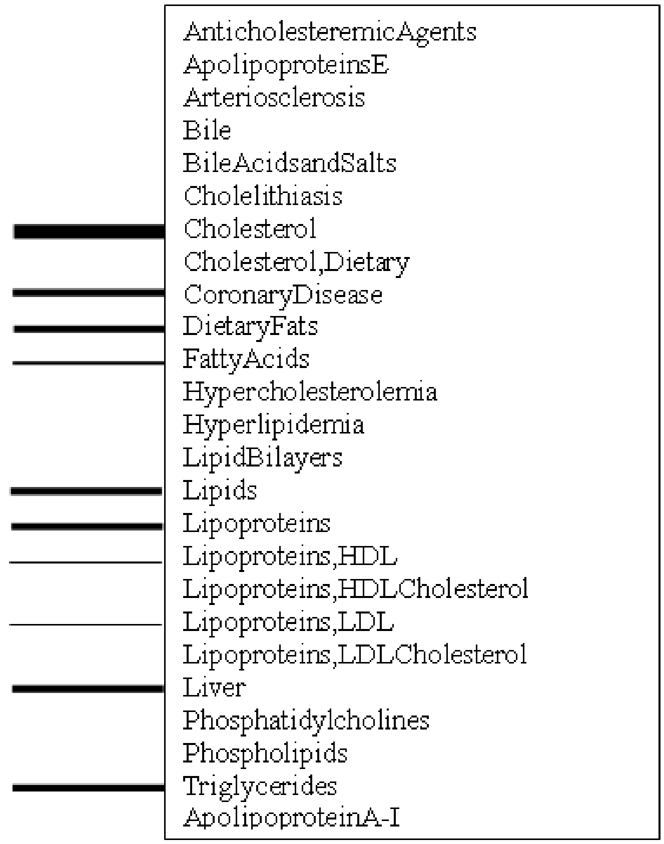
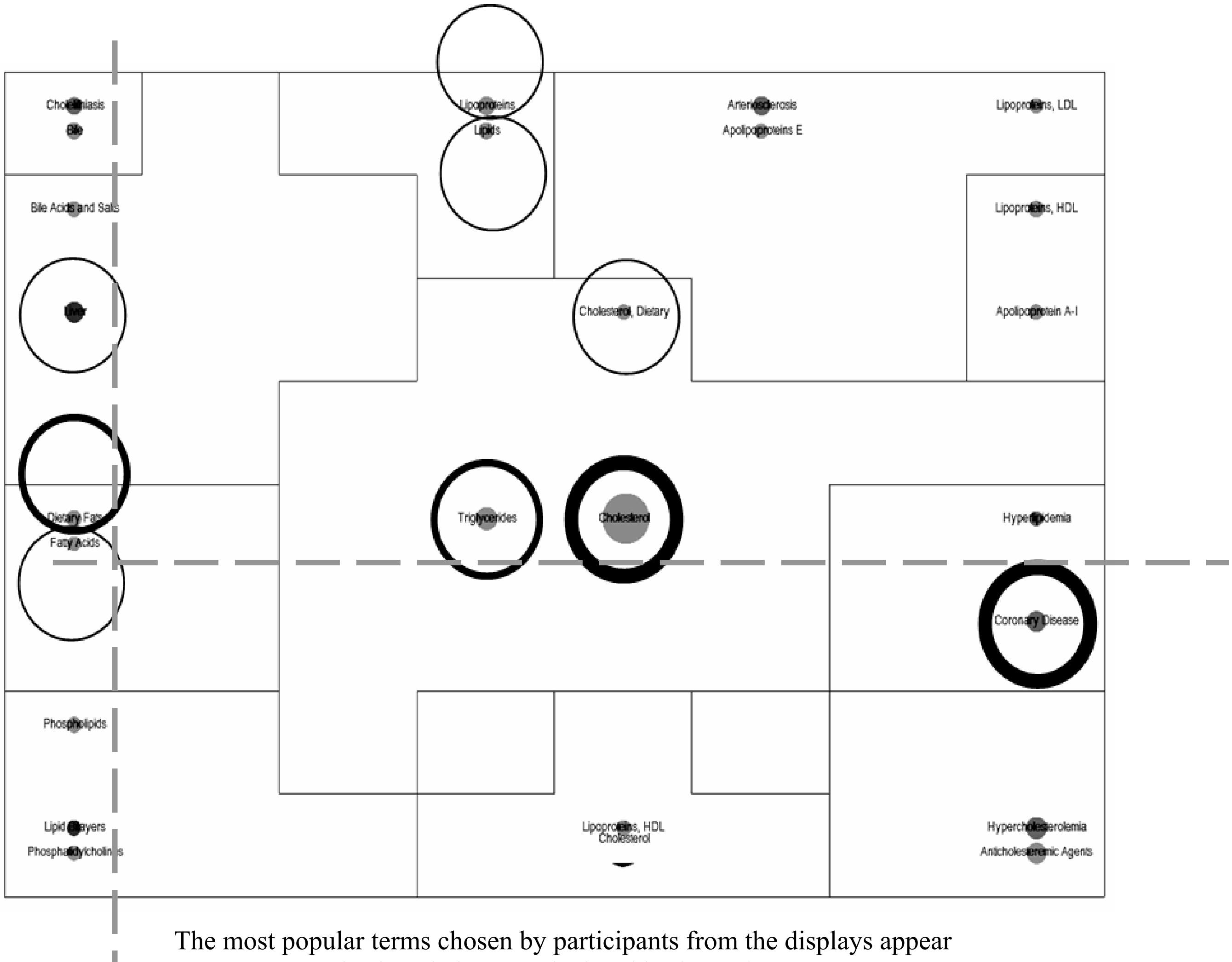


Figure 31. List Display and Top 10 Terms for Cholesterol



The most popular terms chosen by participants from the displays appear to be found along vertical and horizontal axes.

Term	*Frequency*
cholesterol	11
coronary disease	10
triglycerides	7
dietary fats	7
heart disease*	5
liver	4
lipoproteins	4
lipids	4
fatty acids	4
cholesterol, dietary	4
**Not map terminology*	

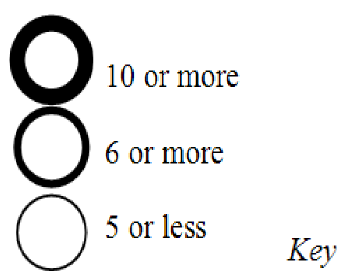
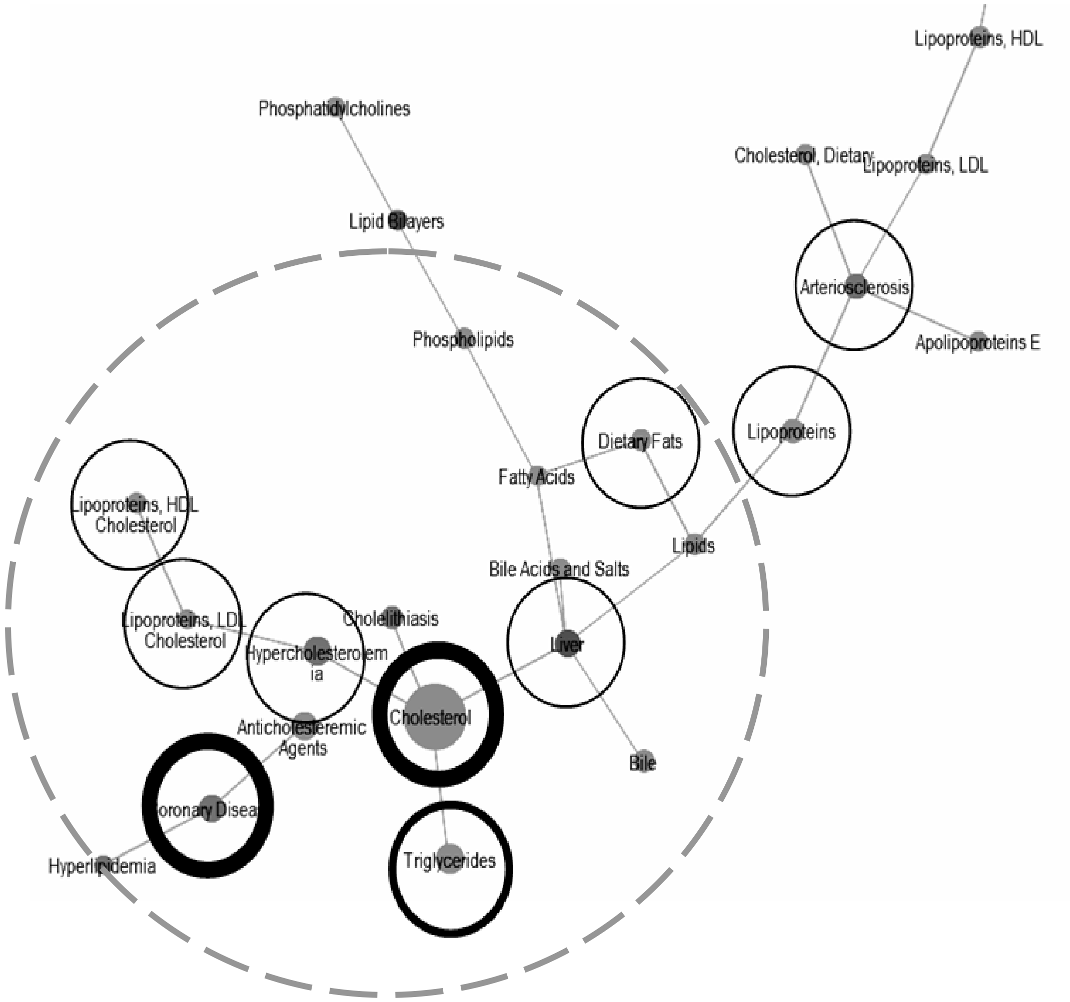


Figure 32. SOM Display and Top 10 Terms for Cholesterol



The most popular terms used by participants appear to be clustered around a central concept.

Term	*Frequency*
coronary disease	13
cholesterol	13
triglycerides	11
liver	6
lipoproteins	5
LDL	5
HDL	5
dietary fats	5
arteriosclerosis	5
hypercholesterolemia	4

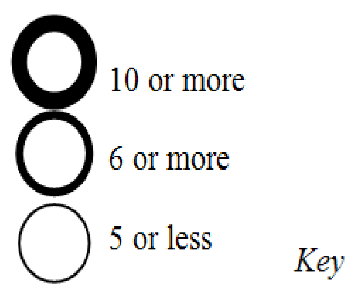


Figure 33. PFNET Display and Top 10 Terms for Cholesterol

Please see Appendix A for PFNET and SOM top ten term comparison of depression.

The differences between the maps in term arrangement, discussed within the framework of Gestalt principles of visual perception, are further illustrated by the previous display comparisons. We argue that for the novice searcher, the PFNET display is richer in visual stimuli and the visual cues of the PFNET display combined with the textual information help participants recognize term relationships more readily than SOM participants and in turn help them choose more terminology from that primary interface level than the SOM group. Recognizing information generated by a visualization is easier than recalling that information by the user (Card, Mackinlay, & Shneiderman, 1999). If a participant is able to infer meaning from the visual cues and in turn recognize or understand the relationships or the concept itself, it would make sense those participants would more readily use those terms. The average person can "recall only a fraction of the terms used to represent a concept"; however, as Bates noted those same people can "recognize a full screen of variants in an instant" (Bates, 1998). In contrast, a novice searcher using a display which has fewer visual cues, which in turn needs more interpretation, may not be able to understand those concepts or the relationships between them to the same degree. And, that person might continue through deeper levels of the display interface in order to locate terms they would be satisfied using.

Along with evaluating how participants using these displays behaved, we also asked participants questions about satisfaction with the terminology they chose. We find that both the SOM and PFNET displays groups were equally satisfied with their chosen terminology after seeing the displays; though the SOM participants drilled down deeper into the displays to choose that terminology. Ultimately, it is possible that participants using a more ambiguous display might rely on their own terms instead of using those from the system. In further support of that theory, we found that post-test SOM participants had personal terminology as part of the top ten term/term phrases. This terminology was not found on the displays.

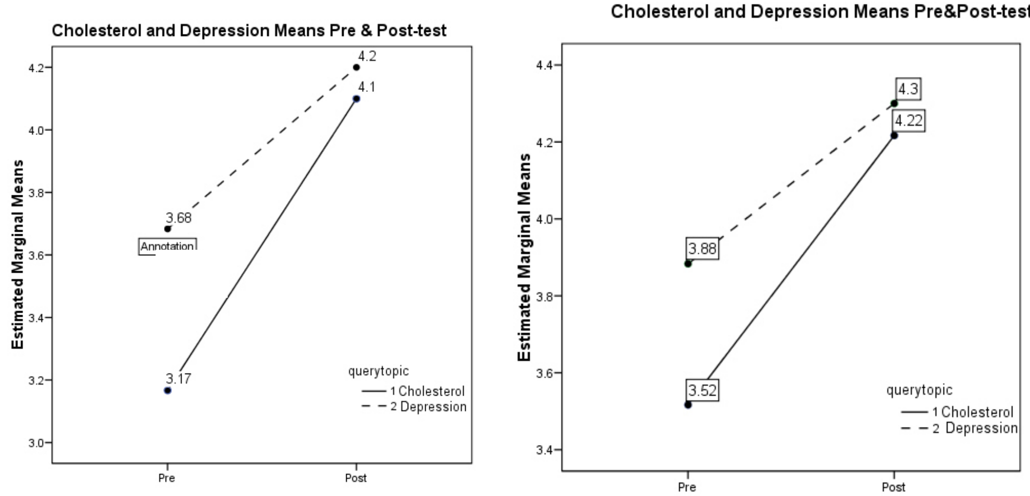
Bringing in the thread of system supported query formulation; we find that term arrangement on the display does impact novices by providing richer clues which influence them to choose terminology from a primary display level. In contrast, SOM participants chose terminology from deeper within the interface levels to find terms they were satisfied using to reformulate their query. Also in some instances participants actively chose NOT to use the formal medical terminology (MeSH) the system presented and use their own terminology. While this is not bad, it supports the notion that the terminology presented on the PFNET along with the arrangement supported by visual processing cues contains more information for the novice to recognize and then employ in their selection of terms. From informal observation during the experiment, it was frequently noted that participants sighed quite heavily, made noises which suggested frustration and upon being shown the "other" display format (PFNET) at the end of the experiment made comments like, "cool," "wow" and "neat". In the section on the final questionnaire asking for general comments about the displays, SOM participants used terms like: "I was confused," "difficult to follow," "confusing at times," and "really confusing".

Further weaving this thread back into the idea of query reformulation this would support the use of visual cues in the presentation of system supported query term expansion. It also would support a hybrid use of graphical cues in query expansion techniques. White and Marchionini's (2007) real time query expansion tool showed a simple list of terms that popped up as participants typed in their search information. They also found that timing of the query support likely increased uptake and that could be related to the alternative presentation technique (White & Marchionini, 2007). Instead of only having text, or as in this experiment having a full-screen with a large involved visualization, perhaps a hybrid expansion tool which shows the terms and can also show the connections between the system-suggested terms for reformulation would be viable. "People could manage more powerful searches quickly if an initial submitted term or topic yielded a screen full of term possibilities, related subjects, or classifications for them to choose from (Bates, 1998, p. 1202; Bates, 1986).

Drawing back on the data from our experiment, another primary finding was on self-reported feelings of confidence, satisfaction, success and relevance. After seeing their assigned displays participants, regardless of type of display, reported increased feelings of confidence, satisfaction, success and relevance regarding the terminology they chose. Again, as noted in other research, this finding supports system generated query term expansion tools to help with reformulation (Efthimiadis, 1996; Greenberg, 2001; White & Marchionini, 2007). In this experiment we found that the system generated query expansion not only supported a move toward better queries (more terms and more full and partial MeSH terminology) but it also helped users feel more confident and satisfied with their search terminology, and increased participants' feelings of success and relevance. Though we did not find a difference in how the arrangement of terms on the different displays might differentially impact the novice searcher with these perceptions, we did find that participant feelings of confidence and satisfaction differed significantly dependant upon the query topic they were searching on.

Participants had more feelings of confidence and satisfaction pre-test for the search topic of depression than for the search topic of cholesterol. Our theory was that prior knowledge on the depression topic fed into subjects' feelings of confidence and success. This is not surprising and was confirmed by participant self-perceptions of knowledge on depression which was reliably different than their knowledge of cholesterol. However the key component we would like to highlight with larger implications is that post-test, the mean score for feelings of satisfaction and the mean score for feelings of confidence on the topic of cholesterol (which were significantly LOWER pre-test) shifted to match the scores for satisfaction and confidence on the topic of depression. Not only did participants across all groups feel more confident and more satisfied with the terms they chose after seeing the displays for the topic they were familiar with and one they were not, but participant confidence and satisfaction scores post-test matched those of the topic they initially felt they had more knowledge on. See Figure 34, following.

Satisfaction **Confidence**



(Satisfaction means post-test were Cholesterol M = 4.1; Depression M = 4.2 and for the topic of confidence were Cholesterol M = 4.2; Depression M = 4.2.)
Figure 34. Pre- & Post-Test Means on Self-Perceptions of Satisfaction & Confidence

In addition, the analysis on correction of knowledge post-search was suggestive that knowledge on cholesterol was corrected more after looking through the displays than knowledge of depression though this was not a significant difference.

Let us shift in our discussion from system supported query expansion and query reformulation to briefly address domain novice and the domain expert. In a very basic manner it would be a difference in their knowledge. From the literature we know that domain experts have specific skills that allow them to assess, process, and understand a problem differently than that of a novice (Bransford, Brown, & Cocking, 1999). An expert notices, organizes, processes and interprets information in their environment differently than a novice and they use a variety of resources to find search term alternatives (Fidel, 1991a, 1991b, 1991c). This in turn helps them to be more successful when searching for information. We might further expand upon this to also encompass the affective feelings of the novice versus the experts during the search process. Experts may not experience uncertainty in the same manner that a domain novice would. The domain novice holds only a few constructs while the expert is “rarely at the true beginning” (Kuhlthau, 1993). Research on the affective aspects of the information seeking process have framed it as a process of reducing uncertainty and making sense(Dervin, 1977; Kuhlthau, 1993; Marchionini, 1989), and others have identified sharp increases in uncertainty and decreases in confidence after searches with novices were initiated (Kuhlthau, 1993). Psychologist Kelly’s phase of construction has confusion and doubt as part of new experience also folding into Kuhlthau’s findings of decreases in confidence after search initiation.

The results of this research with novices found that after seeing the displays, and interacting with the system, participants not only felt more satisfaction in their chosen terms, their confidence levels also increased. And, where there was a significant difference pre-test in feelings of satisfaction and confidence for the depression topic, post-test confidence and satisfaction scores for cholesterol, a topic they felt they had less knowledge on and less confidence and satisfaction in their choice of terms, shifted to match that of their scores for depression. This alone has wide implications on many levels. A process which is known to cause uncertainty and confusion in the novice, for our particular experiment, we found that feelings of satisfaction, confidence, relevance and success increased. Not only did all participants across all groups become more confident, satisfied with their terms, and feel they would be more successful and relevant post-display interaction, their feelings of confidence and satisfaction matched the confidence and satisfaction scores of a topic they reported having greater knowledge on.

Further investigation into the influence of visual display formats supporting query term expansion on affective feelings during the constructive information seeking process is warranted by these results.

Lastly, in the information behavior and perceptions data we mentioned trends in the means of our dependent variables that was potentially meaningful. These trends point to the PFNET group as the forerunner of the three displays used in this experiment. To review, the PFNET participants on average: 1) used more full MeSH terms on the post-test; 2) had more full and partial MeSH than the SOM or LIST groups; 3) participants rated their confidence, satisfaction, success and relevance with chosen terms on the post-test highest; 4) the PFNET group participants felt that their post-searching knowledge corrected their prior knowledge more than the other display types; and 5) the means for the PFNET group were higher on search formulation help and support for coming up with search terms.

None of these results were significant and taking these results individually would not warrant further discussion. However, from an aggregate perspective they are meaningful and appear to support the idea that the PFNET and its display format influences the novice searcher differently than the other displays. The rich visual cues of the PFNET display combined with the textual information help participants understand the visual relationships more readily, and in turn participants choose terminology from that primary display. If we think of the PFNET display as a well-designed display for the novice searcher, it might then be claimed that novices seem to be able to "obtain information far more rapidly and accurately from external sources than from their own memories" (Pirolli & Card, 1994; Ware, 2000).

Additionally we know from our perceptions of use data (RQ 4) that SOM participants reliably rated their display the lowest of the three on the question, “the system was easy to understand”. These same participants also reliably rated that the SOM display, “made no sense at all” in comparison to the PFNET group. This self-reported lack of understanding and sense making of the concept arrangement for the SOM display further supports our theory that the visual stimuli for the SOM participants is less rich in perceptual cues, there is less understanding inferred and those participants then search through deeper levels of the display interface to find terms they would choose. Even the LIST group participants reliably rated their display interface easy to understand, more than the SOM participants.

PFNET for the Novice Searcher

We will argue in the following section that the PFNET display is better suited for the novice searcher. Our previous discussions highlighted how the PFNET and the SOM, by arrangement of terms in the visual display, might be interpreted and impact the novice searcher. To contrast our discussion of the novices in this experiment and their reactions and perceptions of the PFNET and the SOM first we will mention some of the basic differences between novices and experts as well as review data from a previous a study which explored these map displays with experts in the humanities. This information will then be used to contrast to the findings of this study.

Experts notice meaningful patterns of information that are not noticed by novices. Experts also have acquired a great deal of content knowledge and that knowledge is organized in ways that reflect a deep understanding of their subject matter. Domain experts have specific skills that allow them to assess, process, and understand a problem differently than that of a novice (Bransford, Brown, & Cocking, 1999). An expert, with an understanding of the domain would already possess the tools to interpret the information as presented and to draw connections based on that knowledge. However, novices lack the understanding and an already constructed mental model of the domain.

Previous research using the same display formats for arrangement of terms (SOM and PFNET) explored how well these displays corresponded to a set of mental maps of experts and whether a particular map was preferred by those experts (Buzydlowski, 2003). These domain experts in the Humanities expressed that the PFNET display acted as a lead; "with the PFNET I found I followed the lines out and out and out" (Buzydlowski, 2003). The same experts also expressed an affinity toward the SOM because it allowed more flexibility for the reader to make their own connections, there were more "open possibilities," and multiple points of contact. This was in contrast to the PFNET, "if you are not clear on the connections it [the PFNET] gives you one" (Buzydlowski, 2003).

It appears that domain experts felt the PFNET led the reader while the SOM allowed for more interpretation. For novices, the SOM appears to rely more on the reader to form their own connections, and novices who use the SOM display, clearly do not have a well-organized understanding of a specific domain and are therefore less able to extract the necessary understanding from their mental models to interpret what they are seeing. SOM participants in this study did comment about the display format and understanding the visual layout: "didn't understand what the white lines represent"; while another noted "[the SOM] does not present information in a way that is concise or help the user understand the relationships between headings"; "the way the topics are grouped is not always clear"; and "I was confused as to how the boxes were divided and why".

In comparison, a participant in the PFNET group for this experiment wrote, "good way to see how relevant the topics are to your main search idea," and another, "it was clear that items/categories of the map were related". It would be a reasonable conclusion that being led and having explicit connections between the terms supports the novice's exploration and information seeking better than a display which leaves more open for interpretation and is perceived as ambiguous by the novice. Novices in our research found the SOM display confusing, and difficult to understand while they rated the PFNET easier to understand, more visually appealing and were more likely to use this display in the future. It is not a difficult leap to say that novices prefer the PFNET display because they understand it better.

Conclusions and Areas for Future Research

In examining the impact of computer generated concept maps on the information behavior, perceptions of performance, knowledge and perceptions of usefulness on the undergraduate student in the information retrieval environment, five main findings resulted from this research: 1)an increase in the number of participant search terms and the use of full and partial MeSH terms occur across all display types (LIST, SOM, PFNET); 2) there is a relationship between the display type and the interface level from which PFNET participants chose terms; 3) searchers' feelings of confidence, satisfaction, success and relevance increase across all groups after system interaction; however, feelings of confidence and satisfaction seem to originate more strongly dependent upon prior knowledge; and 4) while participants' overall information behavior did not seem to be heavily impacted by display type, participants rated the PFNET system more visually appealing, easier to understand and were more likely to use the system again in the future if given the option. Overall, findings suggest that all displays were useful to the participants and that the PFNET display is particularly useful for the novice searcher.

This research focused on the novice searcher in the information retrieval environment using three different arrangements of visualizing information. In order to ascertain the impact of the displays on the novice searcher, participants in this research study were limited to exploring the display and concepts while not allowed to actually retrieve documents. Also, the information task or search query was kept broad in nature in order to allow participants to fully explore the different displays. These factors limited the type of data gathered and also produced some impact on true topic exploration. One suggestion for future iterations of this research would include observing novice searchers using the PFNET and SOM display formats with real time medical information queries. This would allow for more topical exploration and formulation and refinement of topics. Along a similar line, further comparing different levels of topical knowledge might help us understand how that knowledge impacts the exploration of these displays. It would also be helpful to study participants from the broader perspective of academic work. Instead of just looking at one slice of their information behavior, it would be useful to explore outcomes on papers and assignments. Comparing grades and other end-products produced by students using the concept maps and a regular system would also be one future manner of measuring the impact of these display formats on novices.

Also, moving beyond basic information behavior analysis to focus more specifically on data collection measurements which evaluate learning, understanding and meaningful retrieval might give us a better look into the impact of these different displays on the cognitive behavior of the novice searcher. In addition, asking novices to look at displays and interpret the different formats for concept relationships might give us a sharper image of how the visual cues on the displays provide are interpreted. Another manner of analyzing the data would be to look at the full, formal and partial MeSH terms used by participants and map each of those terms to their respective display. It then might be possible, to quantitatively explore participant terms to look for visual relationship on the displays. For example, exploring if the full and partial MeSH terms used post-test are connected directly by a line with the PFNET or within the same area on the SOM. This might help us to further understand how participants chose terms to use for their query reformulations.

Conclusion

It is possible to better acquaint the searcher with the information landscape they are traveling. By providing clues to the information landscape, novices use more terminology and use more system-appropriate terminology through supported query reformulation. This suggests that just having terms generated by the system early on to support query reformulation is more important than the arrangement of those terms.

In the retrieval environment where it is known that feelings of confidence and satisfaction typically decrease at the initial stages of the process, incorporating visualizations into the retrieval context had the opposite effect in our research.

Participants felt more confident, more satisfied and more successful with their terminology and the overall search process. The affective aspect of design has been predominantly neglected in research, particularly when it comes to electronic information environments. Understanding how the affective impacts information behavior and engagement with an information retrieval system in its broader environment is important. The only differing aspect of the display formats used in this experiment were arrangement of terms; however, participants in one particular group rated that display higher overall. It was considered to be more visually appealing, more understandable, and more likely to be used in the future. This suggests system design had an impact on the affective as it related to future system use. More research to understand what constitutes system engagement and how the affective engagement and future should be studied.

Overall the results of this research warrant a caution against implementing information retrieval systems with visualization interfaces before the full impact of these displays can be understood. The basic graphical rule can be simplified with the following statement: if it doesn't add anything, don't use it.

With access to information increasing exponentially, traversing the ins and outs of the digital library and electronic access to information is more circuitous and potentially frustrating for any searcher, particularly the novice. Despite all of the technological and changing computer capabilities, our standard retrieval systems still hide the landscape from the searcher. The system perspective of design has focused on increasing the level of what the system can do automatically for the searcher. This model engages the searcher less and less and smartens the system.

A system should connect with and engage the user more, to help them understand and learn within the context of the information they need. Modern systems should use what we know from research and instead of taking on more of the task of the searcher, the system should work to engage and invoke participation from the user. By engaging the user more in the search process, it increases the potential for meaningful engagement. The model of the passive user; inputting a few terms, looking through a few results, changing a few terms, looking through a few more results, has not shown over time to be all that effective, no matter how much computing capabilities have increased. How do we engage the mind of the searcher through the interface of the system? How do we use technology to elicit active participation and engage Jesse, our novice searcher, not just in the task of finding information, but in a meaningful, affective, cognitive and reflective manner (Norman, 2004)? Just because we have more access to information, it does not mean we have good access.

Bibliography

Agosto, D. E. (2001). Bounded rationality and satisficing in young people's web-based decision-making. Journal of the American Society for Information Science, 53(1), 16-27.

Allen, B. (1994). Cognitive abilities and information system usability. Information Processing and Management, 30(2), 177-191.

Allen, B. (1998). Information space representation in interactive systems: Relationship to spatial abilities. Paper presented at the Digital Libraries, Pittsburgh, PA.

Austin, B. (2001). Mooers' Law: In and out of context. Jounal of the American Society for Information Science, 52(8), 2.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern Information Retrieval. New York: ACM Press, Addison-Wesley.

Barnett, S., & Ceci, S. (2002). When and where do we apply what we learn? A Taxonomy for far transfer. Psychological Bulletin, 128(4), 612-637.

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

Baruchson-Arbib, S., & Bronstein, J. (2002). A view to the future of the library and information science profession: A Delphi study. Journal of the American Society for Information Science and Technology, 53(5), 11.

Bates, M. J. (1989). The Design of browsing and berrypicking techniques for the online search interface Online Review, 13, 407-424.

Bates, M. J. (1986). An Exploratory paradigm for online information retrieval. Paper presented at the Proceedings of the Sixth International Research Forum in Information Science (IRFIS 6), Frascati Itals.

Bates, M. J. (1977). Factors affecting subject catalog search success. Journal of the American Society for Information Science, 28(3), 161.

Bates, M. J. (1979b). Idea tactics. Journal of the American Society for Information Science, 30, 280-289.

Bates, M. J. (1998). Indexing and access for digital libraries and the internet: Human, database, and domain factors. Journal of the American Society for Information Science, 49(13), 20.

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

Beaulieu, M. (1997). Experiments with interfaces to support query expansion. Journal of Documentation, 53(1), 8-19.

Belkin, N. J. (1988). On the nature and fuction of explanation in intelligent information retrieval. Paper presented at the SIGIR '88 Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval.

Belkin, N. J. (1991). A methodology for taking account of user tasks, goals and behavior for design of computerized library catalogs. 23(1), 61-65.

Belkin, N. J., Cool, C., Kelly, D., Lin, S.-J., Park, S. Y., Perez-Caballo, J., et al. (2001). Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval. Information Processing and Management, 37, 403-434.

Belkin, N. J., & Croft, W. B. (1987). Retrieval techniques. In M. E. Williams (Ed.), Annual review of information science and technology, (Vol. 22, pp. 109-145): Elsevier Science Publishers

Belkin, N. J., & Marchetti, P. G. (1990a, 1990). Determining the functionality features of an intelligent interface to an information retrieval system: SIGIR '90: Proceedings of the 13th annual international ACM SIGIR conference on Research and development in information retrieval. p. 151-177

Belkin, N. J., Michell, B. G., & Kuehner, D. G. (1980). Representation of texts for information retrieval. Paper presented at the Proceedings of the 18th annual meeting on Association for Computational Linguistics.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982a). Ask for information retrieval: Part I background and theory. Journal of Documentation, 38(2), 61-71.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982b). Ask for information retrieval: Part II results of a design study. Journal of Documentation, 38(3), 145-164.

Belkin, N. J., & Vickery, A. (1985). Interaction in information systems: A review of research from document retrieval to knowledge-based systems. London: British Library.

Blandy, S. G., & Libutti, P. A. (1995). As the cursor blinks: Electronic scholarship and undergraduates in the library. Library Trends, 44(2), 279-305.

Borgman, C., Hirsh, S., Walter, V., & Gallagher, A. (1995). Children's searching behavior on browsing and keyword online catalogs: The science library catalog project. Journal of the American Society for Information Science, 46(9), 663-684.

Borgman, C. L., Belkin, N. J., Croft, W. B., Lesk, M. E., & Landauer, T. K. (1988). Retrieval systems for the information seeker: Can the role of the intermediary be automated? Paper presented at the CHI '88: Proceedings of the SIGCHI conference on Human factors in computing systems., 51-54.

Börlünd, P., & Ingwersen, P. (1997). The Development of a method for the evaluation of interactive information retrieval systems. Journal of Documentation, 53(3), 225-250.

Brajnik, G. (1999). Information seeking as explorative learning. Multimedia Information Retrieval Applications.

Brajnik, G., Mizzaro, S., & Tasso, C. (1996). Evaluating User interfaces to information retrieval systems: A Case study on user support. SIGIR 96, 128-136.

Bransford, J., Brown, A., & Cocking, R. (Eds.). (1999). How people learn: Brain, mind, experience, and school. Washington, D.C.: National Academy Press.

Buzydlowski, J. W. (2003). Comparison of self-organizing maps and pathfinder networks for the mapping of co-cited authors. Unpublished Dissertation, Drexel University, Philadelphia, PA.

Buzydlowski, J. W., White, H. D., & Lin, X. (2002, 2002). Term co-occurrence analysis as an interface for digital libraries: Visual interfaces to digital libraries. Paper presented at the [JCDL 2002 Workshop].

Campbell, K. S. (1995). Coherence, continuity and cohesion: Theoretical foundations for document design. Hillsdale, NJ: Erlbaum.

Card, S., Mackinlay, J., & Shneiderman, B. (Eds.). (1999). Readings in information visualization - Using vision to think: Morgan-Kaufmann.

Card, S. K., Robertson, G. G., & Mackinlay, J. D. (1991). The Information visualizer: An Information workspace. Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology. ACM, 181-186.

Carlson, S. (2001, November 16). As Students work online, reading rooms empty out-leading some campuses to add Starbucks. Chronicle of Higher Education. A35

Case, D. (2002). Looking for information. San Diego, CA: Academic Press.

Chang, K. E., Sung, T. T., & Chen, I.-D. (2002). The Effect of concept mapping to enhance text comprehension and summarization. The Journal of Experimental Education, 71(1), 5-23.

Chen, C. (1998). Bridging the gap: The use of pathfinder networks in visual navigation. Journal of Visual Languages and Computing, 9, 267-286.

Chen, C. (2000). Empirical evaluation of information visualizations: An introduction. International Journal Human-Computer Studies, 53, 631-635.

Chen, C., & Czerwinski, M. (2000). Empirical evaluation of information visualizations: An introduction. International Journal of Human-Computer Studies, 53(5), 631-635.

Chen, C., & Dhar, V. (1991). Cognitive process as a basis for intelligent retrieval system design. Information Processing and Management, 27(5), 405-432.

Chen, C., & Yu, Y. (2000). Empirical studies of information visualization: A Meta-Analysis. International Journal Human-Computer Studies, 53, 851-866.

Chen, H., Houston, A. L., Sewell, R. R., & Schatz, B. R. (1999). Internet browsing and searching: User evaluations of category map and concept space techniques. Journal of the American Society for Information Science, 49(7), 582-603.

Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. (2001, 2001). Using information scent to model user information needs and actions and the Web: CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems. 490-497.

Chi, E. H., Pirolli, P., & Pitkow, J. (2000, 2000). The scent of a site: a system for analyzing and predicting information scent, usage, and usability of a Web site. CHI '00, Proceedings of the SIGCHI conference on Human factors in computing systems. 161-168.

Chi, M. (1978). Self-Explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), Advances in Instructional Psychology (Vol. 5, 161). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-Explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145-182.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Cleveland, W. S. (1993). Visualizing data. Summit, New Jersey: Hobart Press.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences (2nd ed.). New York: Academy Press.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. International Journal-Human Computer Studies, 41, 801-849.

Dervin, B. (1977). Useful theory for librarianship: Communication, not information. Drexel Library Quarterly, 13, 16-32.

Dervin, B. (1992). From the mind's eye of the user: Sense-making qualitative-quantitative methodology. In J. D. Glazier & R. R. Powell (Eds.), Qualitative research in information management (pp. 61-84). Englewood, CO: Libraries Unlimited.

Dervin, B. (1998). Sense-making theory and practice: An overview of user interests in knowledge seeking and use. Journal of Knowledge Management, 2(2), 36-46.

Dervin, B., & Nilan, M. (1986). Information needs and uses. In M. E. Williams (Ed.), Annual Review of Information Science and Technology (ARIST) (Vol. 21): Knowledge Industry Publications, Inc.

Doyle, L. B. (1961). Semantic road maps for literature searchers. Journal of the Association for Computing Machinery, 8, 553-578.

Durrance, J. C. (1988). Information needs: Old song, new tune. In rethinking the library in the information age: a summary of issues in library research (pp. 159-177): Office of Educations Research and Improvement, Department of Education.

Efthimiadis, E. N. (1994, May 10-12). A study of end-user behavior in searching CD-ROM bibliographic databases. Paper presented at the Proceedings of the 15th National Online Meeting, New York.

Efthimis N. Efthimiadis. (1996) Query Expansion. In: Martha E. Williams (ed.), *Annual Review of Information Systems and Technology (ARIST)*, v31, pp 121-187, 1996.

Efthimiadis, E. N. (2000). Interactive query expansion: a User-based evaluation in a relevance feedback environment. Journal of the American Society for Information Science and Technology, 51(11), 989-1003.

Ellis, D. (1989). A Behavioural approach to information retrieval system design. Journal of Documentation, 45(3), 171-212.

Ellis, D. (1992). The Physical and cognitive paradigms in information retrieval research. Journal of Documentation, 48(1), 45-64.

Fenichel, C. H. (1979). Online information retrieval: Identification of measures that discriminate among users with different levels and types of experiences. Drexel University, Philadelphia.

Fidel, R. (1984). Online searching styles: A case-study-based model of searching behavior. Journal of the American Society for Information Science, 35(4), 211.

Fidel, R., & Soergel, D. (1983). Factors affecting online bibliographic retrieval: A Conceptual framework for research. Journal of the American Society for Information Science, 34(3), 163.

Fowler, R. H., & Dearholt, D. W. (1990). Information retrieval using pathfinder networks. In R. W. Schvaneveldt (Ed.), Pathfinder Associative Networks: Studies in Knowledge Organization (pp. 165). Norwood, NJ: Ablex Publishing Corporation.

Fowler, R. H., Fowler, W. A. L., & Wilson, B. A. (1991). Integrating query, thesaurus, and documents through a common visual representation. Paper presented at the Proceedings of the Fourteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 143-151, October 13-16, Chicago, Illinois.

Franklin, B., & Phum, T. (2004). Library usage patterns in the electronic information environment. Information Research, 9(4), Paper 187.

Froehlich, T. J. (1994). Relevance reconsidered- Towards an agenda for the 21st century: Introduction to special topic issue on relevance research. Journal of the American Society for Information Science, 45(124-134).

Graham, M., Kennedy, J., & Benyon, D. (2000). Towards a methodology for developing visualizations. International Journal Human-Computer Studies, 53, 789-807.

Greenberg, J. (2001a). Automatic query expansion via lexical-semantic relationships. Journal of the American Society for Information Science, 52(5), 402-415.

Greenberg, J. (2001b). Optimal query expansion (QE) processing methods with semantically encoded structured thesauri terminology. Jounal of the American Society for Information Science, 52(6), 487-498.

Hascoet-Zizi, M., Ahlberg, C., Korfhage, R., Plaisant, C., Chalmers, M., Rao, R. (1996, November 6-8). Where is information visualization technology going? Paper presented at the Proc. of User Interface Software and Technology (UIST) '96, Seattle, Washington, USA.

Haskell, E. H. (2001). Transfer of learning: Cognition, instruction and reasoning. New York: Academic Press.

Hearst, M., & Karado, C. (1997). Cat-a-Cone: An interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Philadelphia, PA, 246-255.

Hearst, M. A. (1995). TileBars: Visualization of term distribution information in full text information access. Proceedings of the SIGCHI conference on Human factors in computing systems, Denver, Colorado, United States, 59-66.

Hearst, M. A. (2003). Information visualization: Principles, promise, and pragmatics. Paper presented at the CHI 2003 Tutorial Fort Lauderdale, Florida.

Hewins, E. T. (1990). Information need and use studies. In M. E. Williams (Ed.), Annual Review of Information Science and Technology (ARIST) (Vol. 25, pp. 145–172): Elsevier Science Publishers.

Hirsh, S. G. (1996). The Effect of domain knowledge on elementary school children's information retrieval behavior on an automated library catalog. Unpublished doctoral dissertation, University of California, Los Angeles.

Houston, A. L., Chen, H., Schatz, B. R., Hubbard, S. M., Sewell, R. R., & Ng, R. D. (2000). Exploring the use of concept spaces to improve medical information retrieval. Decision Support Systems, 30, 15.

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. Journal of the American Society for Information Science, 44(3), 161-174.

Ingwersen, P. (1982). Search procedures in the library: Analyzed from the cognitive point of view. Journal of Documentation, 38(3), 165-191.

Jansen, B. J., Spink, A., Bateman, J. and Saracevic, T. (1998) Searchers, the subjects they search, and sufficiency: a study of a large sample of Excite searches. Proceedings of WebNet 98 Conference, Orlando, FL, November 1999.

Koenemann, J., & Belkin, N. J. (1996). A case for interaction: a study of interactive information retrieval behavior and effectiveness. Paper presented at the Conference on Human Factors in Computing Systems, Proceedings of the SIGCHI conference, Vancouver, British Columbia, Canada.

Koffka, K. (1935). Principles of Gestalt Psychology. New York: Harcourt-Brace.

Korfhage, R. (1991). To see, or not to see- is that the query? In: Proceedings of the Fourteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval 1991. pp. 134-141. Chicago, Illinois.

Koshman, S. (2006). Visualization-based information retrieval on the Web. Library & Information Science Research, 28, 192-207.

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

Kuhlthau, C. (2004). Seeking meaning: A process approach to library and information sciences (2nd ed.): Libraries Unlimited.

Lin, X. (1992). Visualization for the document space. Proceedings of the 3rd conference on Visualization '92, October 19-23, 1992, Boston, Massachusetts.

Lin, X. (1993). Self-organizing semantic maps as graphical interfaces for information retrieval. Unpublished doctoral dissertation. University of Maryland at College Park, College Park, MD, 1993

Lin, X. (1996). Graphical table of contents: DL '96: Proceedings of the first ACM international conference on Digital libraries.

Lin, X., Soergel, D., & Marchionini, G. (1991, 1991). A self-organizing semantic map for information retrieval: SIGIR '91: Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval.

Lin, X., White, H. D., & Buzydlowski, J. (2003). Real-time author co-citation mapping for online searching. Information Processing & Management, 39(5), 689-706.

Lin, X., White, H. D., Buzydlowski, J. 2001b. Associative Searching and Visualization. The Book of Abstracts, SSGRR 2001, International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet. August 6-12, L'Aquila, Italy. 16 (full text on cd-rom).

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. Journal of the American Society for Information Science, 54-66.

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

McCagg, E. C., & Dansereau, D. F. (1991). A convergent paradigm for examining knowledge mapping as a learning strategy. Journal of Educational Research, 84(6), 317-324.

McCain, K. W. (1986). Co-cited author mapping as a valid representation of intellectual structure. Journal of the American Society for Information Science, 37(3), 111-122.

McCain, K. W. (1990). Mapping authors in intellectual space: a technical overview. Journal of the American Society for Information Science, 41(6), 433-443.

McCain, K. W., White, H. D., & Griffith, B. C. (1987). Comparing retrieval performance in online data bases. 23(6), 539-553.

McDonald, S., & Stevenson, R. J. (1999). Spatial versus conceptual maps as learning tools in hypertext. Journal of Educational Multimedia and Hypermedia, 8(1), 43-64.

McGaghie, W., Crimmon, D. R., Mitchell, G., & Thompson, J. A. (2004). Concept mapping in pulmonary physiology using pathfinder scaling. Advances in Heath Sciences Education, 9, 225-240.

Mooers, C. (1960; 1997). Mooers' Law; Or why some retrieval systems are used and others are not, reprinted in Bulletin of the American Society for Information Science Oct/Nov 1996; (23)1.

Morse, E., Lewis, M., & Olsen, K. A. (2000). Evaluating visualizations: using a taxonomic guide. International Journal Human-Computer Studies, 53, 637-662.

Norman, D. (1993). Things that make us smart. Reading, MA: Addison-Wesley.

Norman, D. (2004). Emotional design: Why we love or hate everyday things. New York: Basic Books.

Novak, J. (1991). Clarify with concept maps. The Science Teacher, 58(7), 45-49.

Novak, J. (1993). How do we learn our lesson. The Science Teacher, 60(3).

Novak, J. (1998). Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. Science Education, 86(4), 548-571.

Novak, J. D., & Canas, A. J. (2006). The theory underlying concept maps and how to construct them. Florida Institute for Human and Machine Cognition.

Pirolli, P. (1997). Computational models of information scent-following in a very large browsable text collection. Proc. Conference on Human Factors in Computing Systems, CHI'97, Atlanta, March 1997, ACM Press, pp.3-10

Pirolli, P., & Card, S. K. (1998). Information foraging models of browsers for very large document spaces. Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '98); 1998 May 24-27; Aquila; Italy. NY: ACM; 1998; 83-93.

Pirolli, P., Card, S. K., & Wege, M. M. V. D. (2000). The effect of information scent on searching information: visualizations of large tree structures. Proceedings of Advanced Visual Interfaces Conference (AVI 2000); 2000 May 23-27; Palermo, Italy. NY: ACM; 2000; 161-172.

Pirolli, P., Card, S. K., & Wege, M. M. V. D. (2001). Visual information foraging in a focus + context visualization, in *Proceedings of CHI '01* (Seattle, WA, April 2001), ACM Press, 506-513.

Pirolli, P., Card, S. K., & Wege, M. M. V. D. (2003). The effects of information scent on visual search in the hyperbolic tree browser. ACM Transactions on Computer Human Interaction. 2003 March; 10 (1): 20-53.

Prabha, C., Connaway, L. S., Olszewski, L., & Jenkins, L. (2007). What is Enough? Satisficing information needs. Journal of Documentation, 63(1), 74-89.

Riley, K., & Parker, F. (1998). Parallels between visual and textual processing. IEEE Transactions on Professional Communication, 41(3), 175-185.

Robertson, S. E., & Hancock-Beaulieu, M. M. (1992). On the evaluation of IR systems. Information Processing and Management, 28(4), 457-466.

Saracevic, T. (1975). RELEVANCE: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, November-December, 321-343.

Saracevic, T. (1995). Evaluation of evaluation in information retrieval. In Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, Seattle, Washington, United States, SIGIR 1995 ACM, 138-146.

Saracevic, T., & Kantor, P. (1988a). A Study of Information seeking and retreiving II: Users, questions, and effectiveness. Journal of the American Society for Information Science, 39(3), 177-196.

Saracevic, T., & Kantor, P. (1988b). A Study of information seeking and retreiving III: Searchers, searches, and overlap. Journal of the American Society for Information Science, 39(3), 197-216.

Saracevic, T., Kantor, P., Chamis, A., & Trivison, D. (1988). A Study of information seeking and retreiving I: Background and methodology. Journal of the American Society for Information Science, 39(3), 161-176.

Savolainen, R. (1993). The Sense-making theory: Reviewing the interests of a user-centered approach to information seeking and use. Information Processing & Management, 29(1), 13-28.

Schamber, L. (1994). Relevance and information behavior. In M. E. Williams (Ed.), Annual Review of Information Science and Technology (ARIST) (Vol. 29, pp. 3-48). Medford, NJ: Learned Information, Inc.

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A Re-examination of relevance: Toward a dynamic, situational definition. Information Processing and Management, 26, 755-775.

Scholl, B. J. (2001). Objects and attention: The state of the art. Cognition, 80, 1-46.

Schvaneveldt, R. W. (1990). Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex.

Shute, S. J., & Smith, P. J. (1993). Knowledge-based search tactics. Information Processing and Management, 29(1), 29-45.

Sihvonen, & Vakkari, P. (2004). Subject knowledge improves interactive query expansion assisted by a thesaurus. Journal of Documentation, 60(6), 673-690.

Simon, H. (1996). Sciences of the Artificial (3rd ed.). Cambridge, MA: MIT Press.

Spink, A. (1994). Term relevance feedback and query expansion: Relation to design. In Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, p.81-90, July 03-06, 1994, Dublin, Ireland

Sugar, W. (1995). User-centered perspective of information retrieval research and analysis methods. In M. E. Williams (Ed.), Annual Review of Information Science and Technology (ARIST) (Vol. 30). Medford, N.J.: Information Today, Inc.

Sullivan, M., Borgman, C. L., & Whippern, D. (1990). End-users, mediated searches and front-end assistance programs on dialog: A comparison of learning, performance and satisfaction. Journal of the American Society for Information Science, 41(1), 27-42.

Suttcliffe, A. G., Ennis, M., & Hu, J. (2000). Evaluating the effectiveness of visual user interfaces for information retrieval. International Journal Human-Computer Studies, 53, 741-763.

Suttcliffe, A. G., Ennis, M., & Watkinson, S. J. (2000). Empirical studies of end-user information searching. Jounal of the American Society for Information Science, 51(13), 1211-1231.

Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. College and Research Libraries, 29: 178-194.

Vakkari, P. (2002). Subject knowledge, source of terms and term selection in query expansion. Proceedings of the 24th ECIR, Berlin and Heidelberg.

Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal a longitudinal case study. Information Processing and Management, 39, 445-463.

Veerasamy, A., & Belkin, N. J. (1996). Evaluation of a tool for visualization of information retrieval results. Proceedings of the 19th annual international ACM SIGIR conference on research and development in information retrieval ACM, 85-92.

Ware, C. (2000). Information Visualization: Perception for design. Morgan Kauffman, San Francisco, CA

Westerman, S. J., & Cribbin, T. (2000). Mapping semantic information in virtual space: Dimensions, variance and individual differences. International Journal Human-Computer Studies, 53, 765-787.

White, H. D., Lin, X., & Buzydlowski, J. (2004). An Associative information visualizer. Paper presented at the INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'04).

White, H. D., & McCain, K. W. (1989). Bibliometrics. In M. E. Williams (Ed.), Annual Review of Information Science and Technology (ARIST) (Vol. 24, pp. 110-184): Elsevier Sceince Publishers B.V.

White, H. D., & McCain, K. W. (1997). Visualization of literatures. In M. E. Williams (Ed.), Annual Review of Information Science and Technology (ARIST) (Vol. 32, pp. 99-168). Medford, NJ: Information Today.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: an author co-citation analysis of information science, 1972-1995. Journal of the American Society for Information Science, 49(4), 327-355.

White, H. D., & McCain, K. W. (2000). In memory of Belver C. Griffith. 51(10), 959-962.

White, R., & Marchionini, G. (2007). Examining the effectiveness of real-time query expansion. Information Processing and Management, 43(3), 685-704.

Wildemuth, B. (2004). The Effects of domain knowledge on search tactic formulation. Journal of the American Society for Information Science, 55(3), 246-258.

Williams, J. G., Sochats, K. M., & Morse, E. (1995). Visualization. In M. E. Williams (Ed.), Annual review of information science and technology (ARIST, pp. 525): American Society for Information Science.

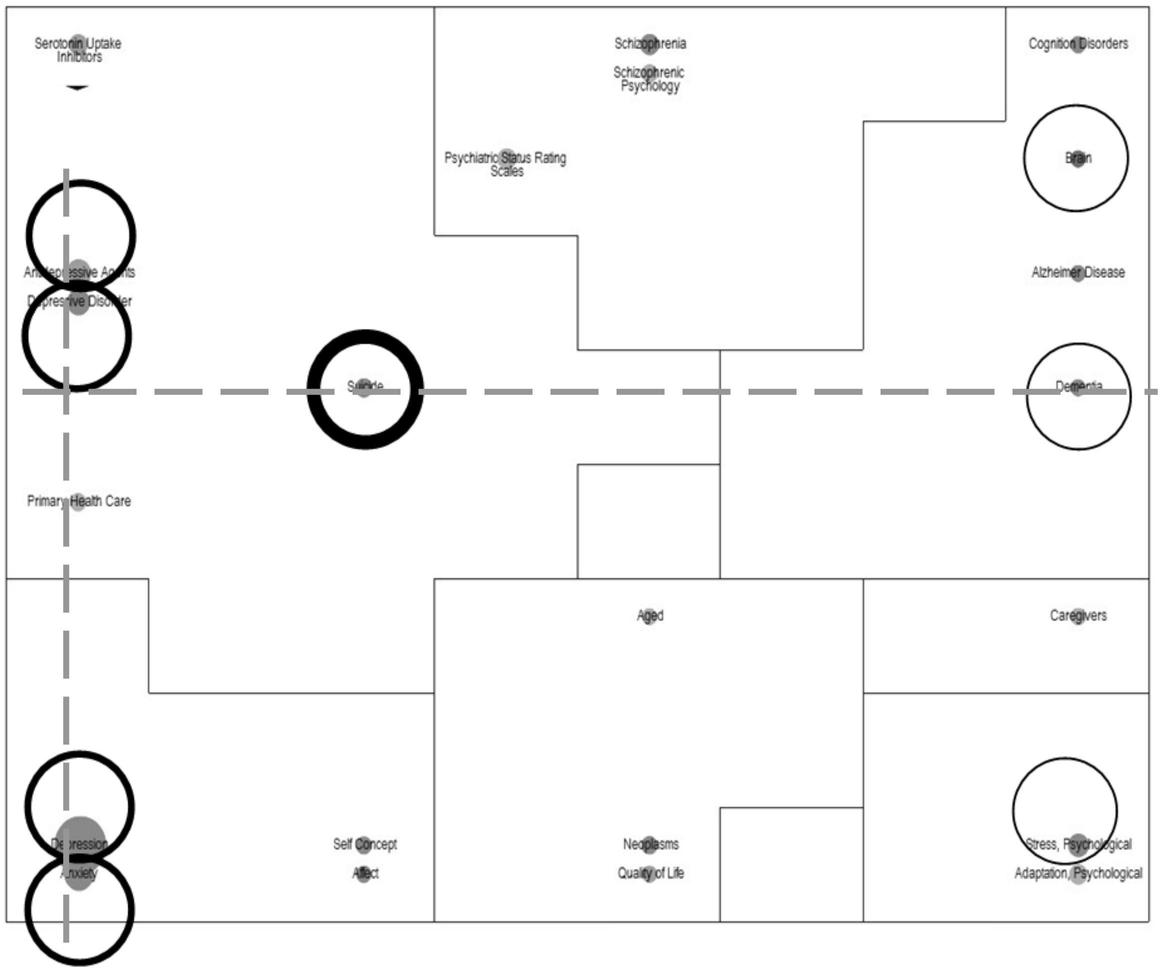
Wilson, T. D. (1981). On UsesStudies and information needs. Journal of Documentation, 37(1), 3-15.

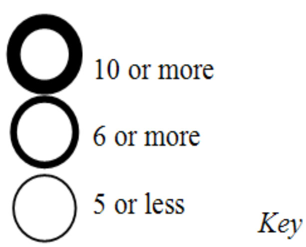
Wilson, T. D. (1999). Models in information behaviour research. Journal of Documentation, 55(3), 249-270.

Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., et al. (1995). Visualizing the nonvisual: Spatial analysis and interaction with information from text documents. IEEE Information Visualizations, 95.

Zipf, G. K. (1949). Human behavior and the principle of least effort. Cambridge, Mass.: Harvard University.

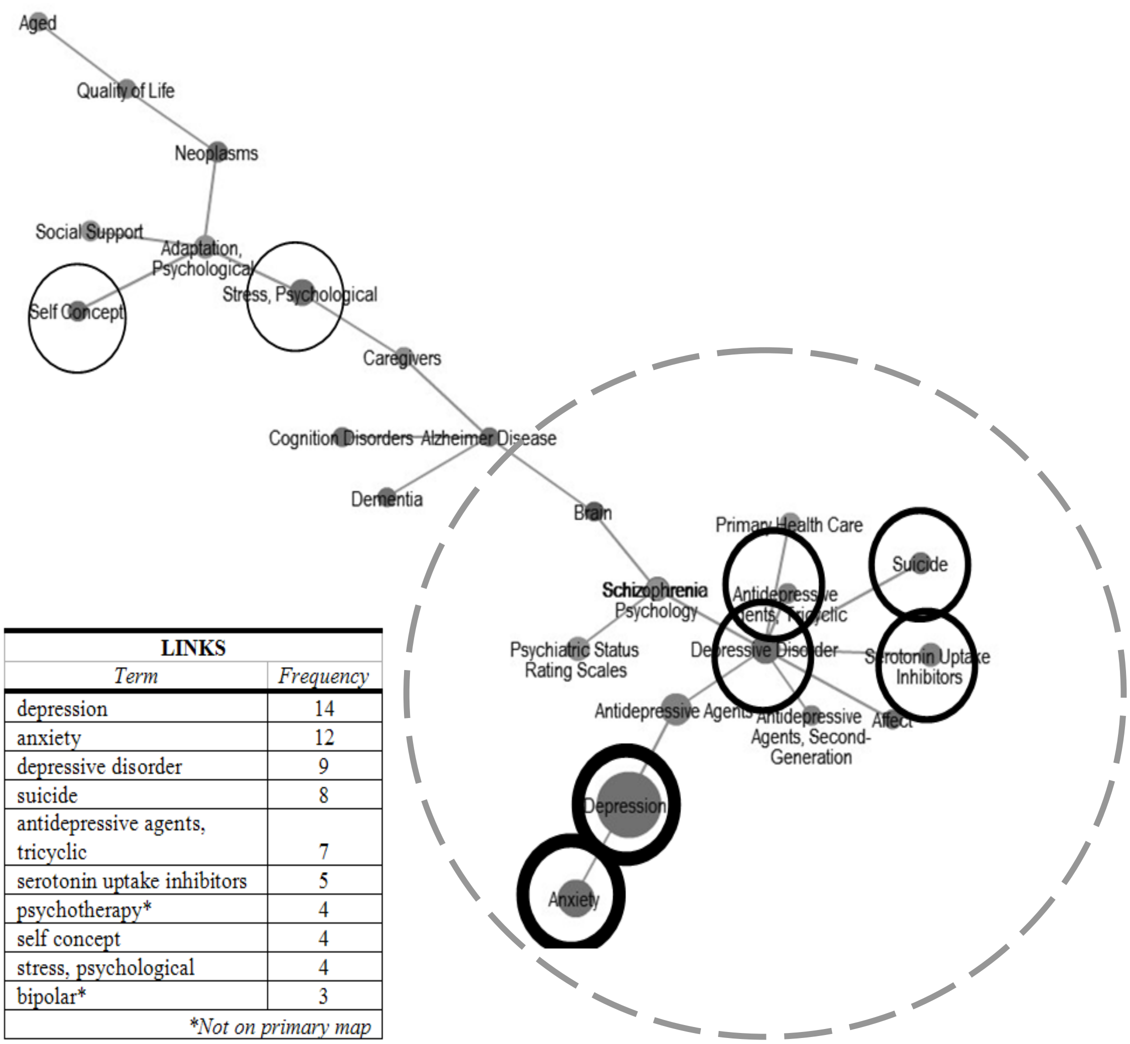
Appendix A: SOM and PathFinder Network of Depression with Top Ten Post-test Participant Terms



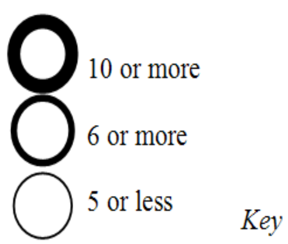


Kohonen SOM of Depression Showing the Top Ten Terms Used by Participants Post-Test

SOM	
Term	*Frequency*
suicide	10
depression	9
antidepressive agents	8
anxiety	7
depressive disorder	6
bipolar disorder*	4
stress	3
antimanic agents*	2
brain	2
dementia	2
**Not on primary map*	



LINKS	
Term	*Frequency*
depression	14
anxiety	12
depressive disorder	9
suicide	8
antidepressive agents, tricyclic	7
serotonin uptake inhibitors	5
psychotherapy*	4
self concept	4
stress, psychological	4
bipolar*	3
	**Not on primary map*



Pathfinder Network of Depression Showing the Top Ten Terms Used by Participants Post-Test

Appendix B: Institutional Review Board Policy & Documents

University of Maine at Augusta
INFORMED CONSENT

Subject's Initials ________ Page 1 of 2

1. Participant Name: ____

2. Title of Research: The Impact of Concept Maps on the Information Behavior and Learning of Novices in the Information Retrieval Context

3. Investigator's Name: Jodi C. Williams

I am being asked to read the following material to ensure that I am informed of the nature of this research study and of how I will participate in it, if I consent to do so.

Signing this form will indicate that I have been so informed and that I give my consent. Federal regulations require written informed consent prior to participation in this research study so that I can know the nature and risks of my participation and can decide to participate or not participate in a free and informed manner.

PURPOSE
I am being invited to participate voluntarily in the above-titled research project. The purpose of this project is to understand the impact on learning and retrieval that concept maps have when used during the information searching process.

SELECTION CRITERIA
I am being invited to participate because I am a student at the University of Maine at Augusta and not an expert at searching for information using library databases.

PROCEDURE(S)
If I agree to participate, I will be asked to consent to the following:

- Fill out questionnaire about my computer use
- Fill out questionnaires about searching
- Write down terms I might use when searching
- It should take about 25 minutes
- The information collected will be completely anonymous.

RISKS
There are no known risks for participating in this research.
BENEFITS
A participant may benefit by gaining knowledge about searching library databases.

CONFIDENTIALITY
Consent forms will be collected and secured. No personal identifiable information will be placed on collected survey documents. Participants will only be referred to by numbers. Only the principal investigator will have access to the data and information collected.

Subject's Initials ________   Page 2 of 2

PARTICIPATION SUBJECT COMPENSATION
None.

CONTACTS
I can obtain further information from the principal investigator, Jodi C. Williams or Grace Leonard, Dean at (207)621-3341 or (207) 621-3257. If I have questions concerning my rights as a research subject, I may call the UMA Provost's office (207) 621-3106.

AUTHORIZATION
BEFORE GIVING MY CONSENT BY SIGNING THIS FORM, THE METHODS, INCONVENIENCES, RISKS, AND BENEFITS HAVE BEEN EXPLAINED TO ME AND MY QUESTIONS HAVE BEEN ANSWERED. I MAY ASK QUESTIONS AT ANY TIME AND I AM FREE TO WITHDRAW FROM THE PROJECT AT ANY TIME WITHOUT CAUSING BAD FEELINGS.

This consent form will be filed in an area designated by the human subjects committee with access restricted to the principal investigator, _ Jodi C. Williams**_ or faculty sponsor, and authorized representative of the** _Natural & Social Sciences_College**. I do not give up any of my legal rights by signing this form. A copy of this signed consent form will be given to me.**

______________________ ______________________
Subject's Signature Date

I have carefully explained to the subject the nature of the above project. I hereby certify that to the best of my knowledge the person who is signing this consent form understands clearly the nature, demands, benefits, and risks involved in his/her participation and his/her signature is legally valid. A medical problem or language or educational barrier has not precluded this understanding.

__________________________________ Date:
Signature of Investigator

__________________________________ Date:
Signature of Faculty Sponsor

University of Maine at Augusta
HUMAN SUBJECTS PROJECT APPROVAL FORM (APPLICATION)

Please complete the following information and submit this form with any other applicable documents. (For simple anonymous surveys, completion of this form, a separate paragraph description stating the purpose of the research and a sample of the survey tool are usually adequate for approval)

PRINCIPAL INVESTIGATOR'S NAME:___Jodi C. Williams_________________________

MAILING ADDRESS: ______Augusta, ME 04330________________

TELEPHONE: ____621-3341_______________ E-MAIL:__jodi.williams@maine.edu______

STATUS (Check One):

____X___ Faculty Member, _L&IS, Natural and Social Sciences_ (Program or Department)

______ UMA Department Representative, ______________________ (Program or Department)

_______Student, ___________________ (Major), ______________Number of credits completed

TYPE OF PROJECT:

_______ Thesis/Project Chairperson _______________________

______ Class Project Course ____________________________

Faculty Sponsor___________________

______ UMA Research Project Department Head__________________

__X__ Other (Specify) Dissertation Research Responsible Party: _Michael Atwood, Ph.D.
Drexel University________

PLEASE CHECK ALL INFORMATION THAT IS ATTACHED

___X___Copy of proposal (Proposal may be entered into the boxes on the Consent Form Checklist or may stand alone as a separate document)

__X____Disclaimer/Consent Form (if applicable

_______Assent Form (if applicable)

___X___Instruments/Questionnaires (if applicable)

_______Recruitment Advertisement/Script (if applicable)

X______Approval letter from project/thesis Advisor or Department Head as needed (Signatures on this form are sufficient for simple anonymous surveys)

_______Checklist(s)

RESEARCH AUTHORIZATION:

____

Signature of Principal Investigator (date)

____

Signature of Faculty Member or Department Head Sponsoring Project (date)

____

Recommendation by UMA IRB Chair (date)

____

Approval by President's Designee (date)

____

Response Transmitted and Document Filed in Provost's Office by: (date)

Appendix C: Experiment Instructions

1. Introduction & Orientation

Good afternoon/morning. My name is Jodi Williams and I will be serving as the test monitor for all of the sessions. I'd like to thank you for participating in this study

The purpose of this study is to measure the impact different types of displays have on your information behavior and learning.

You will be asked to look at two displays with a hypothetical information need and answer some brief questions. In no way are we evaluating your performance. This test will only measure your reaction to the systems through questions. We are interested in learning about your reactions and use of the different systems you explore.

All data collected from test sessions will be held completely confidential and results will be anonymous. Please try to perform the tasks in a manner that you normally would in a school environment and feel free to ask questions if you need to. Be aware, however, that we cannot answer some questions because they may ultimately negate the very information that we are trying to obtain.

The entire process will take place here in this room and you will need no other tools than a pen or pencil and the handouts we provide. Restrooms are located outside this room. Please let me know should you need a break before we have finished.

The experiment includes exploring a display type provided to you and your reaction to that display system. those in relation to a medical topic you would search on for a class project. You do not need to do any research, we are interested only in your reactions to the maps. For each task, you will be given a set of documents that will help in your hypothetical searching. Again, we are evaluating the impact of the tools and not your abilities.

As the test monitor, I will be serving as a neutral observer and may take notes during the study. A research assistant may also be present to observe the session. The entire test session should last no more than 30 minutes.

There are a few forms for you to complete before we get started. One is a consent form which gives us your permission to observe and collect data from the session. There is also a general background questionnaire to collect basic information. Test participation is voluntary and no compensation will be awarded.

All forms are completely anonymous. For each search task there are a set of questions to answer. If you finish on any section early, please be patient and wait for your fellow participants to finish. We will go through the steps together for each part.
Before we begin, do you have any questions?

<u>2. Data Gathering: HANDOUT: Handout and collect Informed consent.</u>

<u>3. Hand out:</u> General Information questionnaire 5 minutes

<u>4. Tool Introduction</u>
Imagine you are enrolled in a health class here at UMA, for a presentation you are asked to use a library database to search for information on a topic like diabetes. Please click on the internet explorer window at the bottom of the desktop. This is a system that allows you to see how a term like diabetes is related to other medical terms. It arranges the related terms in different formats. Please take a look at the screen in front of you. By using your mouse, and double-clicking on the words you can bring up other terms related to the word you clicked on.

<u>5. Hand out:</u> Diabetes Test Query
Please spend the next 5 minutes exploring the system in front of you on a hypothetical search topic. This practice question will set the scene for searching and help you become acquainted with the other questions we will use during this experiment.

<u>6. Exploration Session 1: Cholesterol, Depression</u>
We will now begin the experiment. For the tasks, you will be given a set of documents that will help in your searching. Please remember, we are testing the systems and not your abilities

- **<u>Hand out: Query Scenario</u>**
  Please take a look at the query in front of you and make sure you understand what it is asking..
- **<u>Hand out: Pre-Tool Test</u>**
  - Take pre-tool test
- **<u>Exploration</u>**
  Now using the display in front of you, please take a look at and explore the tool based on the written task scenario you were given. Please do not go to any other websites and use only the tool in front of you. Please click only on the terms as presented on the screen.
- **<u>Hand out: Post-Tool Test</u>**
  Now that you have finished exploring, please fill out the following.
  - Take Post-Tool test

<u>7. Exploration Session 2: Depression, Cholesterol</u>
We will now begin our second exercise.

- **<u>Hand out: Query Scenario</u>**
  Please take a look at the query in front of you and make sure you understand what it is asking you to do.
- **<u>Hand out: Pre-Tool Test</u>**
  - Take pre-tool test

- **Exploration**
  Now using the display in front of you, please take a look at and explore the tool based on the written task scenario you were given. Again, please do not go to any other websites and use only the tool in front of you. Please click only on the terms as presented on the screen.
- **Hand out: Post-Tool Test**
  Now that you have finished exploring, please fill out the following.
  - Take Post-Tool test

8. Hand out: Visualization Reaction questionnaire 10 minutes

9. Conclusion

Thank you for your participation in this study. Findings and recommendations will be compiled and shared with the research team, as well as with the study participants who are interested.

Appendix D: General Information Questionnaire

University of Maine at Augusta/Drexel University

Demographics: *Please check the most appropriate selection.*

1. Gender:
   - ❑ Female
   - ❑ Male

2. Age range:
   - ❑ under than 21
   - ❑ 21-34
   - ❑ 35-49
   - ❑ 50-64
   - ❑ 65 or older

3. Please indicate what approximate year you are in college
   - ❑ Freshman
   - ❑ Sophomore
   - ❑ Junior
   - ❑ Senior

4. Indicate your level of medical knowledge.
   - ❑ Extensive
   - ❑ Moderate
   - ❑ Little
   - ❑ None

Internet/Web Experience: *Please check the most appropriate selection*

5. Do you use online search engines (i.e. Google, Yahoo, AltaVista, etc.)?
   - ❑ Yes
   - ❑ No

6. Do you use online resources from the library to find articles and books (Ursus, Minerva, Proquest, Ovid, etc.)?
   - ❑ Always
   - ❑ Sometimes
   - ❑ Occasionally
   - ❑ Never

<u>Related Knowledge/Experience</u>: *Please check the most appropriate selection*

7. Indicate your level of familiarity/experience with Medical Subject Headings (MeSH) indexing terms.
   - ❑ Extensive
   - ❑ Moderate
   - ❑ Little
   - ❑ None

8. Indicate your level of experience searching medical literature.
   - ❑ Extensive (on a weekly basis)
   - ❑ Moderate (a few times a month)
   - ❑ Little (couple times a year)
   - ❑ None

Appendix E: Visualization Reaction Questionnaire

University of Maine at Augusta/Drexel University

General Information/Search Process: *Please check the most appropriate selection*
INSTRUCTIONS: *Please answer the following questions about the tools you used while searching. The tool refers to the additional resource you were given to explore the medical topic.*

Place an X in the box that best fits your thoughts.	Strongly Agree 1	Somewhat Agree 2	Somewhat Disagree 3	Strongly Disagree 4
1. Using the system made coming up with search terms easier.				
2. I understood the medical topics better using the system.				
3. The system helped me to formulate my search.				
4. The system was easy to understand.				
5. Seeing the organization of the medical topics was helpful.				
6. The system was visually appealing.				
7. The system made no sense at all.				
8. I would have preferred not using the system.				
9. I would use the system again if I had the option.				
10. I learned about the medical topic with the system.				
11. I know more about Medical Subject Headings.				
12. The system helped me remember information I already knew about the medical topic.				
13. After using the system I decided to change what I was looking for.				

14. If you changed what you were looking for after using the tool, please explain why in the space below.

____

____

____

15. Do you have any general comments about the tool you were given to use?

____

____

____

You are Finished!

Thank you for your participation in this research study! Your input is greatly appreciated and will help us better understand the impact of concept maps on searching for articles in the information retrieval environment.

Appendix F:Task Scenario & Pre-Test-CH

University of Maine at Augusta/Drexel University

For your health class you have been asked to give a 30 minute presentation on cholesterol. You need to find out as much about the topic as possible and locate articles that contain the information needed.

INSTRUCTIONS: Please fill in with the appropriate information on the form below.

INSTRUCTIONS: What terms or term phrases would you use to search for information on the above topic? (*For example, what words would you type into the computer if you searched using a library database?)*

______________________ ______________________ ______________________

______________________ ______________________ ______________________

______________________ ______________________ ______________________

Perceptions: *Please circle the most appropriate selection based on the statements below.*

1. I am satisfied with the terms I used above.	1 2 3 4 5 (not at all … very)
2. I am confident these terms will help me locate resources for the assigned task.	1 2 3 4 5 (not at all … very)
3. I would be successful locating information for this project using the search terms above.	1 2 3 4 5 (not at all … very)
4. I feel the terms I used above are relevant to the search topic.	1 2 3 4 5 (not at all … very)

Please state what you are looking for?

____

____

Describe in the space below what you already know about this topic.

____

____

____

____

1. On a scale of 1 to 5, how would you rank the amount of knowledge you possess on this topic? | 1 2 3 4 5 |
little personal knowledge — considerable personal knowledge

-Wait for instructions before continuing on to the next part-

Appendix G: Task Scenario & Pre-Test-DE

University of Maine at Augusta/Drexel University

INSTRUCTIONS: Please fill in with the appropriate information on the form below.

For a psychology class you have been asked to write a 20 page paper on depression. You need to find out as much about the topic as possible and locate articles that contain the information needed.

INSTRUCTIONS: What terms or term phrases would you use to search for information on the above topic? (*For example, what words would you type into the computer if you searched using a library database?)*

________________________ ________________________ ________________________

________________________ ________________________ ________________________

________________________ ________________________ ________________________

Perceptions: *Please circle the most appropriate selection based on the statements below.*

Statement	1 (not at all)	2	3	4	5 (very)
1. I am satisfied with the terms I used above.	1	2	3	4	5
2. I am confident these terms will help me locate resources for the assigned task.	1	2	3	4	5
3. I would be successful locating information for this project using the search terms above.	1	2	3	4	5
4. I feel the terms I used are relevant to the search topic.	1	2	3	4	5

Please state what you are looking for?

____

____

Describe in the space below what you already know about this topic.

____

____

____

____

1. On a scale of 1 to 5, how would you rank the amount of knowledge you possess on this topic? | 1 2 3 4 5 |
   little personal knowledge — considerable personal knowledge

-Wait for instructions before continuing on to the next part-

Appendix H: Post-Test-DE

University of Maine at Augusta/Drexel University

Topic Information: *Now that you have explored the topic, answer the following questions.*

After exploring the tool, what terms or term phrases would you use now to search for information on this topic? (*For example, what words would you type into the computer if you searched using a library database?)*

_______________________ _______________________ _______________________

_______________________ _______________________ _______________________

_______________________ _______________________ _______________________

Perceptions: *Please circle the most appropriate selection based on the statements below.*

Statement	not at all				very
1. I am satisfied with the terms I used above.	1	2	3	4	5
2. I am confident these terms will help me locate resources for the assigned task.	1	2	3	4	5
3. I would be successful locating information for this project using the search terms above.	1	2	3	4	5
4. I feel the terms I used are relevant to the search topic.	1	2	3	4	5

Did you use any terms from the computer tool you explored? ___yes ___no

Please state what you were looking for?

____

____

Describe what you know about depression now that you have explored it using the tool.

____

____

____

____

(Please mark the best choice.)

This system helped me: ___learn something new ___remember things I had forgotten
___both (learn & remember) ___neither (learn or remember)

Perceptions: *Please circle the most appropriate selection based on the statements below.*

1. I feel I know more about the topic than I did before searching. | 1 2 3 4 5 | not at all … very
2. My post-search knowledge has corrected what I knew before searching. | 1 2 3 4 5 | not at all … very

Appendix I: Post-Test-CH

University of Maine at Augusta/Drexel University

Topic Information: *Now that you have explored the topic, answer the following questions.*

After exploring the tool, what terms or term phrases would you use now to search for information on this topic? (*For example, what words would you type into the computer if you searched using a library database?)*

______________________ ______________________ ______________________

______________________ ______________________ ______________________

______________________ ______________________ ______________________

Perceptions: *Please circle the most appropriate selection based on the statements below.*

Statement	1 (not at all)	2	3	4	5 (very)
1. I am satisfied with the terms I used above.	1	2	3	4	5
2. I am confident these terms will help me locate resources for the assigned task.	1	2	3	4	5
3. I would be successful locating information for this project using the search terms above.	1	2	3	4	5
4. I feel the terms I used are relevant to the search topic.	1	2	3	4	5

Did you use any terms from the computer tool you explored? ___yes ___no

Please state what you were looking for?

____

____

Describe what you know about depression now that you have explored it using the tool.

____

____

____

____

(Please mark the best choice.)

This system helped me: ___learn something new ___remember things I had forgotten
___both (learn & remember) ___neither (learn or remember)

Perceptions: *Please circle the most appropriate selection based on the statements below.*

1. I feel I know more about the topic than I did before searching. | 1 2 3 4 5 | not at all … very
2. My post-search knowledge has corrected what I knew before searching. | 1 2 3 4 5 | not at all … very

Appendix J: Pilot Study

Changes Due to Pilot Study

Before the pilot studies were conducted, there was a walkthrough experiment with two participants for the purpose of testing each phase of the experiment with the instructions to make sure that there were no inconsistencies or confusion on the part of the participants. It was also to elicit general feedback about the participants experience in the research overall. This preliminary walkthrough consisted of three task scenarios and was 2+ hours in length. After discussion with the participants and upon further reconsideration about cognitive load, the decision was made to eliminate one of the task scenarios so that there were only two searches during the data gathering of the experiment. Following this walkthrough, any confusion or inconsistency in the instructions or questionnaires mentioned by the subjects was examined and changes were made based on their suggestions.

There were two primary pilot studies for this experiment. The first tier had three participants and the second tier had four. All participants were undergraduate students at the University of Maine at Augusta asked at random to participate in the pilot study. These students were representative of the intended participant audience for the full experiment and measures were taken so that they will not be included in the actual experiment.

The three participants in the first pilot study were representative of each facet of the dependent variable (Map1, Map2, List). Again, participants were volunteers at the University of Maine at Augusta and followed all steps of the actual experiment. This first pilot study verified instructions as well as time to complete tasks and the process for the card sorting exercise.

The first noted change for this pilot study was the elimination of the third retrieval which brought the experiment time down from 120 minutes to 90 minutes. Also, with respect to discussion about cognitive load, the second tier pilot study participants felt that they were not taxed mentally with the two queries.

Another change, larger in impact after the first pilot test, was the decision to add a fourth group to the experiment. In order to fully measure the impact concept maps have on information retrieval and learning, it was decided there should be a group using nothing but the standard bibliographic interface with no supplemental search tool. The three initial groups were List, Map_1, and Map_2.

If this research does indeed seek to measure the impact of concept maps on learning that takes place during the retrieval process then the interaction of a user with PubMed, without any supplemental search tools, is important to that task. Also, depending upon how broadly one defines visualization, providing a simple term list to a participant in an experiment is also a form of providing a visualization tool. Adding the No Tool group will allow a comparison of the term list group to the no tool group to see if there are any differences in our dependent variables (Map_1, Map_2 and List). This also provides the means to determine if there is a difference in learning between subjects with the maps, and those with the list of terms. The data collected will further be able to show if it was just the simple list of terms on a page, or it was actually the relationships between concepts illustrated on the maps which impacted learning.

The last primary change occurring as a result of the pilot studies was to the instructions for the card sort. In the second pilot test, participants were unsure if they could leave concepts in an "unknown" pile. This question was asked by the first participant after the instructions were read, but before the card sorting started. The two other participants did leave concepts aside in an unknown pile, but asked after the experiment was completed and during the interview if this was ok.

After the last changes to the experiment a final pilot test was run with 4 participants, 1 for each group of the independent variable (No Tool, List, Map1, and Map2). Some of the preliminary data for the pilot studies is presented at the end of this methodology section after the research questions and variables are introduced. General descriptive statistics are presented to support chosen methods for the primary research questions.

Vita

Jodi C. Williams
Augusta, ME (207) 621-3341, jodi.Williams@maine.edu

Education: Ph.D., Drexel University, 2007
M.L.S., Southern Connecticut State University, 1999
B.A. - English Literature, Southern Connecticut State University, 1997

Professional Experience: Dr. Williams has taught Library and Information Science both online and on-ground since 1999. Currently she is Assistant Professor & Coordinator of the Information & Library Services program at the University of Maine at Augusta(UMA). She works with other ILS faculty at UMA to offer educational opportunities to students in Maine, the United States, and around the world using various forms of technology. Her research interests include design of information systems and human-computer interaction, system engagement and the affective aspects of system design.

Selected Publications & Presentations:

Williams, J.C. (2006) "Support Staff in the Library; A presentation on the history, role and current issues." Maine Library Association Conference

Atwood, M. Gross, M., McCain, K. & J. Williams (2005) "Science of Design: Why We Need It and Why It is so Difficult to Achieve." Human Computer Interaction Consortium. Winter Workshop, Colorado.

Williams, J. (2004) "Mentoring in the Library: Inspiring Growth and Sharing Ideas… What we need to know and where we can learn it!" New England Technical Services Librarians Conference. Holy Cross, Worcester, MA.

Atwood, M.E., McCain, K.W., & Williams, J.C. (2002) "How does the design community think about design?" In *Proceedings of Designing Interactive Systems 2002,* New York: ACM, 2002, pp125-132.

Williams, J. Goodrum, A. (2001) "Scholarly Publishing on the Web: Link Analysis of the Top 200 Highly Cited Computer Science Articles on the WWW". Proceedings of the 64TH ASIS&T Annual Meeting. Vol. 38, 2001

1998-1999 Graduate Research Fellow: Southern Connecticut State University "Exploring The Patterns Of Information-Seeking Behavior And The Pattern Of Information Use Among Undergraduate Students At Southern Connecticut State University"