The Construction of Mental Models of Information-rich Web Spaces: The Development Process and the Impact of Task Complexity

Yan Zhang

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Information and Library Science

Chapel Hill

2009

Approved by:

Barbara Wildemuth

Gary Marchionini

Paul Solomon

Diane Kelly

Javed Mostafa

Peiling Wang

Abstract

YAN ZHANG: The construction of mental models of information-rich web spaces: The development process and the impact of task complexity
(Under the direction of Barbara Wildemuth)

This study investigated the dynamic process of people constructing mental models of an information-rich web space during their interactions with the system and the impact of task complexity on model construction. In the study, subjects' mental models of MedlinePlus were measured at three time points: after subjects freely explored the system for 5 minutes, after the first search session, and after the second search session. During the first search session, the 39 subjects were randomly divided into two groups; one group completed 12 simple search tasks and the other group completed 3 complex search tasks. During the second search session, all subjects completed a set of 4 simple tasks and 2 complex tasks. Measures of the subjects' mental models included a concept listing protocol, a semi-structured interview, and a drawing task.

The analysis revealed that subjects' mental models were a rich representation of the cognitive and emotional processes involved in their interaction with information systems. The mental models consisted of three dimensions (structure, evaluation and emotion, and (expected) behaviors); the structure and evaluation/emotion dimensions consisted of four components each: system, content, information organization, and interface. The construction of mental models was a process coordinated by people's internal cognitive structure and the external sources (the system, system feedback, and tasks) and a process distributed through time, in the sense that earlier mental models impacted later ones. Task complexity also impacted the construction of mental models by influencing what objects in the system were perceived and represented by the user, the specificity of the representations, and the user's feelings about the objects.

Based on the study results, recommendations for employing mental models as a tool to assist designers in constructing user models, eliciting user requirements, and performing usability evaluations are put forward.

Dedication

To my father.

Acknowledgements

The writing of a dissertation could be a solitary journey. But many people made mine an enjoyable process. I want to first thank my advisor, Barbara Wildemuth, for being positive, encouraging, and inspiring. This work would not have been possible without her guidance. My genuine appreciations also go to other members of my dissertation committee, Gary Marchionini, Paul Solomon, Diane Kelly, Javed Mostafa, and Peiling Wang for their input and support.

In shaping my research questions, I benefited a lot from the discussions with Dr. Peter Ornstein in the Department of Psychology at UNC, and John Schumacher at IBM Lotus Software Group. They helped me see the meaning and significance of this research from other perspectives.

I also want to express my appreciation to the SILS community, the faculty, staff, and students. My journey of pursuing a PhD would be less rewarding without your company. For many kinds of support that I have received for crafting my dissertation, I am particularly grateful to Deborah Barreau, Rachael Clemens, Jane Bahnson, Sanghee Oh, Sarah Redeem, Cassidy Sugimoto, Amber Cushing, Weimao Ke, Dana Hanson-Baldauf, Carolyn Hank, Dihui Lu, Jungsun Oh, Songphan Choemprayong, Laura Sheble, and Leo

Cao.

Finally, I am grateful that I have a loving and supportive family. My mom and dad shared with me their persistence, courage, and humbleness, and my husband was unconditional in his support of my work.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Mental models theory is a psychological theory that has been widely employed to explore people's cognitive processes involved in their use of mechanical or computer systems. It is generally accepted that mental models help people understand a system, navigate through the system, and predict the system's behavior in future instances (Norman, 1983). Therefore, in the fields of human-computer interaction (HCI) and interactive information retrieval (IIR), research on mental models is expected to inform the design of information systems that are easy to use, intelligent in supporting users to form an appropriate understanding of the system, and effective in reducing unnecessary human errors.

The existing research on mental models of IR systems has mainly focused on one of a few areas: the characteristics of mental models, such as accuracy and completeness (e.g., Dimitroff, 1990), transfer of mental models between systems (e.g., Cool, et al., 1996), and mental models' effects on people's information searching behavior and performance (e.g., Borgman, 1986). These studies improved researchers' understanding of the mental model concept. However, they produced few readily usable tools for system and user interface (UI) designers, which directly leads to the uncertainty of the theoretical and practical meaning of the concept in the application-oriented HCI and IR fields.

Being dynamic is a feature that characterizes mental models. However, one common limitation of the existing studies is that they treat mental models as a static construct. Recognizing this problem, this study focused on exploring the dynamic processes of users constructing mental models of an IR system. Through this investigation, the usefulness of mental models as a theoretical construct and as a practical method or tool for system design was reexamined. To fulfill this purpose, this study investigated two themes: (1) the dynamic process of users constructing mental models of an information-rich web space during a search session, and (2) the impact of a major contextual factor, task complexity, on the model construction.

Examining the construction process of mental models will reveal whether people's mental models change over time and what types of changes they undergo, what aspects of the system can be easily represented and what aspects cannot. Such knowledge could be a particularly rich source for generating practical implications for system design, user modeling, and user instruction. The decision to focus on mental models construction in a short time period corresponds to the phenomenon that, in the web environment, people do not want to spend a long time figuring out the details about an IR system. If they cannot find the needed information quickly, they give up on the system. Therefore, knowledge about the development of people's mental models during such a short time is valuable for web-based system design.

Task is a variable that has consistently been reported to have significant impact on people's information searching behavior. Examining the impact of task complexity on the construction of mental models of a system places the investigation of mental models construction in a context of specific tasks.

As noted above, an information-rich web space, a special type of IR system, was used as the platform for this study. Information-rich web spaces are websites that contain a large amount of information (often thousands or millions of web pages) in a particular domain or across domains, such as wikipedia, aclu.org, and the Internet Movie Database. Information-rich web spaces are of interest because they are becoming more and more pervasive on the web. As the web keeps expanding, government agencies, military organizations, and companies are looking more and more to such hypermedia-based solutions for disseminating information (Farris, Jones, & Elgin, 2002). Meanwhile, this type of website has different structures and encourages different search strategies than traditional IR systems, and so requires additional investigation (Shneiderman, 1997).

As an inter-disciplinary study, the research will contribute to psychology, HCI, and information science (IS). For psychology, exploring the construction of mental models is useful for understanding the development of mental representations, as well as the role of mental models in problem solving. For HCI, the exploration could provide system and UI designers a model-based view of end users of information systems. Different from discrete design guidelines, such a structured view of users could help designers better understand user needs and requirements and form a holistic view for a design. For IS, investigating mental model construction in the context of solving particular tasks helps researchers understand the process of users making sense of an IR system, therefore, increasing the understanding of cognitive mechanisms underlying users' information behavior.

Chapter 2: Literature review

This chapter reviews mental models from various aspects. It begins with a discussion of how mental models have been defined within psychology. Next, the application of mental models in HCI is discussed. Then, we focus on the research on mental models of IR systems, followed by a review of techniques that have been used to elicit users' mental models and to explore the effects of mental models on people's information search behavior. This chapter ends with a review of the roles of two variables, spatial ability and task complexity, in people's interactions with IR systems, which establishes the rationale for controlling subjects' spatial ability and exploring the impacts of task complexity on their mental models in the current study.

2.1 Mental models in psychology

2.1.1 Mental models in logical reasoning and language comprehension

The term "mental model" was coined by Craik in his 1943 book, *The Nature of Explanation*. He proposed that mental models are internal representations of external objects and phenomena; and mental models consist of words, numbers, and other symbols. Johnson-Laird, in his 1983 book, *Mental Models*, further developed the construct of mental model into a mental model theory as an effort to explain human beings' deductive reasoning and fallacies that occurred in the reasoning.

Cognitive psychologists have traditionally argued that people employ formal *inference rules*, such as:

If A then B

A

Therefore B

in their reasoning (e.g. Braine, 1978). When they fail to recover the correct logical form of the premises, or fail to construct a mental derivation of the conclusion, they make errors. However, this formal logic argument ignores the influence that the semantic meaning of premises has on people's cognitive reasoning (Byrne, 1992). Research in decision making has consistently shown that, when the content of an abstract logical inference is substituted with real-life content, subjects' decision accuracy rate increases (e.g., Stanovich, 1999).

To account for the influence of meaning in logical reasoning, Johnson-Laird (1983) proposed that, instead of employing formal logic, people construct mental models to understand and interact with the external world. A *mental model* is "a representation of the way the world would be if the premises were true" (Byrne, 1992: 12). Based on this argument, people will construct an internal model of the situation that the premises describe in a deductive reasoning task. When the content of the premises is familiar, the models can be easily "fleshed out" by incorporating any additional information implied by the situation. People make valid inferences when their conclusion is true in every model that can be constructed of the premises; and they make errors when they fail to consider all the possible models of the premises. This proposal is supported by the empirical evidence that the fewer models of the premises that need to be constructed, the fewer errors people make in an inference (Johnson-Laird, Byrne, & Schaeken, 1992).

An example extracted from Byrne (1992) illustrates the role of mental models in deductive reasoning. To make the inference:

> If there is a triangle then there is a circle
>
> There is not a circle
>
> Therefore, there is not a triangle

Based on the mental model theory, the meaning of the conditional premise is represented by a set of models:

The first conditional premise is represented as:	[Δ]	o
The second conditional premise is represented as:	¬ o	

To make the correct inference, the models must make explicit the alternative states of affairs. The fully fleshed-out models for the premise are:

Δ	o
¬ Δ	o
¬ Δ	¬ o

When the information from the second premise is added to this set of models, the models with a circle are eliminated, which leaves the third model:

¬ Δ	¬ o

This model supports the conclusion.

Mental model theory also helps explain the cognitive processes of language comprehension. Unlike other psychological approaches, which "emphasize the words themselves, as entries in a mental dictionary, as linked semantic entities, or rules for specifying relations between words" (Ehrlich, 1996: 224), mental models theory emphasizes the content of the words, and asserts that people construct mental models based on the content (Wason & Johnson-Laird, 1972). Meanwhile, mental models share a similar relation-structure to that of the reality they imitate (Johnson-Laird, 1983). Both the semantic and structural similarity of mental models to reality enables people to derive meanings from mental models. Johnson-Laird (1981: 117) has stated that:

> The psychological theory of meaning that I wish to advance assumes that the mental representation of a sentence can take the form of an internal model of the state of affairs characterized by the sentence.

2.1.2 Mental models in representing domain knowledge

In the same year in which Johnson-Laird published his mental model theory, Gentner and Steven (1983) edited and published a highly influential book, also titled *Mental Models*. In this book, mental models were defined as people's mental representations of domain knowledge that provides the basis for people to make inferences about the domain (Roger, 1992). For example, diSessa (1983) explored undergraduate students' mental models of simple physical mechanisms, such as springiness. Forbus (1983) simulated people's qualitative reasoning about space and motions. Gentner and Gentner (1983) were concerned about people's understanding of electricity. Two schools of thought emerged from these studies: one argues that people develop coherent but naïve theories of physical phenomena, while another argues that people possess only fragmented knowledge which by nature is a set of loosely connected ideas (Roger, 1992).

Unlike Johnson-Laird's concept of mental model, which has been addressed primarily as a temporary structure in short term memory in the process of carrying out an inference, Gentner and Steven's concept of a mental model was mainly examined as a permanent knowledge structure stored in long-term memory (Byran, 1992).

In summary, the concept of mental models has been discussed at two levels in psychology: the knowledge representation level and the inference and decision making level. At the knowledge representation level, a mental model "mirrors" the perceived structure of the external system being modeled (Johnson-Laird, 1983), so that meaning can be derived from actual states of affairs as they exist in the world. At the inference and reasoning level, because mental models are simplified and incomplete versions of reality, running mental models is less cognitively demanding. As a result, people can make inferences about the world efficiently.

It is worth noting that the notion of mental models is a part of a continual theoretical development in attempts to explain the human mind and human behavior (Wilson & Rutherford, 1989). Similar cognitive structures proposed to account for knowledge representation and information processing include schemata, scripts, and frames.

2.1.3 Schemata, Scripts, and Frames

2.1.3.1 Schemata

Schemata are regarded as building blocks of cognition and fundamental elements upon which all information processing, such as perception, comprehension, remembering, learning, and problem solving, depends (Rumelhart, 1980). Rumelhart and Ortony (1977:101) defined schemata (the singular is schema) as:

> data structures for representing the generic concepts stored in memory. They exist for generalized concepts underlying objects, situations, events, sequences of events, actions, and sequences of actions. Schemata are not atomic. A schema contains, as part of its specification, the network of interrelationships that is believed to generally hold among the constituents of the concept in question. Schemata, in some sense, represent stereotypes of these concepts.

A schema has a set of prescribed variables and relationships between the variables. For example, the buy schema involves five variables: purchaser, seller, money, merchandise, and bargaining and the relationships between the five variables. The variables in a schema could have different values in different instances, but the structure of the schema is stable. Schemata could be organized in hierarchical structures or in network structures (Rumelhart, 1980).

Schemata are active (Rumelhart, 1980). People assimilate new information into their knowledge base by fitting the information into an existing schema (top-down activation). When no appropriate schemata exist in a person’s mind, new schemata can be created to accommodate the new information. New schemata can be created by modifying the existing schemata or by induction (pattern recognition) (Rumelhart & Norman, 1978).

When people apply an inappropriate schema to the current situation, the schemata might distort information (Grunig, Ramsey, & Schneider, 1985).

2.1.3.2 Scripts

The concept of scripts was proposed in Artificial Intelligence (AI) for computer simulation of human beings' intellectual activities, such as comprehending semantic meanings of texts, or making decisions in a specific situation. In a script, tasks of understanding are often broken into discrete and serially executed components that are suitable for computer modeling and processing (Schank, 1999). Schank and Abelson (1977) defined that:

> A script is a structure that describes appropriate sequences of events in a particular context. A script is made up of slots and requirements about what can fill those slots. The structure is an interconnected whole, and what is in one slot affects what can be in another. Scripts handle stylized everyday situations. They are not subject to much change, nor do they provide the apparatus for handling totally novel situations. Thus, a script is a predetermined, stereotyped sequence of actions that defines a well-known situation.

Scripts outline high-level background information and procedural information that people would encounter in specific contexts. For example, a restaurant script from a customer point of view includes props (tables, menu, food, check, and money), roles (customer, waiter, cook, cashier, and owner), the conditions of the customer (hungry and has money), and a series of actions (entering, ordering, and eating). The series of actions can be further divided into sub-actions such as attending eyes to tables, choosing food, signaling to waiters, and so on. Thus, scripts encapsulate explicit knowledge so that people can predict particular aspects of a specific context, such as knowing what to do in restaurants. In some script applications, plans and goals are used in parallel with scripts as high-level structures to control understanding (Carbonell, 1979; Wilensky, 1978).

Similar to computers retrieving information from hard disks, people retrieve a script through indexes. A script is called into play by triggering script headers. In the restaurant example, the headers for the restaurant script are concepts related to hunger, restaurants, and so on in the context of planning actions for getting fed (Schank & Abelson, 1977).

2.1.3.3 Frames

A frame is also a data structure that has been proposed for computer simulation of human beings' intellectual activities. Minsky (1975) proposed that, to account for the effectiveness of common sense thoughts, the "chunks" of reasoning, language, memory, and perception must be very structured. Within these chunks, the factual and procedural contents must be "intimately connected in order to explain the power and speed of mental activities" (Minsky, 1975: 211). He proposed the concept of a *frame* to represent these mental "chunks":

> A frame is a data structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed.

The essence of the concept is that when one encounters a new situation, one selects a frame from memory and adapts it to fit reality (Minsky, 1975). A frame consists of a network of nodes and relations. The top levels of a frame are fixed nodes, representing things that are always true about an entity or a situation; whereas the lower level nodes represent features that are less essential. Each node has one or more default values. With these default values, frames are able to fill in additional information that is not explicitly encoded in the current situation to help people generate expectations for the situation (Wilson & Rutherford, 1989). Once a frame is selected to represent a situation, a matching process is initiated to assign values to the nodes of the frame. When the selected frame cannot fit the reality (or the current goal), a replacement frame will be initiated.

In the literature, no clear lines have been drawn between the three concepts: schemata, scripts, and frames. They were used in an interchangeable manner. Rumelhart (1980) pointed out that the three concepts are so closely related that a discussion of any of one of them will serve as an introduction to the others. Also, historically, the terms were closely related. Minsky (1975) has partially credited the idea of *frame* to Bartlett (1932), and Rumelhart (1975) has attributed the term *schema* to Bartlett, as well. Although Schank and Abelson (1977) used the term "scripts", they pointed out simultaneously that such structures have been called "frames" and "schemata".

2.1.4 Mental models and schemata

In order to be a meaningful concept, mental models must not be redundant with the mental structures that have just been reviewed. Wilson and Rutherford (1989) pointed out that the theoretical uniqueness of mental models lies in their computational ability. The computational ability (or runnability) allows mental models to generate meanings dynamically, thus giving mental models superior explanatory power over schemata, scripts, and frames, which only provide related background knowledge for the current situation.

Nevertheless, mental models are closely related to schemata. Most researchers believe that mental models arise from schemata and are the running mode of schemata. Rumelhart (1984) described a mental model as the total set of schemata instantiated at the time. Johnson-Laird (1983) suggested that schemata provide the procedures from which mental models are constructed. Norman and Bobrow (1979) contended that schemata are data structures in memory, whereas mental models are the utilization of such information in a computationally dynamic manner. Brewer (1987) held a similar view.

Despite the effort to differentiate mental models from schemata, it is hard to draw a clear line between them (Wilson & Rutherford, 1989). Johnson-Laird (1983) admitted that the issue of the distinction between mental models and schemata was not strictly resolvable, since the set of all possible schemata and the set of all possible mental models have not been specified.

2.1.5 Mental imagery: the form of mental models

There is a continuous debate, known as the “analogy-propositional” debate, about the nature and form of human beings’ mental experience. The analogy school argued that mental imagery is a distinct, pictorial, and non-language-like form of representation (e.g., Kosslyn, 1994). The propositional school argued that mental representations are non-pictorial; rather, they are detailed descriptions of the reality (Pylyshyn, 1981; 2002).

Later on, some researchers proposed that there are two separate cognitive subsystems, the imagery system specialized for the representation and processing of information concerning nonverbal objects and events, and the verbal system specialized for dealing with language (Paivio, 1986).

Mental models are essentially one type of mental representation, and thus cannot escape the picture-description debate surrounding mental imagery. In mental models studies, some researchers have asserted that mental models may be non-verbal, pictorial, or image-like (Rouse & Morris, 1986); and mental models are spatially arrayed corresponding to their real-life counterparts (Doyle & Ford, 1998). Other researchers have argued that mental models might be propositional representations (Wilson & Rutherford, 1989). Consistent with the dual processing theory, Glenberg and Langston (1992) contended that the elements in mental models point to both propositional and spatial information in long-term memory.

2.1.6 Situated cognition: alternative approaches to mental models

As a form of mental representation, mental models are inevitably challenged by the difficulties associated with the fundamental question of cognitive science: how does the mind work? To use mental models theory as a paradigm for HCI and IR research, it must be assumed that the mind works by constructing a representation of the world, either in language like description, in pictorial like depiction, or in both forms. What if the mind does not work in this way?

Recently, alternative approaches to this symbolic computational theory of cognition, including “dynamical systems theory” and “situated” or “embodied” cognition, have emerged. These alternative approaches call into question the basic assumption of the computational cognitive theory that mental contents are represented in computational units (data structures), and manipulated in the cerebral computer (Pylyshyn, 1999; Thelen & Smith, 1994). Related work on both robotic and human vision also has suggested that perception is best understood as ongoing, directed exploratory activity, rather than as the processing of sensory input into a detailed inner representation (Landy et al., 1996; O’Regan & Noe, 2001).

2.2 Mental models in HCI

2.2.1 Mental models defined in HCI

In cognitive science, a mental model is a knowledge representation that supports cognitive reasoning and decision making. In HCI, the term, mental model, has been defined in the context of other models involved in HCI processes (Norman, 1983):

System *conceptual model*. A conceptual model provides an appropriate representation of the system. It is accurate, consistent, and complete.

System image. The system image is the look and feel of a system and is characterized by the display, the documentation, and system messages. System image is also called the *User Interface (UI) designer’s model of the system*, through which UI designers communicate the system conceptual model to end users (Ehrlich, 1996).

The *users' mental models* of the system. Users' mental models of the system represent how users understand the system and how they perceive the functions of the system.

Scientists' conceptualizations of users' mental models. How researchers perceive and represent users' mental models of the system.

System's model of the user. The system's model of the user is the model constructed inside the system of the user as it runs through different sources of information such as profiles, user settings, logs, and even errors (Fischer, 1991).

Since Norman's denotation, it has been widely accepted that mental models refer to mental representations that users employ to understand and further operate the system (Brehmer, 1987; Conant & Ashby, 1970; Halasz & Moran, 1983; Norman, 1986). Within this scope, mental models have been defined further from different perspectives with different focuses and degrees of specificity. Rouse and Morris (1986: 351) defined that:

> Mental models are the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states.

Another widely recognized definition of mental model was proposed by Carroll and Olson (1988: 51):

> The user's mental model of a system is a rich and elaborate structure, reflecting the user's understanding of what the system contains, how it works, and why it works that way. It can be conceived as knowledge about the system sufficient to permit the user to mentally try out actions before choosing one to execute.

Rouse and Morris's definition of mental models focuses on the functional aspects of mental models, while largely ignoring mental models' role as a knowledge representation structure; whereas Carroll and Olson point out that the mental model is both a knowledge structure and a dynamic instrument that enables users to interact with a system. However, both definitions do not cover some important characteristics of mental models. For example, they do not define the structural features of mental models of computer systems. It has been widely agreed in cognitive psychology that the structure of a mental model "mirrors" the perceived structure of the external system being modeled (Doyle & Ford, 1997).

2.2.2 Types of mental models in HCI

Young (1983) proposed eight tentative views of a mental model of an interactive device: strong analogy, surrogate, mapping, coherence, vocabulary, problem space, psychological grammar, and commonality. The eight viewpoints are distributed on a continuum from assimilation to accommodation. Models at the assimilatory end (e.g., strong analogy models) tend to view the device in terms of its relationships to other systems that have been familiar to the user. At the accommodatory end, the emphasis is more on an understanding of the device in its own right.

Young (1983) elaborated on two kinds of mental models – surrogates and task-action mapping – based on his research on people's use of pocket calculators. A surrogate model is defined as a simulator of the target device, accounting for the device's working mechanism. For example, a surrogate model of a calculator could be a four-stack structure. This structure assimilates the internal structure of the calculator. A surrogate model is only a partial representation of the system's mechanisms. Unlike a surrogate model, a task/action model encompasses a core set of corresponding relationships between tasks and actions. New tasks are expressed as variants of the core tasks. Sequences of actions corresponding to the new tasks are derived from the core actions that correspond to the core tasks.

diSessa (1986) distinguished between structural and functional models in terms of contextual specificity. Structural models contain information about the internal structure of the system and the models are independent of specific tasks. There is only one structural model for the system and it is universally applicable. Functional models contain information about how to use a selected set of functionality to perform a specific task. The functional models are task related and reflect the relationship between goals and means.

diSessa (1986) pointed out that his structural model is similar to Young's surrogate model. Both models focus on representing the structure of the system without considering contextual variables, such as tasks. The isolation from context limits their power to direct users to accomplish certain tasks. diSessa's functional models are similar to Young's task/action model. Both are rule-based models that represent the mapping between tasks and actions to achieve the tasks. The dual structural and functional aspects of mental models are acknowledged by Nielsen (1990) and supported by Marchionini's (1989a) empirical study on students' use of print and electronic encyclopedias.

In addition to structural and functional models, diSessa (1986) further proposed the concept of *distributed models*. Saying that the model is distributed means that it is not organized around a single, coordinating principle or metaphor but instead consists of a number of distinct facts or (causal) explanations about the object it describes. The facts or explanations are accumulated through experience (Nielsen, 1990). These partial or fragmented models serve as foundations for users to rationalize variations of constructs that they encounter while using the system.

Carroll and Olson (1988) divided mental models into three separate categories: surrogate models, metaphors, and “glass box” machines (Duboulay, O’Shea, & Monk, 1981). A surrogate model is identical to the surrogate model proposed by Young (1983). A metaphor model is a direct comparison between the target system and some other systems already known to the user. It is very similar to Young’s strong analogy model. A typical example is “a text editor is a typewriter”. Glass box models lie in between surrogates and metaphors. They are mimics of the target system, offering a semantic basis for understanding the system, as do metaphors.

Rasmussen (1979, 1986) developed a taxonomy of mental models of systems. His taxonomy includes five types of models, moving from concrete to abstract: physical form, physical function, functional structure, abstract function, and functional meaning or purpose. Based on Rasmussen’s taxonomy, Rouse and Morris (1986) proposed that mental models could represent a system purpose (why a system exists), function (how a system operates), state (what a system is doing), or form (what a system looks like).

The exploration of types of mental models can potentially improve our understanding of models people use to reason about systems, the limits on the complexity of the models, and the interplay of multiple models in users' learning and performance (Riley, 1986). However, some researchers have pointed out that classifying an abstract concept like mental models could be artificial and misleading because the models represent the same reality (Johnson-Laird, 1989).

2.2.3 Characteristics of mental models in HCI

Incompleteness is an inherent feature of mental models (Norman, 1983). A mental model's content is only a partial representation of the environment and its scope is limited (Sanderson, 1990). It is worth noting that incompleteness is not necessarily a shortcoming of mental models. Keeping the model to a manageable size grants mental models cognitive feasibility in people's information processing, because human beings' memories and processing capabilities are limited. Many studies provide empirical evidence for the incompleteness of mental models. For example, Makri et al. (2007) found that students developed rudimentary mental models of the digital libraries that they access. Their understanding of the system was limited.

People's mental models are often naïve and not "scientific", in the sense that mental models are often not consistent with the normative conceptual model of the system. When using a system, people tend to speculate about the system's underlying mechanisms based on their own observations. As long as the system behaves as expected, users would assume that their mental models are valid, even though the system's behaviors are not generated by the speculated mechanisms. There is also a tendency for users to establish causal relationships based on co-occurrence of events, even though the co-occurrence might be random and for reasons that differ from the ones believed by the user (Besnard, Greathead, & Baxter, 2004).

Mental models often involve misconceptions or errors (Young, 1983). For example, Bayman and Mayer (1983) found that, in learning a programming language, novice users tended to develop conceptions of the statements that either failed to include the main idea or that included misconceptions. Misconceptions have also been found in people's mental models of the Internet (Papastergiou, 2005) and Web search engines' query processing mechanisms (Muramatsu & Pratt, 2001).

It has been postulated that mental models allow users to predict the status and feedback of an interactive system, as well as to plan methods for novel tasks (Card & Moran, 1986). If this postulation is true, a mental model has to be dynamic, that is, to be able to "run" in response to environmental changes. "Running" a model is a dynamic process of building, trying out, and changing the mental model (Ehrlich, 1996). Being able to "run" is a key feature that distinguishes mental models from other static knowledge structures, such as schemata and scripts.

Although runnability is a critical feature of mental models, people's ability to run a model is severely limited (Norman, 1983). This is reflected in the recurring observation that users have difficulties in recovering from an error (e.g., Moray, 1987). They repeat the same operations even when the system behaves abnormally (Chen & Dhar, 1991). People's limited ability to construct and run a model is also reflected in people's inclination to trade off extra physical actions for less cognitive effort. Sasse (1997), for instance, found that subjects tended to specify their own formulae instead of using system built-ins when they use MS Excel.

Mental models do not have firm boundaries. People's existing mental models of one domain may influence the construction of their mental models of another domain. Norman (1983) observed people who had experience with several different calculators and found that they tended to mix up the features of different calculators. They were often unsure which feature applied to which calculator and had various superstitions about the operation of the calculator. In some cases, the existing models can be easily transferred to a new domain. Marchionini (1989a) found that some students were able to adapt their mental models of the print encyclopedia and further developed distinct mental models for the electronic encyclopedia.

In most studies, mental models were discussed as if they were single entities (Staggers & Norico, 1993). Yet some authors have postulated the possible existence of multiple mental models (e.g., Borgman, 1984; deKleer & Brown, 1983; Moray, 1987). The use of multiple models is affected by the complexity of the task and the complexity of the system/device itself. For example, Williams, Hollan, and Stevens (1983) found that individuals held multiple models for a heat exchanger – both a temperature and a heat flow model. Neither model individually can fully explain the behavior of the device.

2.2.4 Utility of mental models

Research has demonstrated that mental models help people learn a new system and further use the system in a more productive way. Meanwhile, mental models are able to communicate system design parameters to help designers produce more usable and enjoyable computer systems and to facilitate user training (Wahlström, 1988).

2.2.4.1 Learning a new system

Mental models' role in learning and using a new device has been demonstrated by some early experimental studies in HCI. In these studies, participants are usually divided into two groups. One group (model group) receives training that emphasizes the conceptual model of a system. Members in the group are then assumed to develop mental models of the system. The other group (procedure group) receives only procedural knowledge of how to use the system. Performance of the two groups is then compared to illustrate the impact of mental models. For instance, in a study investigating people's use of calculators, Halasz and Moran (1983) found that, although there were no differences between the groups in their accuracy of solving routine or even some complex problems, the conceptual model group did outperform the procedural group when it came to solving problems that require the invention of new methods and more cognitively intense problem-solving.

Kieras and Bovair (1984) came to a similar conclusion in a study exploring the role of mental models in learning to operate a simple control panel device. They found that the model group learned the procedures more quickly, retained the knowledge more accurately, and used it more adaptively than the rote learners (who were only taught the procedures of operating the device). Furthermore, the model group was able to infer the procedures from the conceptual model of the device. To exclude the possibility that the model group outperformed the procedure group because they were given more information, Kieras and Bovair conducted another experimental study, demonstrating that simply giving more information was not a sufficient condition for improving performance. Specific information about the device topology and functioning is essential to support direct inferences about the behavior of the device.

Fein, Olson, and Olson (1993) extended Kieras and Bovair's effort to look at whether a mental model provides benefit to users learning and operating a complex device. They created three groups by imposing three training conditions: rote, explicit model, and full model. Test results showed that both the model groups outperformed the rote group on time spent on tasks, success rate on retrieval tasks, and success rate on transfer tasks.

These empirical results provide support for the benefit of mental models for learning new devices (Ehrlich, 1996). But why do better mental models lead to better performance? In an effort to explain mental models' impact in learning, Brown (1986) argued that, to be able to support high-level cognitive activities, such as reasoning, planning, and coping with new situations, users have to go beyond the simple procedural knowledge to understand the reason for and the interrelationships between the operations that form a procedure. A mental model provides a stable and robust basis for such an understanding. Bibby (1992) echoed Brown's argument by stating that knowledge of the deep structure of a system that a mental model encompasses enables users to reason about the functioning of a device. Thus, mental models serve as a basis for taking sensible actions during the interaction with a system.

2.2.4.2 Improving efficiency of problem solving

Mental models not only help users learn to use a new system, but also help users who have had experience with a system understand the system at a more abstract level and use the system in a more efficient manner. This utility is reflected in a large number of studies on strategic differences between novice and expert users of a system. In these studies, all subjects had experience with the target system, while some had more sophisticated mental models than the others. For example, by interviewing and observing five nursing PhD students using the SPSSX statistical package, Staggers and Norcio (1993) found that novice users have a limited knowledge repository and repertoire of problem solving strategies. They tended to depend on notes when performing tasks. Expert users, however, were well organized and confident. They used trial-and-error strategies to solve emerging problems.

To explore the effects of mental models on users' behaviors when using Excel spreadsheets, Sasse (1997) asked two groups of users with different levels of mathematics and computing background (mental models) to describe and use Excel. The math and computing savvy group (comparison group) described the system at a conceptual level, whereas the other group (main group) gave a purely procedural introduction to the system. The two groups also showed different behaviors; subjects in the main group tended to trade off physical efforts against cognitive efforts by specifying their own formulae instead of using system built-ins.

Dimitroff (1992) conducted a study to examine the relationship between users' mental models of an online library catalog and their success in using the system. She found that the completeness of mental models had a clear impact on users' searching performance. Students with more complete mental models made significantly fewer errors and found significantly more items. Observing public library users using the Web and/or a Web based library catalog, Slone (2002) found that users' mental models affected their search approaches, Web sites visited, and sources used. Users with immature mental models of the Internet relied more heavily on the online catalog and off-line sources.

2.2.4.3 Enhancing a system's usability

Hammond et al. (1983) interviewed software designers and found that, in making certain decisions, designers rely on some "psychological theories" of the user and user performance. Providing them with some type of user model is a better way to improve interface design than context-free design guidelines or task-action analysis methods, such as GOMS, a design tool representing a set of tasks as Goals, Operators, Methods, and Selection rules of a particular system (Card, Moran, & Newell, 1983).

Mental models can also inform the design of new system mechanisms or functions. In their study of game playing, Graham, Zheng, and Gonzalez (2006) demonstrated mental models' potential in improving game playability and usability. In the study, they observed that, when playing a real-time strategy game, a player's first experience is based on the available mental models of the surface characteristics of the environment. As experience increases, they tend to move away from these surface characteristics. Based on this phenomenon, the authors proposed that mental models are a good descriptor of users and mental model shifts might serve as a benchmark for designing game progression.

Weaknesses in users' mental models can guide system developers in coming up with designs that are able to reduce unnecessary human errors. It is well documented that, when an error occurs, users often do not analytically explore the environment and accordingly update their mental models. Rather, they persistently try to fit data to their existing models and reject any information that is not consistent with their expectations (e.g., Norman, 1988). This phenomenon is termed "cognitive lockup" or "confirmation bias". It has been suggested that intelligent display and decision aids can be designed to overcome this shortcoming so as to facilitate users' recovery from errors (Besnard, Greathead, & Baxter, 2004; Moray, 1987).

Although taking users' mental models into consideration during the software design process is theoretically appealing, completely relying on mental models could be misleading in practice (Carroll and Thomas, 1982). In exploring the effectiveness of striving for optimum compatibility with users' initial conceptualization in system design, Wright and Bason (1982) constructed a data analysis software package based on users' specifications of their problems. Meanwhile, they constructed the same system based on system designers' reinterpretation of the problems. They found that bending the system to the preconceptions of the user is not necessarily the most viable approach. The design based on designers' reinterpretation sometimes provided more expedient operational routes that turned out to be acceptable to end users.

2.2.4.4 Facilitating user training

It is very common for users to develop misconceptions about a system and it is often hard for users to jump out of fallacies (e.g., Norman, 1988). In their investigation of programmers' misconceptions of BASIC programming, Bayman and Mayer (1983) concluded that explicit training, including the introduction of a concrete model of computers and key transactions for each statement, is needed to encourage users to develop an appropriate mental model of the behavior of the programming language. For many complex systems, training is necessary for efficient operation, even when the system usability has been achieved. For example, Borgman (1996) proposed that current library online catalog systems can be made more effective through training that provides users with a conceptual framework rather than a set of procedures for searching.

The effectiveness of providing users with a conceptual model of the system for enhancing their performance has been illustrated by many studies (e.g. Halasz & Moran, 1983). Reciprocally, knowledge about users' mental models can be used to inform the design of better training or teaching materials. For example, in their investigation of the use of a phone system, Hanisch, Framer, and Hulin (1991) pointed out that instructional aids and training programs should stress features that are likely to be misperceived by users and take into consideration the discrepancies between novices' and experts' representations of the system.

2.2.4.5 User modeling

User models are a system's representation of the characteristics of its users (Allen, 1997; Fischer, 2001). The process of creating and maintaining an up-to-date user model by a system is known as user modeling (Brusilovsky & Millan, 2007). User modeling allows a system to be adaptive to users' current status in knowledge or preferences so as to create a friendlier computing environment for the user. Different information sources could contribute to user model construction: the user's domain knowledge, interests, goals, background, individual traits (e.g., cognitive styles and learning styles), the user's actions or action patterns (e.g., document reading, saving and printing), and the context of the work (e.g., computer platform and user location) (Kelly & Belkin, 2002; Kobsa, 2001).

The main purpose of user modeling is to make the system more adaptive and more usable to people. Mental models are users' representations of the system. Therefore, users' mental models could be a valuable source to inform the system of the users' current status of knowledge about the system (Streitz, 1988). For example, in their effort to design a more playable and more adaptive strategy game, Graham, Zheng, and Gonzalez (2006) modeled subjects' real-time mental models of the game and further used the mental models to direct the progression of the game.

A major challenge in developing effective adaptive systems is that the adaptive behavior of a system empowered by user models may cause disruptions to the user's mental models of the system, therefore causing difficulties for the user in performing certain tasks. Knowledge about mental models can be employed to reduce such disruptions. For example, Hui, Partridge, and Boutilier (2009) proposed a probabilistic model to assess the amount of disruption of an adaptive system that changing function locations (in order to make access more convenient) has on the user's mental models of the location of those functions.

2.3 Mental models in IR

As an important concept in both psychology and HCI, mental models have been extensively researched in various domains in the ILS field, such as online library catalogs (Borgman, 1986; Dimitroff, 1990; Kerr, 1990; Slone, 2002), experimental retrieval engines (Cool, et al., 1996; Savage-Knepshield, 2001), commercial databases (Katzeff, 1990; Zhang, 1997), a digital encyclopedia (Marchionini, 1989a), digital libraries (Makri, et al., 2007), college websites (Otter & Johnson, 2000), the internet and the web (Bruce, 1999; Kerr, 1990; Slone, 2002; Thatcher & Greyling, 1998; Zhang, 2008a), and web search engines (Efthimiadis & Hendry, 2005; Muramatsu & Pratt, 2001). Westbrook (2006) explored mental models of academic information seeking process held by a group of graduate students in a reference course.

The purpose of studying mental models of IR systems is twofold: (1) to understand the underlying cognitive processes of people's behaviors while using IR systems, and (2) to use this understanding to design more usable systems and more effective instructional materials. Thus, mental models are usually not studied as an isolated concept in the ILS field. Their influence on people's information searching behavior and performance is of particular interest to researchers (e.g., Borgman, 1984). In this section, what we know about mental models of IR systems will be reviewed from various aspects.

2.3.1 Content and structure of mental models

A mental model is an unobservable mental construct, which dynamically changes with the environment. This inherent feature of mental models prevents researchers from illustrating them in a tangible fashion. Thus, the content and structures of mental models are often inferred from subjects' verbal accounts, drawings, ratings on related concepts, or behaviors. The validity of these approaches is not conclusive; nevertheless, they provide useful insights to enhance our understanding of this fundamental issue in mental model research.

In a study of undergraduate students' mental models of the web as an IR system, Zhang (2008b) identified that the mental models have four components: building elements (information, technologies, and people), functions (e.g., information access and shopping), attributes of the Web (e.g., infinite), and feelings toward the web (positive, negative, and neutral). Thatcher and Greyling (1998) found that people's mental models of the internet include elements like computers, modem, servers, webpages, various information subjects (e.g, psychology and biology), various functions supported by the Internet or the Web (e.g., shopping and entertainment), types of information (e.g., news, maps, and movies), end users, search engines, and so on. The structures of mental models often appeared as linear communications between computers or users, hierarchical organization of computers or information resources, or networked connections among servers, computers, users, and information sources. In some cases, mental models only included a collection of functions or information sources, without any structural organization.

In studying subjects' use of an experiment IR system, Savage-Knepshield (2001) observed that subjects' mental models of information retrieval is an integration of schemata and scripts. Schemata included task specific and system specific variables, such as ranked results, information seeking strategies, and goals. Also included were fixed IR components, such as a computer, search engine, query box, and retrieved results list. For example, one subject exhibited use of the following schema: search topic (Alzheimer's disease), results (relevant), data (new, bad terms, good terms), and the ability to reformulate a query. When schemata are organized into "a temporally organized, well-structured sequence of events", they form an instantiation of a script (Savage-Knepshield, 2001: 130).

Similar to other areas, researchers in ILS often construct a conceptual or an expert model of a system and evaluate users' mental models against the predefined conceptual models. For example, Dimitroff (1992) constructed an eight-component scale based on system documentation and her own experience: contents of the database, interactive nature of the system, multiple files, multiple fields within each record, multiple indexes and/or inverted indexes, Boolean search capability, keyword search capability, and use of controlled vocabulary. The completeness of users' mental models of the system was measured against the scale. Zhang (1997) arrived at nine essential concepts and three attributes of IR systems by consulting a group of experts. The nine concepts -- browsing, classification, data structure, document content, feedback, information need, interface, query, and search -- cover important components of an IR system. The three attributes: format/process, targeted/untargeted, and specific to IRs/applicable to all information systems, are three dimensions on which the concepts were judged. Subjects' mental models of the testing system were then represented by their ratings on the concepts and attributes.

Efthimiadis and Hendry (2005) created a conceptual model for search as a benchmark for analyzing subjects' mental models of web search engines. The model divided search into three phases with various processing components embedded in each phase:

Indexing (Components: content, spidering/crawling, parsing, inverted index creation, link analysis, and storage)

Searching (Components: user, user needs, query, and results)

Matching (Components: query processing, matching, accessing inverted file, and ranking).

Users' drawings of their perceptions about web search engines were evaluated against this predefined scheme.

The assumption underlying this indirect approach is that users' mental models of the IR system share both content and structure with the conceptual model. It is worth noting that this assumption is not necessarily correct. End users' mental models might contain content that is not included or reflected in the predefined "normative" model. Users' mental models may also exhibit different structures from the "normative" model. Evaluating users' mental models against the predefined norms might miss some unique content or structures in subjects' mental models.

2.3.2 Factors affecting mental models of IR systems

2.3.2.1 Individual differences

Users come to a particular system with different mental models and assumptions (Cooper, 1995; Marchionini, 1995; Tognazzini, 1992). These preexisting knowledge structures have significant effects on their mental models of an IR system. In their study of lostness in a hypertext system, Otter and Johnson (2000) identified a significant correlation between the accuracy of subjects' mental models and their level of familiarity with hypertext systems. Papastergiou (2005) found that high school students who had been taught about the internet at school had significantly better mental models than those who had not; and students who had used the internet at home had significantly better mental models than those who had not. Thatcher and Greyling (1998) also found that experienced users hold more complete and detailed mental models of the internet than inexperienced users. Nevertheless, it is worth pointing out that frequent use of a system is a necessary rather than a sufficient condition for detailed and complete models.

Other individual differences have effects on mental models, as well. Zhang (1997) found that users' educational status and academic background were related to users' mental models of an IR system. In studying users' mental models of the web, Zhang (2008a) found that male subjects were more likely to have a technical view than female subjects, and conversely, female subjects were more likely to develop a process view of the web than male subjects. More research is needed to explore the effects of individual differences, such as age, learning style, and spatial ability, on users' mental models of IR systems.

2.3.2.2 Environmental factors

Empirical studies showed that providing users with an explicit conceptual model of an IR system would enhance their ability to construct mental models of the system. For example, Savage-Knepshield (2001) found that subjects who were exposed to explanations of how the experimental IR system's features operated were more likely to create representations of the new features in their mental models. However, it is worth pointing out that training is not a necessary condition for users to develop mental models of IR systems. Neumann and Ignacio (1998) observed that novice users of digital library systems learn to use the systems in a systematic and planned manner and they form their mental models of different interfaces by structured trial and error. Neither is training a sufficient condition for developing a mental model. Some subjects who had received training were not able to develop an adequate mental model for interacting with an experimental IR system (Savage-Knepshield, 2001).

2.3.2.3 System images

Clear system feedback sometimes is more crucial than providing users with a conceptual model of the system (Norman, 1983; Savage-Knepshield, 2001). Muramatsu and Pratt (2001) reported that when subjects are provided with visible feedback about the opaque query transformation process of web search engines, it was easier for them to construct mental models of search engines in terms of query processing. Savage-Knepshield (2001) observed that sometimes subjects' mental models degrade over time when they are using a system. She attributed this mental models degradation to the unintuitive feedback that the system provides in response to the user's actions.

2.3.3 Construction of mental models of IR systems

Savage-Knepshield (2001) described the formation of mental models based on previous knowledge as a top-down process: users come to the system with preexisting knowledge about IR. When using the system to conduct a specific task, their preexisting mental model for comparable tasks was elicited for modification. During the interaction with the system, they gradually added task-specific information and specific system features to their mental models of IR.

There are generally three patterns for people to adapt to a new IR system. The first was fitting new systems to old mental models. Users exhibiting this pattern made minimal use of new system features and tried to apply routine search behaviors in the new environment. The second pattern was combining old and new models of systems. Users with this pattern made partial use of the new system features, representing an incomplete model of the new system. The third pattern was effectively using new features provided by the new system. Users with this pattern were able to interact with the new system in ways supporting the formation of an accurate model of the system. This type of user tended to adapt to the new system environment more quickly (Cool et al., 1996; Marchionini, 1989a).

Katzeff (1990) provided an account of mental models construction process by analyzing subjects' think-aloud protocols during their interactions with a commercial database. Three phrases/states of the mental model construction process were identified: construction, testing, and running. The three states do not necessarily occur sequentially. In the construction phase, the system's feedback in response to the user's input is interpreted by the user and then incorporated into his/her mental model of the system. The mental model formed by the user is then tested in the user's further interactions with the system. If the model cannot explain the system feedback satisfactorily, the model will be rejected. At the running phase, the model is used to predict system feedback. Meanwhile, the model becomes sticky and users tend to interpret outcomes to suit the model.

2.3.4 Characteristics of mental models of IR systems

The general characteristics of mental models reported in HCI studies are also observed in people's mental models of IR systems. People tended to have incomplete mental models of an IR system (e.g., Dimitroff, 1991). Mental models are often simple, sometimes rudimentary (e.g., Papastergiou, 2005; Makri, et al., 2007). Mental models are not necessarily correct. They sometimes include misunderstandings of system relations or system functioning (e.g., Brandt & Uden, 2003). Even so, mental models are important mechanisms that help people navigate through IR systems, predict system behavior, and complete information searching tasks (Ehrlich, 1996).

Savage-Knepshield (2001) observed that subjects relied on their mental models to decide what their next steps would be, given a set of possible system responses. This observation confirmed Johnson-Laird’s conjecture that mental models resulting from perception and comprehension are the basis of thinking and reasoning. She also pointed out that mental models are neither inherently metaphorical nor do they include mental imagery (i.e., mental models are not visualizations of the system), based on the analysis of subjects’ verbal accounts of how the experimental IR system operated. Her observations also did not provide evidence for the existence of multiple mental models, a notion suggesting that people might need multiple different mental models to fully understand a complex device (deKleer & Brown, 1984; Johnson-Laird, 1983; Moray, 1987; Williams, Holand, & Stevens, 1983).

Mental models might have an affective dimension. In investigating users’ mental models of the web, Zhang (2008b) found that subjects inevitably expressed their feelings when they mentioned certain elements of the web. Bruce (1999) also found that mental models were able to reflect users' satisfaction with using the Internet: a conceptualization of the Internet as an information store or library was more likely to arouse higher levels of satisfaction with the Internet than the conceptualization of the Internet as connectivity and interconnectedness. More empirical studies are needed to explore the affective dimension of mental models.

2.3.5 Mental models' effects on information searching behavior

Mental models have significant effects on users' information searching behavior and performance. Better mental models often lead to better task performance. Kerr (1990) assessed 99 students using system cues (textual, graphic, color) to navigate through a videotext information system and concluded that the presence or absence of physical cues was less important to successful searching than the user's ability to represent internally the structure of the information. In this study, users with more detailed and complete impressions of the database searched faster. Dimitroff (1992) found that students with more complete mental models of an OPAC system made significantly fewer errors and found significantly more items.

However, the evidence for mental models' positive effect on users' performance is not conclusive. Savage-Knepshield (2001) found that, although subjects who possessed higher congruency in their mental models achieved better performance in recall, precision, and number of documents saved than those who possessed mental models with lower congruency, the differences were not statistically significant. Zhang (2008a) also did not find statistically significant differences between groups with different mental models of the web (i.e., technical, functional, process, and connection view) on starting an interaction, navigation, query construction, and search patterns.

Borgman (1986) reported an interaction between mental models (imposed by training conditions) and the type of task performed. The model-based training led to improved performance on complex tasks that required some problem solving and creativity; whereas on routine or simple tasks, model-trained subjects' performance was equal or inferior to non-model-trained subjects. She speculated that it may not be necessary to invoke the model for simple tasks and the model-based training provides no advantage for simple tasks. The same interactive effect between mental models and task complexity was found in people's use of calculators (Halazs & Moran, 1983).

Mental models also affect users' selection of information resources and use of new system features. In exploring users' web searching patterns, Slone (2002) found that mental model was one of the factors that determined users' search approaches, web sites visited, and sources used. Users with immature mental models relied more heavily on the web online catalog or off-line sources. Cool et al. (1996) observed two searchers with similar routine search behaviors using a new IR system. They found that the person with a stronger mental model of searching used more new features in more exploratory ways than the person with a less well formed mental model.

2.3.6 Mental models and IR system design

A system that matches users' mental models or can explicitly assist users in building mental models would be easy to learn and easy to use. Thus, knowing more about mental models can potentially help the design of systems with good usability. However, how to transfer knowledge of mental models to IR system design in a systematic manner remains a challenge facing researchers and practitioners alike.

The design guidelines that have been provided by mental models research are often very general. In empirical studies, researchers often concluded that providing users with appropriate clues is of central importance for system design. For example, Dimitroff (1992) suggested that systems should be able to provide appropriate cues to help users develop a usable mental model so that users' difficulties with subject search can be reduced. Waern (1985) suggested that designers need to make sure that the relationships between goals and methods are consistent with users' prior task knowledge. Unfortunately, Dimitroff did not specify which cues to incorporate in the system design and Waern did not specify what aspects of task knowledge are relevant. With more specificity, Savage-Knepshield (2001) pointed out that providing users with visible feedback in response to their actions (i.e., to make explicit the causality relationship between users' actions and system feedback) and make some encapsulated processes transparent would help users develop accurate mental models of the system. However, this proposal is still very general.

Nevertheless, some studies provided specific suggestions that correspond to specific systems. In studying high school students using an encyclopedia in both print and electronic form, Marchionini (1989a) found that the efficiency of mental model building depends on the level of detail transferred. The students who transferred only content knowledge rather than details of organization and access were quicker on building unique mental models for the electronic system. Based on this result, he suggested that selection and use of metaphors for new systems are important for system designers and instructors; metaphors that highlight the similarities between traditional and electronic systems must be augmented by instructions that focus on the unique characteristics of the electronic system. In observing users using a commercial news articles database, Katzeff (1990) found that the most salient cues of the database were concerned with the order in which articles and pieces of articles were presented. These cues were not clearly presented in the system. To improve the usability of the system, the presentation of search results needs to be improved.

2.3.7 Summary

Mental models of IR systems have been studied for several decades. Significant attention has been focused on the impact of mental models on information searching behavior and little effort has been dedicated to exploring the working mechanisms of mental models themselves. To gain a deeper understanding of mental models and their impact on IR, more studies are needed to explore fundamental questions such as what elements are included in people's mental models of IR and what factors impact the construction of mental models. Answers to these questions will help develop a common theoretical ground based on which comparisons can be made across different studies. Further substantial development on studies of mental models of IR hinges on this development.

IR is a practical field. Mental models are studied mainly to inform the design of better IR systems and to improve the effectiveness of user instruction. Giving users well-organized and conceptually-based instruction have been proven to be helpful for the development of more accurate and more complete mental models of IR systems. However, on the system design side, no systematic design principles, guidelines, or tools have been produced based on mental model research. Similar to the broader field of HCI, how to apply empirical findings in the practice of IR system design remains a big challenge for IS researchers.

2.4 Methods for studying mental models

2.4.1 Mental models elicitation methods

Mental models are abstract and, as a result, difficult to measure. The most used methods are interviews and think-aloud protocols, drawing, and naturalistic observation. A host of other techniques have also been used to elicit and represent mental models, such as repertory grid technique, concept listing, pair-wise rating, and concept mapping.

2.4.1.1 Interviews and think-aloud protocols

Researchers have employed both direct and indirect probing strategies in interviews. Direct probing asks participants to describe the system or how the system works (e.g. Borgman, 1984; Dimitroff, 1992; Muramatsu & Pratt, 2001; Slone, 2002). For example, Slone (2002) asked subjects the question, “Can you describe the Internet to me?”. Indirect probing asks subjects to provide analogies or metaphors to the system under study (Bruce, 1999; Sasse, 1997). For example, Bruce (1999) asked subjects to articulate their conceptualization of the Internet by completing the sentence “Internet is like a…”. The indirect approach assumes that mental models are transferable from one system to another. In both cases, answers to the interview questions were transcribed and content analyzed to represent subjects’ mental models.

Both probing strategies are unstructured, imposing little restriction on users’ articulation of their mental models. Another form of interviews, semi-structured interviews, provide a loose framework for users to express their mental models, while allowing subjects to express ideas in any way they want. Zhang (2008b) generated a set of interview questions intended to probe into four aspects of subjects’ mental models of the web (i.e., information sources, information organization, search mechanism, and interface). Subjects expressed their thoughts in the framework established by the interview questions. However, the researchers could adjust the order of the questions based on subjects’ responses so that subjects were able to talk freely and express any ideas that came into their minds. New dimensions, such as affective aspects of mental models, were able to emerge from the interview. Mental models also can be elicited by structured interviews. Sharit, et al. (2008) designed 10 interview questions to access subjects' mental models of the internet, web browsers, and search engines. For every subject, the questions were asked in the same order and the responses to each question were scored immediately by the researcher.

Think-aloud protocols are another frequently used method to get people's verbal accounts of their mental models (e.g. Clement, 1983; Katzeff, 1990; Makri, et al, 2007). For example, Katzeff (1990) asked subjects to think aloud when they performed search tasks in a news database. The subjects' mental models were constructed by analyzing the reasoning process reflected in the think-aloud comments. An example of a subject testing an incorrect mental model and finally forming a correct model is shown by the transcript: "If I write "next"… I will probably get the same article… now we'll see … No, it is the next article … I should have done this "expand" in between in order to stay within the same article…".

A variant of a think-aloud protocol, the teaching back technique, also has been used in mental models study. In a teaching back scenario, participants are asked to teach another person how to use the system by verbalizing their knowledge about the system (Sasse, 1991). The most frequently asked questions are of two types: "what is?" and "how to?" (Van der Veer & Puerta Melguizo, 2001). Sasse (1997) used this approach to elicit subjects' mental models of a spreadsheet application.

2.4.1.2 Drawing

Drawing is a primitive form of communication. Gray (1990) pointed out that drawings can be used to elicit and illustrate structural aspects of users' mental models. Denham (1993) used drawing as a means to examine children's conceptions of computers. To investigate how people conceptualize Web search engines, Efthimiadis and Hendry (2005) asked 279 students at the University of Washington to draw sketches of how an Internet search engine works. Zhang (2008a) asked 44 undergraduate students to draw their perceptions about the Web. In a proposed framework for IR systems evaluation and design, Pejtersen and Fidel (1998) pointed out that questions like, "Please draw a diagram or picture of it," can help evoke mental models of the Internet.

Empirical research also suggests that drawing is an effective method to represent mental models (Kerr, 1990). Thatcher and Greyling (1998) used drawing to elicit and categorize people's conceptualizations of the Internet. Six categories of mental models were derived: interface and utilitarian functionality, central database, user to the world, simple connectivity, simple modularity, and networking. The categories were found to be significantly correlated with the subjects' experience with the Internet. Papastergiou (2005) collected sketches from 340 high school students in Greece. Similar categories of mental models, from utilitarian to structural ones, were identified. Significant correlations were found between students' drawings and their answers to a set of internet related questions.

A variant of drawing is concept mapping. Concept mapping has been widely used to elicit people's cognitive structure and mental models in psychology and education research. For example, Chang (2007) used the concept mapping technique to externalize students' mental models regarding the homeostasis of blood sugar and concluded that concept mapping is able to differentiate students' mental models and it is a workable method for representing mental models of complex and abstract concepts. However, there is a paucity of research using the method to represent users' mental models of information systems.

2.4.1.3 Observation

Observing users' interactions with the system when they conduct searching tasks is another commonly used method to study mental models. This type of study usually analyzes errors, command patterns, or behavior patterns as a means of gaining insights into the mental models of the subjects. For example, Huang (1992) reported a study exploring subjects' pause behavior when searching the DIALOG system to shed light on subjects' mental models (although mental models was not the main theme of the research). Chen and Dhar (1990) identified users' misconceptions of information retrieval systems by observing 30 subjects performing searches in an online library catalog. By this means, gaps in subjects' mental models of the system were identified.

Behavioral data gathered by observation are more objective and more reliable than verbal accounts gathered from the user (Norman, 1983). However, the conceptualization derived from observation does not reveal much information about a user's cognitive model: the researcher will see what subjects did, but know nothing about why they did it (Sasse, 1991). Researchers have to construct users' reasoning processes (mental models) based on the observed behaviors. Making accurate inferences about users' mental models from the behavioral data is challenging (Marchionini, 1989a). Researchers with different backgrounds might have different interpretations of the same sequence of behaviors.

Because of the inherent constraints of this method, there are very few studies utilizing the single observation method to represent mental models. Rather, observation is used to complement verbal accounts (think-aloud protocols and interviews) in most of the mental model studies (e.g. Borgman,1989; Katzeff, 1990; Kerr, 1990; Makri et al., 2007; Marchionini, 1989a; Muramatsu & Pratt, 2001).

2.4.1.4 Other methods: repertory grid technique (RGT), concept listing, and pair-wise rating.

The repertory grid technique (RGT) generates a list of elements and a list of constructs based on the elements. An *element* is defined as things or events under investigation, such as books or search engines. A *construct* is an attribute of an element. It is a bipolar dimension, where each pole represents the extreme of a particular view or observation of an element. For example, "has cache feature" and "no cache feature" consist of the two ends of a construct to evaluate Web search engines (Crudge & Johnson, 2004).

Zhang (1997) employed the RGT to represent users' mental models of an IR system. In his study, nine concepts (elements) provided by experts were used to represent components of IR systems and three attributes of the concepts (constructs) represented properties of those components. The nine concepts were browsing, classification, data structure, document content, feedback, information needs, interface, query, and search. The three attributes were three dimensions on which concepts were judged: form/process, targeted/untargeted, and specific to IR systems/applicable to all information systems. Mental models thus were represented and measured by subjects' rating on nine concepts on the dimensions represented by the three attributes.

Wang et al. (2004) used the concept listing method to explore the development of students’ knowledge structure of a subject domain over a semester. In the study, a group of students enrolled in an information organization class were asked to list terms related to the class subject during a timed session over a semester. They found that the number of terms and concepts increased over the semester and the quality of the vocabulary also increased. More importantly, terminologies listed by some subjects formed distinct clusters on particular topics in information organization.

In the pair-wise ratings method, a predefined set of central concepts of a domain will be created and participants will be asked to rate the similarity or relatedness for each possible pair of concepts in the total concept pool. These ratings are transformed into a matrix, which is then analyzed by Pathfinder, a graph theoretic technique that derives network structures from rated data (Schvaneveldt, 1990). The concepts set can be created by domain experts or derived from system documents. The pair-wise rating method has been used to elicit subjects’ mental models of mobile phone networks (Langan-Fox et al., 2006; Hanisch, Framer, & Hulin, 1991).

2.4.1.5 Summary

Each technique has its limitations in representing mental models. Verbal accounts depend on people's ability to articulate their mental models and mental models inferred from the information retrieved from memory do not necessarily reflect the way that the information is stored in memory (Cacioppo, von Hippel, & Ernst, 1997). Drawing and concept mapping allow subjects to present their thoughts in an integrated manner, but the interpretation of drawings or concept maps is subject to researchers' understanding. Meanwhile, drawings are not effective in representing abstract concepts (Zhang, 2008b). Observation allows researchers to see the real behavior of users, but inferring the cognitive mechanism underlying the behavior is very subjective. More structured methods, such as structured interviews, RGT, pair-wise ratings, and concept mapping, restrain the emergence of dimensions and might misrepresent mental models.

In fact, researchers often employ multiple methods in combination to explore mental models. Muramatsu and Pratt (2001) used comments from post-session interviews, along with video-taped search sessions, to infer subjects' mental models of web search engines' query processing mechanisms. Zhang (2008b) combined data from drawings, drawing descriptions, and interviews to represent subjects' mental models of the web. It was found that data collected by drawing and by interview methods supplemented each other and, together, they provided a more holistic representation of mental models.

2.4.2 Experimental studies of mental models

In mental model studies, the experimental approach does not elicit and capture descriptions of mental models; rather, it is employed to solve a different set of questions. For example, experiments can be used to understand whether subjects trained by the conceptual model of a system perform better than subjects who did not receive the same training; whether one form of instructional material is better than the others on cultivating correct mental models; and whether a particular mental model of a system is able to better support learning, problem solving, or other kinds of reasoning about the system (Bibby, 1992).

A typical research design can be illustrated by Kieras and Bovair (1984), one of the earliest experimental investigations in mental models. The study consisted of three experiments. In the first experiment, one group of subjects learned a set of operating procedures for a device by rote, and the other group was presented the device model (in the form of a diagram) before receiving the identical procedure training. Subjects were required to conduct a set of tasks after the training. It was found that the device model group learned the procedures faster, retained them more accurately, executed them faster, and was able to simplify inefficient procedures more often than the rote group. In the second experiment, subjects were required to think-aloud when they used the device to accomplish a task. The think-aloud protocol revealed that the model group was able to infer the procedures much more easily than the rote group, which led to more rapid learning and better recall performance. The third experiment was designed to investigate which type of knowledge/information accounts for better performance and better retention of the operating procedure information. Four groups of subjects were formed based on the presence or absence of two factors: fantasy context and specific control information. The study demonstrated that the important content of the device model was the information about the functioning of the components and the device topology, and not the motivational aspects, component descriptions, or general principles.

In experimental studies different mental model groups were often created by imposing different training conditions. The validity of this approach is uncertain because it is based on two questionable assumptions. The first is that users in the model-instruction groups will develop mental models of the system, whereas users in the procedure-instruction will have no such mental models. This assumption is problematic because studies have shown that some users are able to construct a model of a system in the absence of any instruction at all (Neumann & Ignacio, 1998; Shrager & Klahr, 1983). The second assumption is that the internally represented product of the instruction is synonymous with a mental model (Bibby, 1992). This second assumption is also problematic because instruction is not the only possible source for mental model construction. System image and feedback also could be sources for mental model development (Norman, 1983).

Experimental methods are also limited in providing information for understanding fundamental aspects of mental models, such as their form and development. Thus, as research advances toward the goal of eliciting information to provide insights into mental models' form, content and construction process, classical experiments are recommended to be used in combination with other methods (Rutherford & Wilson, 1992).

2.5 Task complexity and spatial ability

Two important experimental variables are introduced in this section. Task may be the single most important factor in information behavior research. This study explored the impacts of task complexity on mental model construction. Thus, the literature on task complexity is reviewed in this section. In the existing literature, researchers have often argued that people with higher spatial ability are able to develop a better mental model of an information system and, therefore, achieve better performance. Thus, subjects' spatial ability was controlled in this study and spatial ability in information behavior studies is briefly reviewed in this section.

2.5.1 Task complexity

2.5.1.1 Search tasks

Information seeking or searching actions are initiated to fulfill individuals' information needs. Tasks reflect people's information needs (Belkin et al., 1982; Ingwersen, 1992; Ingwersen & Järvelin, 2005). Thus information related behaviors need to be investigated and explained within the context of tasks (Byström & Hansen, 2005; Wildemuth & Hughes, 2005; Vakkari, 2003). Tasks also play an essential role in designing system functions and assessing system usability (Nielsen, 1989). Information systems are designed to serve specific purposes by supporting certain tasks or subtasks. Task analysis is an important analytic tool for system design. GOMS, for instance, a widely recognized cognitive model to inform system design in HCI, functions by representing a set of tasks as Goals, Operators, Methods, and Selection rules (Card, Moran, & Newell, 1983).

Tasks have different levels of granularity. In some studies, tasks are conducted to meet an overarching goal, such as writing a research proposal. The information searching process to accomplish this type of tasks is often very complex. A number of different subtasks, such as selecting the topic, articulating the problem, selecting sources, or gathering information, can occur (Allen, 1996). Research questions concerning this type of task are preferably addressed by longitudinal studies.

Another type of task is often associated with specific communication media or information systems (Allen, 1996), and they are conducted to fulfill specific search goals. These tasks are often termed search tasks (Ingwersen, 2005; Vakkari, 2003; Wildemuth & Hughes, 2005). Search tasks could be natural search goals generated by subjects. The fully self generated tasks reflect searchers' real information needs and maintain the context for information searching behavior. Search tasks also could be assigned by researchers. Assigned tasks do not reflect the information needs of the subject, but they provide a useful means for researchers to control the effects of the tasks on search performance (Hancock-Beaulieu, et al., 1996). In most IR experiments, search tasks are assigned tasks. In order to sustain the merits of both types of tasks, simulated search tasks have been proposed (Borlund, 2000; Borlund & Ingwersen, 1997). In such tasks, researchers fabricate a realistic (though not real) scenario that may lead to information searching. Simulated search tasks allow individuals to interpret the situation and choose aspects of interest to them, as in real life tasks (Bilal, 2002; Vakkari, 2003).

2.5.1.2 Task complexity

Search tasks can be categorized based on different criteria in information behavior research, such as general vs. specific tasks (e.g., Qiu, 1993), topical vs. factual tasks (e.g., Kim, 2000), research vs. fact based tasks (Bilal, 2001), known item search vs. subject search (Kim & Allen, 2002), open vs. closed tasks (e.g., Marchionini, 1989b), and simple vs. complex tasks (e.g., Borgman, 1986; White, Ruthven, & Jose, 2005). Among these classifications, task complexity is the one that has been most implemented and discussed in the literature.

Different authors have used the term task complexity to refer to different constructs. Campbell (1988) defined task complexity based on four task related characteristics: multiple potential paths to a desired end-result, the presence of multiple desired outcomes, the presence of conflicting interdependence among paths to multiple outcomes, and uncertainty regarding paths. Rasmussen, Pejtersen, and Schmidt (1990) defined the complexity of a task based on the amount of information to be considered, the number of goals to be fulfilled, and the coupling of goals and contextual constraints. Byström and Järvelin (1995) defined task complexity as the predeterminability of information requirements (what information is required), processes (how to find the information), and output of the task (how to assess relevance). In a less complex task, the types of task results, the associated work processes, and the output of the task are known by the searcher in advance; whereas in a complex task, none of the aspects are determined in advance. Marchionini (2006) defined the complexity of a task in terms of its cognitive demand. Based on the cognitive activities involved in tackling a task, he classified tasks as lookup tasks and exploratory tasks. Lookup tasks often have a definite answer and are less cognitively demanding, whereas exploratory tasks have less definite answers and require more complex cognitive processes such as analyzing, synthesizing, and evaluating information. Due to a higher demand on cognitive resources, exploratory tasks are more complex than lookup ones.

Task complexity is also a psychological experience (Campbell, 1998). Personal factors, such as the subjects' prior knowledge and their search strategies, can affect their assessment of the complexity of a task (Vakkari, 1999). Bell and Ruthven (2004) considered task complexity as a measure of people's uncertainty with a search task and pointed out that task complexity is a dynamic construct. It can be amplified or reduced by factors such as the searcher's interest in the topic.

It is apparent that task complexity is a multidimensional construct. In empirical IR studies, different authors have characterized task complexity in different ways. Bilal (2000, 2001) used the number of facets involved in and the cognitive demand of a task to denote task complexity. Complex research tasks had more facets and required more critical thinking than fact-based simple tasks. Bell and Ruthven (2004) implemented tasks with three levels of complexity based on criteria proposed by Byström and Järvelin (1995): subjects' pre-knowledge about information requirements, search processes, and outcome of the task. White, Ruthven, and Jose (2005) determined task complexity by the number of potential information sources and types of information required. Capra et al. (2007) created three types of tasks with increasing levels of complexity based on three factors: the number of facets to be combined to get the result, the extent to which higher level thinking is required, and the navigation path to the result page.

2.5.1.3 Effects of task complexity on mental models

Characteristics of tasks have effects on mental models (van der Velden & Arnold, 1992). Savage-Knepshield (2001) has investigated the effects of task combinations on mental models. She found that subjects who performed the same task over two different trials (one week apart) did not increase their mental models on accuracy and completeness. However, improvement on mental models' congruency was observed on subjects who performed different tasks in the two trials.

Mental models are dynamic. They are developed, validated, and modified during people's interactions with systems as they complete particular tasks. These tasks provide a context for people's information behavior, as well as for the construction of mental models. In other words, the development of a mental model is, at least, partially embedded in the context of the current task. Certain aspects of the task will inevitably affect the development of mental models. However, these effects are heavily underresearched. Compared to the large amount of research on mental models' effects on subjects' performance on different types of tasks (e.g., Borgman, 1986), the research on tasks' effects on mental models is sparse. Given the significant role that tasks play in information searching and task performance (e.g., Byström & Järvelin, 1995; Kellar, Watters, & Shepherd, 2007; Solomon, 2002), it is worthwhile to explore the effects of task complexity on people's mental models of the system.

2.5.2 Spatial ability

2.5.2.1 Spatial ability, information searching behavior, and mental models

Spatial ability is the ability to generate, retain, and manipulate abstract visual images (Lohman, 1979). It is the cognitive ability that is most frequently reported to be associated with information retrieval (Westerman, Collins, & Cribbin, 2005). Studies showed that people with higher spatial ability often took less time to complete their information browsing or information searching tasks (Campagnoni & Ehrlich, 1989; Chen & Rada, 1996; Pak, Rogers, & Fisk, 2006; Stanney & Salvendy, 1995; Vicente & Williges, 1988; Vicente, Hayes, & Williges, 1987), and were less likely to get lost in navigation in traditional 2D interfaces (Campagnoni & Ehrlich, 1989). Similar results were observed in 2.5D and 3D interfaces. Subjects with higher spatial ability made fewer navigation errors, traveled less distance, and completed more experimental trials within the same time limit in

2.5- and 3D interfaces (Modjeska & Chignell, 2003; Czerwinski & Larson, 1997). High spatial ability subjects were also more effective in using direct manipulation UIs (Swan & Allan, 1998), and more active in exploring categories represented on the interface (Czerwinski & Larson, 1997).

People construct and employ mental models to direct their interaction with systems (e.g., Norman, 1983; Carroll & Olson, 1987). It is very possible that spatial ability affects people's information searching behavior and performance through its impact on mental models (Chen, 2000; Dillon, 2000). It is likely that people with higher spatial ability are better at constructing mental models of an information system and using the models to direct their navigation and, therefore, are able to achieve better performance (e.g., Campagnoni & Ehrlich, 1989; Sein et al., 1993). Given the possible significant effects of spatial ability on mental model construction, this factor will be measured and controlled in the study to ensure that the comparison on variables (such as mental model construction) is performed between subjects with comparable spatial ability.

2.5.2.2 Measurement of spatial abilities in information behavior studies

Spatial ability is a multifaceted construct. The three basic spatial ability factors are spatial relations, spatial orientation, and visualization (Carroll, 1993). In the context of exploring people's use of various interfaces, the most measured factor is *spatial visualization*: the ability to manipulate or transform the image of spatial patterns, into other arrangements (Ekstrom, et al., 1976: 173). Another factor of spatial ability, *spatial* *orientation*, refers to the ability to perceive spatial patterns or to maintain orientation with respect to objects in space (Ekstrom, et al., 1976: 149), and has been occasionally measured, along with spatial visualization, using the ETS Cube Comparison Test (Pak, Rogers, & Fisk, 2006; Stanney & Salvendy; 1995).

It is worth noting that, in the existing literature, the terminology concerned with spatial ability is not used in a consistent manner. The general term, spatial ability, is often used to represent either of the two subfactors, spatial visualization or spatial orientation. Meanwhile, the term spatial visualization is often used in an interchangeable manner with terms such as structural visualization, spatial reasoning ability (e.g., Swan and Allan, 1998), and visualization ability (e.g., Sein, et al., 1993). In this study, spatial ability refers to spatial visualization ability.

In most studies, spatial visualization ability was measured by the ETS VZ-2 paper folding test (Campagnoni & Ehrlich; 1989; Chen, 2000; Chen & Czerwinski, 1997; Pak, Rogers, & Fisk, 2006; Sein et al., 1993; Swan & Allan, 1998; Westerman, 1998; Westerman, Collins, & Cribbin, 2005). The VZ-2 test has been validated across a wide variety of samples and has consistently demonstrated high test-retest reliability (0.84 reported by Eckstrom, et al., 1976). This test also has been administered in an array of information searching studies. Significant effects of spatial ability as measured by the VZ-2 test have been found on people's information retrieval performance on various information retrieval tasks (e.g., Campagnoni & Ehrlich, 1989).

A couple of studies used measurements other than VZ-2 to measure subjects' spatial ability. Modjeska and Chignell (2003) employed the Minnesota Paper Form Board Test produced by the Psychological Corporation to measure subjects' spatial ability. Dahlbäck, Höök, and Sjölinder (1996) tested spatial ability with three different tests (rotation of images, left or right hand identification, and Kohs' blocks test) from the Düremann-Sälde test battery, a Swedish standardized test of cognitive abilities.

Chapter 3: Theoretical framework & research questions

This chapter introduces the research framework that guides this study; it is based on the studies reviewed in the previous chapter. The discussion of the research framework is followed by the research questions that this study intends to explore and the rationale for those questions.

3.1 Theoretical framework

3.1.1 Rationale for the framework

Traditionally, information behavior studies in the ILS field focus on exploring the relationships between people's individual differences (e.g., gender, age, technical aptitude, and learning styles) or environmental/contextual factors (e.g., tasks and work environment) and their information searching behavior and performance (Borgman, 1989). This traditional line of research contributes to the knowledge of human-IR system interaction by demonstrating relations among biological, psychological, behavioral, and social forces as people use IR systems to solve particular problems. Studies in this line often reveal factual knowledge about information behavior; for example, elderly people have more difficulty in using a particular system than young people, or people with high spatial ability navigate more easily in a particular information space than people with low spatial ability. This general approach of exploring relationships between two variables falls short when it comes to explaining the cognitive sources of the differences.

As the field develops, there is an urgent need to pay more attention to the underlying cognitive mechanisms and processes by which psychological and environmental factors influence people's use of information systems. Because mental models are a mental structure representing the structure-relation of the systems and serve as a proxy through which users interact with the systems, they could be a platform for detailed examination of the underlying mechanisms and processes. Full realization of mental models' potential in explaining cognitive processes supporting people's interaction with IR systems requires researchers to investigate mental models in relation to other important variables involved in human-IR system interaction.

As reviewed in Chapter 2, different approaches have been taken to understand mental models and realize the potential of the concept in the fields of HCI and IR. The research reviewed in Chapter 2 fell into one of three sets: (1) features or attributes (e.g., forms, types, characteristics, content and structure) of people's mental models of various computer applications, including IR systems; (2) factors that affect people's mental models of computer systems, primarily including users' individual differences, such as computer experience, cognitive styles, and contextual factors such as training and system image; and (3) mental models' effects on people's behavior when using computer systems (e.g., people's information searching behavior when using an IR system).

The need to inspect mental models in relation to variables of interest to IR and information behavior research and the review of past work on mental models give rise to the research framework (Figure 1, below). This framework not only demonstrates the current status of research on mental models in IR and information behavior research and motivates the research questions in this study, it also allows future research questions to emerge from the relationships that it manifests.

3.1.2 The research framework

Figure 1 illustrates the framework that guides this study.

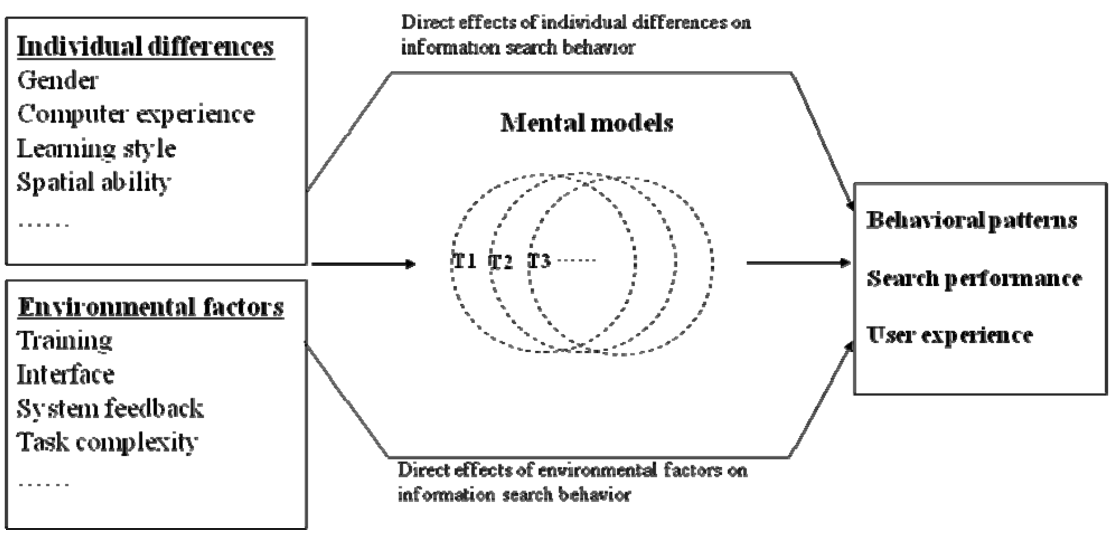


Figure 1. Theoretical framework

As shown in the figure, on the left side are individual differences and contextual factors (independent variables), and on the right side are people's behavior, performance, and experience with IR systems (dependent variables). The bulk of existing user studies in ILS focus on the direct effects of individual differences and contextual factors on people's information searching behavior, performance, and experience, as illustrated by the two direct arrows, one on the top and the other at the bottom of the figure.

When cognitive mental structures, in this case, mental models, are introduced into the picture (as shown by the circles in the middle of the figure), there were two approaches to treat the construct. As has been briefly mentioned, one approach treated mental models as a dependent variable (illustrated by the left-side arrow in the middle of the figure), investigating the impact of individual differences, such as gender, existing knowledge, and computer experience, and environmental factors, such as training conditions and interfaces, on people's mental models of IR systems (e.g., Savage-Knepshield, 2001; Thatcher & Greyling, 1998; Zhang, 2008a). The other approach treated mental models as an independent variable (illustrated as the right-side arrow in the middle of the figure), focusing on investigating the impact of mental models on people's information searching behavior, task performance, or experience with the system (e.g., Borgman, 1986).

There are limitations associated with the current status of research on information behavior in general and mental models of IR systems in particular. As was pointed out at the beginning of the chapter, the investigation of the direct impact of individual differences and contextual factors on information behaviors reveals only factual knowledge about which variable affects which variable, while largely ignoring the underlying mechanisms for such effects. The introduction of mental models into the research signifies researchers' attempts to investigate the underlying cognitive mechanism for certain phenomena concerning people’s use of IR systems. However, the current mental models research treats mental models either as an independent variable or as a dependent variable, failing to coordinate with each other. Mental models are a proxy for people to interact with systems. It is legitimate to speculate that some individual differences’ or environmental factors’ effects on people’s behavior and performance in IR systems might be due to the fact that they affect people’s mental models of the systems; in other words, mental models might mediate individual differences’ and environmental factors’ impact on information searching behavior.

Another shortcoming of the past mental models studies is that few of them treated mental models as a dynamic construct. As has been reviewed in Chapter 2, mental models can be run, and this is an important characteristic differentiating mental models from other mental structures (Ehrlich, 1996). However, in most studies, mental models have been operationalized as a static construct and measured at only one time (e.g., Efthimiadis & Hendry, 2005; Zhang, 2008a), leaving the mental models’ dynamicity largely unexplored. In this framework, mental models (as shown in the middle of the figure) are depicted as a dynamic construct, changing over time as people’s experience with information systems increases (as illustrated by the circles labeled T(ime)1, T(ime)2 and T(ime)3).

Overall, in this framework (Figure 1), mental models are viewed as a mediating factor, mediating the relations between the independent variables and the dependent variables, in the context of people using IR systems. Meanwhile, mental models are viewed as a dynamic construct, changing during people's interactions with a system. Such an integrated view of mental models in IR-human interaction could effectively overcome the limitations of the research on information behavior and mental models in IR outlined in the previous two paragraphs and allow us to pursue a more in-depth understanding of the role mental models play in human-IR system interaction.

3.1.3 Utility of the framework

As has been pointed out, mental models were treated in past studies as either an independent variable or as a dependent variable and were operationalized in different ways. Subsequently, the results from these studies were interpreted in different ways. A theoretical framework, as the one outlined above, could serve as a common ground for research on mental models in human-IR system interaction and, hence, foster the transformation of empirical research into a theoretical understanding of the construct.

In addition, this research framework could help move current user studies beyond merely asking, "*Does* this individual difference or contextual factor lead to different performance with the system?" to asking *how* the individual difference or contextual factor affects people's use of the system. Studies based on this framework may provide a more sophisticated explanation of the interdependencies between psychological and environmental factors, on the one hand, and information searching behavior and task performance, on the other.

This research framework also has practical significance. In information and library science, the exploration of individual differences and contextual factors and their effects on people's mental models of IR systems eventually is expected to inform the design of better IR systems. However, design principles or design guidelines produced by the research often have been about a specific element or function in a particular system. With limited power to inform system design at a higher level, few design tools have been produced based on the existing mental models research.

This framework might help translate the research results to system design at both the element or function level and the overall system level. At the system element level, because mental models mediate the relationships between individual differences/environmental factors and information search behavior, users' mental models could manifest relationships between system elements/functions and certain information behaviors or preferences. Design guidelines can be derived from the newly built connection between the system, the user, and the user's behavior. At the system level, such an integrated view of mental models could inform system designers in a model-based way by providing them with a view of the end users' dynamic cognitive actions in interacting with a system. Providing designers with a model of the user was suggested to be more helpful to software designers than task analysis (Hammond, et al., 1983).

3.2 Research questions and rationale

As has been discussed in the previous section, the current approaches to information behavior research in ILS have limitations. Introducing mental models as a mediating factor between individual differences/environmental factors and search behavior and treating mental models as a dynamic construct to a certain degree could improve the current status of information behavior research. This study investigates the latter issue: the process of mental model construction. The specific research questions are:

1) What changes do people's mental models of an information-rich web space experience during a search session?
2) Does task complexity have an impact on the construction of mental models?

In this section, rationales for focusing on mental model construction and the impact of task complexity are introduced. The additional elements of the theoretical framework will be investigated in future studies.

3.2.1 Construction of mental models of an information-rich web space

If people employ mental models to interact with a system, as suggested by many researchers (e.g., Norman, 1983; Carroll & Olson, 1987), knowledge about how mental models are constructed and modified over time is a particularly rich source for generating practical implications for system design, user modeling, and user instruction. Great attention has been paid to investigating factors that affect the construction of mental models. Prior research results (as reviewed in Chapter 2) have indicated that the construction of mental models is affected by users' existing knowledge structures, training conditions, and the system image. However, these explorations only provide fragmented knowledge about mental model construction, giving us limited insight into the construction of mental models as a process.

Fortunately, some efforts have been made to explore the mental models construction process. For example, Katzeff (1990) described three phases/states of the construction of mental models based on users' think-aloud protocols when performing search tasks: construction, testing, and running. Savage-Knepshield (2001) described mental models construction as a process of incorporating system and task specific features to users' prior knowledge structure. Mayer and colleagues (Mayer, 2002; Mayer & Chandler, 2001) proposed that the construction of mental models of cause-and-effect physical systems consists of two stages: building components models of each major part in the system and building a causal model of the entire system.

Nevertheless, these studies either describe the construction of mental models at too high a level of abstraction to have practical meaning, or fail to incorporate system features and functions into the mental models construction process. In addition, the systems used in these studies are command driven database systems, stand-alone experimental IR systems, or complex physical devices, such as an automobile's brake system. These systems are fundamentally different from the information-rich web spaces in which we are interested in this study.

An information-rich web space is defined as a website that contains a large amount of information (often thousands or millions of web pages) in a particular domain or across domains. With the fast growth of the amount of information online, the number of information-rich websites is increasing and people tend to be more and more accustomed to the information architecture of such sites. People's expectations of IR systems that they encounter later in their lives might be affected by the experience that they have with web-based information-rich spaces. Meanwhile, information-rich web spaces have different characteristics from traditional search based systems. Users tend to show different behaviors in information-rich web spaces than in traditional IR systems. Most prominently, they use a web space for just a short period of time. If they cannot find what they want very quickly, they will switch to another website. Given the growing importance of information-rich web spaces and their differences from heavily researched traditional IR systems, it is necessary for us to understand how people construct mental models of a web space. This leads to the first research question: how do people's mental models of an information-rich web space develop during a search session?

3.2.2 Effects of task complexity on mental model construction

The search task is an important contextual factor in information searching. The current literature consistently demonstrates that tasks have a significant effect on people's information searching behavior and performance (e.g., Byström & Järvelin, 1995). Mental models are developed, validated, and modified during people's interactions with systems to complete particular tasks. These particular tasks provide a context for people's information behavior, as well as for the evolution of their mental models. In other words, the construction of mental models is at least partially embedded in the context of the current task. Certain aspects of the task will inevitably affect the construction and modification of mental models. However, these effects have not been researched.

Task can be characterized by different features, such as size, urgency, difficulty, and complexity (Ingwersen & Järvelin, 2005; Wildemuth & Hughes, 2005). Among these features, task complexity is most discussed and most often investigated in empirical studies. Thus, in this study, we are particularly interested in investigating the effects of task complexity on the construction of mental models of an information-rich website during a search session.

Exploring the effects of task complexity on mental model construction has practical implications for both system design and user instruction. Investigating how tasks affect the way people represent the system can guide designers in shaping the architecture of a system. In particular, designers can pay more attention to aspects of the tasks that impair the sophistication of mental models and figure out ways to reduce the system's cognitive demand. The results from the study will also help to improve user instruction. In the current practice of software or website design, examples or demos are often provided to help novice users learn to use the system. The exploration of the effects of task complexity on people's representations of the system will potentially inform the design of such "out-of-the-box" examples or instructional tools. This leads to the exploration of the second research question: does task complexity have an impact on the construction of mental models?

Chapter 4: Methods

This study has two goals. The first is to investigate the subjects' construction of mental models of an information-rich web space during their interaction with the system. The second goal is to investigate the effects of task complexity as a contextual factor on mental model construction. The information-rich web space used in the experiment is MedlinePlus, a consumer health information website created and maintained by NLM (National Library of Medicine). This chapter describes the research methods and research design used in this study.

4.1 Subjects

A total of 39 (19 males and 20 females) undergraduate students from the University of North Carolina at Chapel Hill participated in the study. The subjects were recruited by emails sent to the undergraduate student mailing list at UNC. The subjects were new users of the platform used in the study, MedlinePlus. Subjects were not majoring in medical related fields, such as nursing, medicine, pharmacy, public health, and biology, due to the possible impact of domain knowledge on people's mental models of systems (e.g., Mayer, 2002). Subjects were assigned to two groups: one performed simple tasks and the other performed complex tasks; see section 4.6 for details of the experimental design.

Previous research suggested that spatial ability might affect people's mental models of information systems (Chen, 2000; Dillon, 2000). To ensure that the two groups were comparable, subjects' spatial ability was measured. The standard paper folding test VZ-2 from the Educational Testing Service's (ETS) *Factor-Referenced Kit of Cognitive Tests* was employed for this measurement. The VZ-2 test has been validated across a wide variety of samples and has consistently demonstrated high test-retest reliability (0.84 reported by Eckstrom et al., 1976). This test also has been administered in an array of information searching or browsing studies (e.g., Chen, 2000; Pak, Rogers, & Fisk, 2006; Swan & Allan, 1998; Westerman, 1998).

Demographic characteristics of the subjects, their experience with the web and with medical information searching, and their spatial ability are reported in Chapter 5. Participation was voluntary. The subjects were given $20 as compensation for their time spent in the study.

4.2 Platform: MedlinePlus

MedlinePlus (www.medlineplus.gov) was selected as the platform for this study as it is a typical information-rich website that is developing rapidly and with which a large portion of the population will come in contact. Meanwhile, medical information searching is an important research topic in ILS because searching for medical information is one of the most popular activities online. Eighty percent of adult Internet users in the US have searched for health information on the Web (Fox, 2005).

MedlinePlus provides authoritative medical information for the general public. It was launched in October 1998 by NLM in response to the intensive use of the MEDLINE database by general consumers via the web. Two sources contribute to the collection of MedlinePlus: the National Institutes of Health's (NIH) publications of consumer health information based on the medical research that NIH sponsors, and publications of professional medical societies and voluntary health agencies without commercial or business motives (Lindberg, 2000). To better serve the public, MedlinePlus also licenses information sources, such as medical dictionaries, encyclopedia, detailed information about prescription drugs, directories of health professionals and hospitals, news feeds, and tutorials. Recently, MedlinePlus started providing health information in over 40 languages. The website's content is updated regularly, but its structure is stable.

MedlinePlus can be freely accessed on the web and its interface has the look-and-feel of general hypertext-based information-rich websites (Figure 2). NLM staff have been working hard to keep up with the fast pace of evolving expectations for web interface design (Marill, Miller, & Kitendaugh, 2006). Unlike PubMed, a search oriented system targeted for physicians and medical researchers, MedlinePlus is a browsing oriented system targeted for health care consumers. Information is mainly organized by health topics in a hierarchical manner. Users can access information by subject or by alphabetical order. Users also can search the site, however, the search function is very simple, similar to general web search engines. Users cannot limit the search to certain fields as they can in some other IR systems.

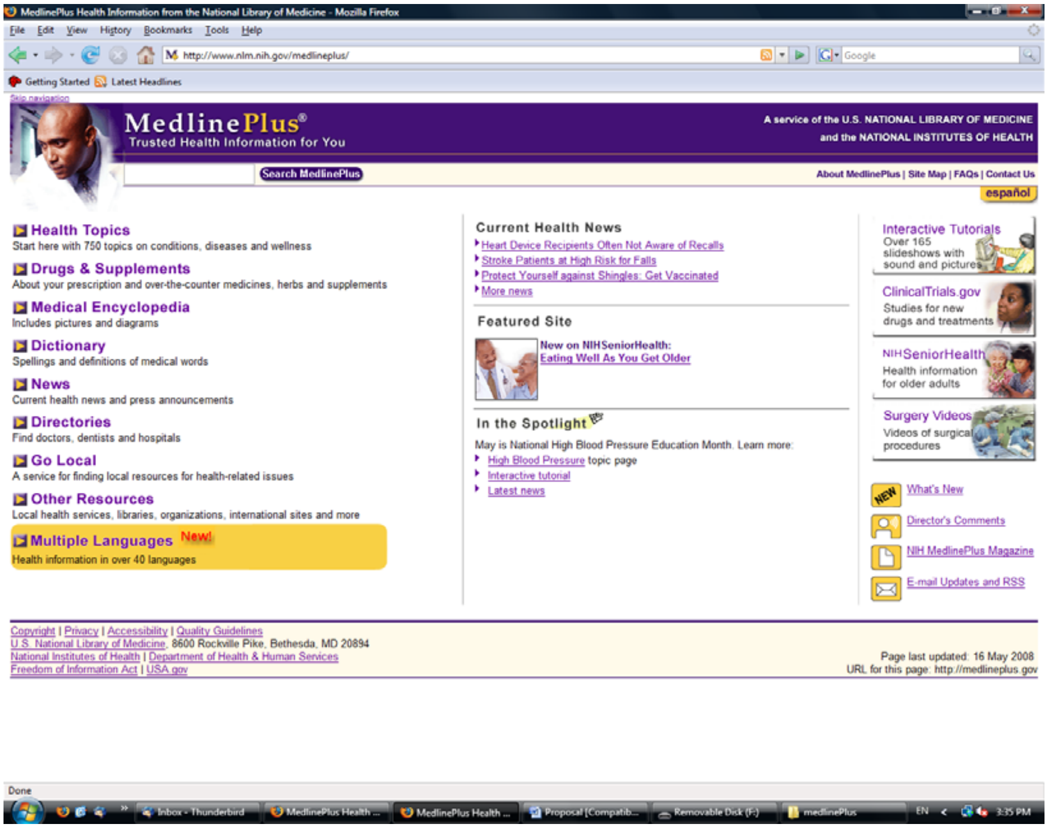


Figure 2. MedlinePlus Homepage

4.3 Tasks

Two types of tasks were defined in the study: simple tasks and complex tasks. The complexity of a task was determined by three factors: (1) the clarity of the information required to answer the question; (2) the distribution of the answer; and (3) the extent to which a higher level cognitive activity, such as synthesizing information from multiple places, was required to complete the task. A simple task is a well-defined question. It is clear what information is required. The answer to a simple task is located on one page and easy to recognize. Little cognitive effort (e.g., information synthesizing) is required.

A complex task is often an open-ended question. The desired information is less clear. Participants often need to decide what aspects of a subject they will cover in the answer. For example, there are many ways to treat high blood pressure, such as diet, exercise, and medicine. It is up to users to seek out different aspects of the possible treatment. Answers to a complex task are often located on multiple webpages. High level cognitive activities, such as comparing, interpreting, and synthesizing information, are involved in tackling a complex task.

To be realistic, the tasks were selected from Yahoo! Answers, a social question and answer website where people post their own questions and/or answer questions posted by others. Some of the questions were directly used for the study, while others were modified to be more suitable for the scope of the system. For every task, a scenario was provided.

To ensure that each task was correctly categorized as a simple or a complex task, two information professionals who specialize in medical information were asked to rate the complexity of each task. They were given the definition of task complexity described above and instructed to perform every task using MedlinePlus. Their decisions were based on both their knowledge about MedlinePlus and their search experience with these tasks. Tasks used in the study were tasks whose complexity level was agreed on by both raters.

The instructions for the task complexity judgment for the invited information professionals were:

> These tasks will be used in a study to explore the effects of task complexity on people's mental models of MedlinePlus and on their information searching performance. Thus, it is important that the

categorization of each task as simple or complex is valid.

> You are asked to be one of the judges to judge the complexity of the tasks because of your experience of working in the medical information area. You will need to read the definition of task complexity provided to you and use MedlinePlus (www.medlineplus.gov) to try out the tasks. Based on the definition and your search experience with the tasks, you will decide the complexity of each task (simple or complex).

In total, 16 simple tasks and 5 complex tasks were selected for the study. A complete list of the tasks can be found in Appendix C. An example of a simple task is:

> Protein is a "building block" nutrient. Your body uses protein to build tissue, such as white and red blood cells, other cells in the immune system, skin, hair, and muscle. Given the importance of protein in your body, you want to find out how much protein an average person needs each day.

An example of a complex task is:

> Imagine that your friend recently was diagnosed with asthma and was put on two inhalers. But he thinks it is chronic bronchitis. So he wants to know what the similarities and differences between asthma and chronic bronchitis are. Also, he wants to know various means to treat or soothe asthma, such as medicine, diet, alternative medicines, exercise, etc. You want to help him by finding as much information as possible in MedlinePlus.

The assignment of the tasks to the two groups is described in section 4.6. After finishing each task, subjects were asked to rate the difficulty of the task (1-very easy, to 5-very difficult), the mental effort required to finish the task (1- very small amount, to 5-very large amount), and their satisfaction with their own performance (1- very disappointed, to 5-very satisfied) on a 5-point scale.

4.4 Measurement of mental models

Mental models are people's mental representations of the states, structures, functions, or behaviors of an information system. As a mental construct, mental models are not directly observable and no agreement has been achieved on the form, symbolic or pictorial, of mental models (Doyle & Ford, 1998). Thus, in the current study, mental models were measured by multiple methods: a concept listing protocol, semi-structured interviews, and drawings. The rationale for and a description of each method is introduced separately, below. The complete set of mental model measurements is attached as Appendix B.

4.4.1 Concept listing

Concept listing provides an efficient means to elicit key concepts that people have about a domain. The protocol is able to generate qualitative data (e.g., concepts in the response, conceptual closeness between terms, emotionality, classification of responses, etc.) and quantitative data (e.g., occurrence and frequency of a particular response, the associative reaction time, response size, response entropy, etc.) (Cramer, 1968). The qualitative data that is especially useful for mental model representation is the temporal sequence of associations—the process of getting from one concept to another, as well as groups or clusters of the concepts. The quantitative data that is informative for mental model representation is the associative reaction time. It could serve as an indication of the closeness of two concepts or two clusters of concepts in subjects' mental models.

In the concept listing protocol, participants were asked to list concepts related to MedlinePlus. Each concept was viewed as a memory node and the list of concepts was viewed as the result of subjects' cognitive process of making sense of and representing the website. Concept listing has been employed to study people's mental models of IR systems (Pejtersen, 1991) and a knowledge domain in IS (Wang et al., 2005). Pejtersen reported that the subjects' responses clearly indicated the various facets in the subjects' conceptual network. Wang et al. reported that terminologies listed by some subjects formed distinct clusters on particular topics in information organization.

Both exploratory studies suggested that, to elicit meaningful concepts about a domain, subjects should be given strong primes. Therefore, in the current study, subjects were primed with clear instructions about what concepts are expected from them. The concept listing instructions were:

> In this task, you are asked to illustrate your understanding of the current system by listing concepts. The concepts could be, but are not limited to, the system's component parts, objects in the system, its working mechanisms, functions, and processes. Remember that you can list any concepts that you believe are important in representing your thoughts about the system.

A computer program was designed for the concept listing protocol (Figure 3). Subjects were instructed to type concepts into the text box in the order that the concepts appeared in the mind. After each input, subjects clicked on the submit button to submit the concept(s). The submitted concept then appeared in the lower part of the screen right away. Concepts listed by subjects were stored in a database. Each concept listing protocol was limited to 5 minutes. To ensure that subjects understood the concept listing protocol and the computer program, a demonstration of using the program to express a student's understanding of the field of psychology was presented to subjects before they proceeded to the concept listing protocol.

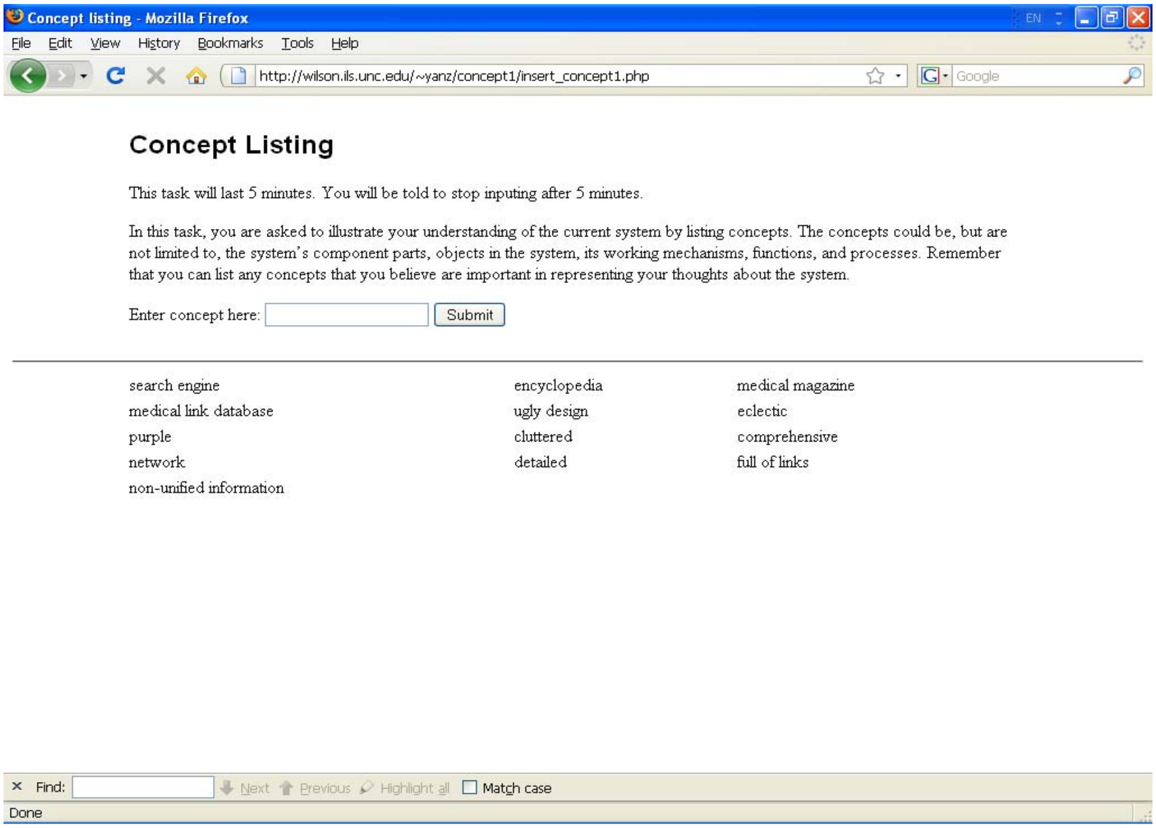


Figure 3. Data collection program for concept listings

4.4.2 Semi-structured interviews

Interviews have been widely used in the current literature to investigate people's mental models of various computer systems (Staggers & Norcio, 1993; Rutherford & Wilson, 1992). It is generally accepted in psychology and behavioral science that verbal accounts are a valid means to represent mental activities. In the current study, interview questions were directed to probe three aspects of mental models: (1) the structure of the target system, (2) functions and working mechanisms of the system, and (3) subjects' procedural knowledge about how to use the system to solve a particular task. The following questions were asked to probe the first two aspects.

1) What information is provided by MedlinePlus?

2) How do you think information in MedlinePlus is organized?
3) How do you think the system works?

To elicit procedural knowledge about how to perform a search in MedlinePlus, subjects were asked to describe their strategies and steps for handling a hypothetical task. The hypothetical task and the instruction for the description are:

> Imagine that you are required to write a paper about hepatitis to arouse public attention to the seriousness of this disease. In the paper, you want to include, but are not limited to, aspects such as what is the difference between chronic and non-chronic hepatitis? What determines this? What are the differences between hepatitis A, B, and C, and what are the treatments for each of them, and can we prevent the development of liver cirrhosis among patients with chronic hepatitis? You decide to use MedlinePlus to collect information for your paper.
>
> What steps would you take in order to find information for your research? Write down each of the steps that you would follow as if you were actually using the system to find related information.

In this study, the results from the semi-structured interviews were used primarily to cross-validate the results of the concept listing protocol.

4.4.3 Drawing

Some researchers have argued that mental models are, by nature, pictorial (e.g., Rouse & Morris, 1986). Several studies have used drawing as a method to represent people's mental models of IR systems (e.g., Efthimiadis & Hendry, 2005). Significant relationships were found between drawings and subjects' computer experience, genders, and scores on system related questions (Thatcher & Greyling, 1998; Zhang, 2008a). These empirical results, from another perspective, suggest that drawing might be a useful method for representing mental models. In this study, the instructions for the drawing task were designed to be general, not restricting the emergence of dimensions of mental models:

Please draw a diagram or a picture of your perceptions about MedlinePlus.

The results from the drawing tasks will not be reported here. A follow-up analysis will be conducted later and the results will be reported elsewhere.

4.5 User experience when using the system

After the first search session, subjects were asked to rate a series of statements about their experience with the system in relation to the following aspects: ease of learning, ease of use, usefulness of the system, understanding of the website's working mechanisms, satisfaction with the information provided by the website, interface design of the website, enjoyment of and engagement with the website, and intention to use the system in the future. The rating scale was a 5-point Likert scale, on which a "1" indicated strong disagreement with the statement, a "3" indicated a neutral rating, and a "5" indicated strong agreement with the statement.

At the end of the questionnaire, subjects were asked to rate their overall experience with the site: whether they are satisfied/dissatisfied, pleased/displeased, contended/frustrated, and delighted/disappointed with the system. The rating scale is a 7-point semantic differential scale (1 = extremely dissatisfied/ displeased/ frustrated/ disappointed, to 7= extremely satisfied/ pleased/ contended/ delighted). The complete user experience questionnaire is attached in Appendix E.

4.6 Research design and experimental procedures

This study used an adapted form of a between-subjects pretest-posttest experimental design. One group of subjects performed a set of simple tasks, followed by a set of tasks containing both simple and complex tasks. The other group performed a set of complex tasks, followed by the same set of mixed tasks. The data collection sessions took place in a private lab in School of Information and Library Science. All subjects were tested individually and each session lasted approximately 2 hours.

The computer for data collection was equipped with the Windows Vista operating system. Internet Explorer (IE 6.0) was used as the default Internet browser because of its wide market coverage. The starting page in IE was set to the homepage of MedlinePlus. The experimental design is illustrated in Figure 4.

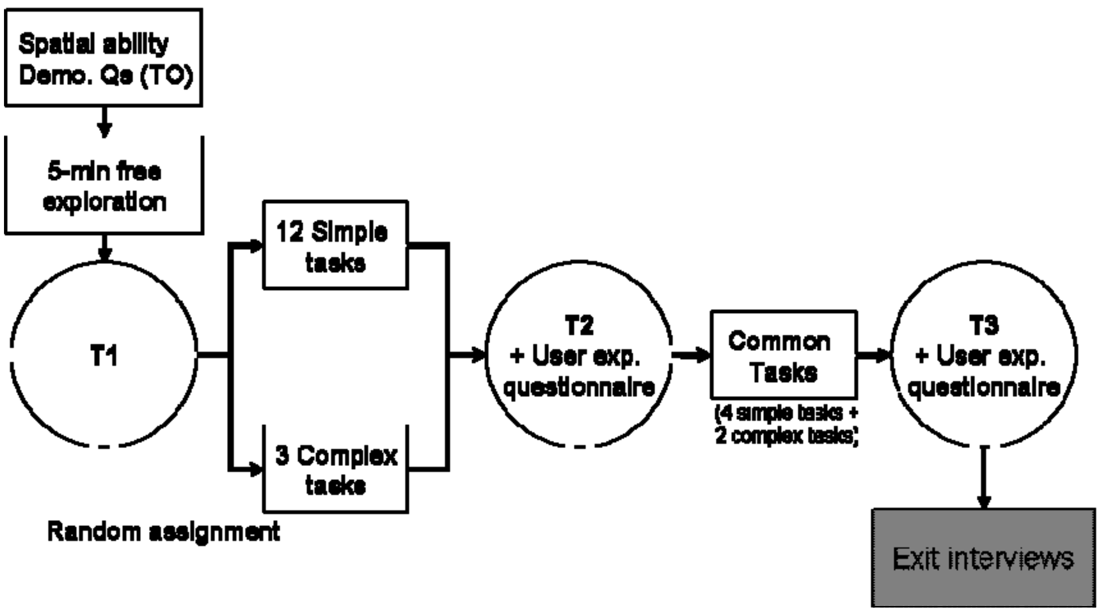


Figure 4. Research procedure[1]

Upon arrival, subjects were welcomed and received a brief introduction to the study. They then were asked to review and sign an informed consent form.

After the consent form, subjects finished the spatial ability test VZ-2, followed by a demographic questionnaire, asking about participants' computer experience, experience with medical information searching, and general impressions of information-rich websites (Appendix A). The questions about information-rich websites are considered a T0 assessment of the subjects' mental models. Then, subjects were directed to the testing computer and spent five minutes exploring the system as they normally do when they encounter a new website. After the exploration, participants finished the three mental model measurements the first time (T1): concept listing, semi-structured interviews, and drawing (Appendix B). Before the concept listing protocol, subjects were presented with an

[1] Data collection at T1, T2, and T3 included the concept listing protocol, a semi-structure interview, and the drawing protocol.

example of how to perform the concept listing using an in-house developed computer program (Appendix B).

Next, participants were randomly assigned to one of the two groups, simple task group or complex task group. The simple task group performed 12 simple tasks (Appendix C), and the complex task group performed 3 complex tasks (Appendix C). The number of tasks (in both this search session and the second search session) ensured that subjects had a good amount of exposure to the system and were able to form mental models of the system. The order of the tasks for each group was randomized. Subjects were instructed to take as much time as they wanted to finish the tasks. After completing each task, subjects rated the difficulty of the task, the mental effort required to accomplish the task, and satisfaction with their own performance with the task (Appendix D) (as has been mentioned in section 4.3). After the search session, subjects completed the mental models measurements (concept listing, interviews, and drawing) the second time (T2), followed by the user experience questionnaire in which subjects provided assessments of their experience with the MedlinePlus website (Appendix E).

After completing the user experience questionnaire, subjects in both groups were asked to perform a common set of tasks. This set of tasks included 4 simple tasks and 2 complex tasks. The simple tasks were always presented to the subjects before the complex tasks; within each type of task, the presentation order was random. Similar to the first search session, after completing each task, subjects rated the difficulty of the task, the mental effort required to accomplish the task, and their satisfaction with their own performance of the task (Appendix D). After this second search session, subjects completed the mental model measurements (concept listing, semi-structured interview, and drawing) the third time (T3), followed by the same user experience questionnaire (Appendix E).

Both search sessions were video captured using Camtasia software. An exit interview (Appendix F) was conducted at the end of the study with each subject. In the interview, subjects provided assessments of the search tasks, the MedlinePlus system, and their search processes. They also were asked to comment on the mental models measurement methods, particularly whether they expected the repetition of the measurements. The answers from the subjects helped validate the repeated measurement of mental models. Upon completion, the subjects were thanked and debriefed about the goals of the study.

4.7 Data analysis

The following types of data were generated in this study:

1) demographic data, such as age, web experience, and use of medical information. In the demographic questionnaire, subjects also described their perceptions and opinions of general information-rich web spaces and their typical strategies for using the system (T0),
2) subjects' spatial ability scores (VZ-2),
3) concepts provided by subjects in concept listing tasks (repeated measurements: T1, T2, and T3),
4) semi-structured interviews (T1, T2, and T3), including users' understanding of the MedlinePlus system and their descriptions of expected strategies to solve a hypothetical task.
5) drawings (T1, T2, and T3),
6) users' experience with the system measured by a user experience questionnaire (repeated measurements: T2 and T3),
7) session length (two search sessions),

8) subjects' evaluations of task difficulty, mental effort, and satisfaction with their own performance for each task, and
9) video-taped search processes.

Among these types of data, drawing and video-taped search processes will not be analyzed in this study.

4.7.1 Quantitative data analysis

Descriptive statistics (mean and S.D.) were calculated for all the quantitative variables produced in the study. *T*-tests were performed to examine the differences between the simple and complex task groups on the following variables: demographic information, including age and web experience, spatial ability, session length, subjects' experiences with the system (T2 and T3 separately), number of concepts contributed by each concept listing protocol (T1, T2, and T3), and subjects' evaluations of task difficulty, mental effort, and satisfaction with their own performance.

Fisher's exact tests were performed to examine the differences between the two groups on the following categorical variables: demographic information, including the frequency of searching medical information online and the use of health information sources. Fisher's exact tests were also performed to examine the differences between the two groups on the strategies that they expected to employ to solve the hypothetical task.

The developmental trend of subjects' experience with the system over time (from T2 to T3) was analyzed using paired-samples *t* tests. Paired-samples *t* tests were also used to evaluate the change in the number of concepts generated over time (T1 to T2 and T2 to T3).

Significant differences for all statistics were those with a p-value of less than .05.

4.7.2 Qualitative data analysis

The concept listing protocols produced lists of concepts concerning various aspects of the system. Content analysis was used to code the data at two levels: the basic dimensions of the mental models, and the components of those dimensions. The coding unit was a concept. In most cases, subjects input a term or a phrase that represented a single concept (such as "alphabetical" and "background research"). One code was assigned to such records (corresponding to the previous examples: information organization: schema, and system: the usage of the site). Sometimes subjects typed in a phrase or a sentence that included multiple concepts (such as easily accessible medical database). Multiple codes were assigned to such a record (e.g., content: subject, and system: evaluation). Thus, the categories and coding scheme emerged from open coding. During the process of generating categories, the constant comparative method (Glaser & Strauss, 1967) was used. When a concept was assigned to a category, it was compared with each of those already assigned to that category.

Some concepts contributed by subjects in the concept listing protocols, such as online, technical, organize, words, practice, questions, mechanical, reemergence, narrow and decisive, were too general or lacked the contextual information needed for the researcher to make reliable interpretations. Such concepts were excluded from the statistical analyses. A second coder was asked to code 10% of all the concepts contributed by the subjects.

Inter-coder reliability was calculated using Krippendorff's alpha (Krippendorff, 2004). Subjects' mental models of MedlinePlus at the three different time points were thus induced, mainly based on the concept listing data.

Subjects' descriptions of their perceptions and opinions of general information-rich web spaces and their typical strategies for using such systems, reported in the semi-structured interviews, were analyzed using content analysis, as were the responses to open-ended questionnaire items. The interviews were video-taped and then transcribed prior to analysis. QSR N6 software was used to assist with content analysis. Open coding was employed and the coding unit was a concept or a theme. Whenever a new concept (such as "videos", "database", "search engine") or a new theme (such as means to use the site and the ways in which information is organized) appeared, it was coded into a free node. Categories emerged by organizing free nodes based on thematic similarity. During the coding process, the constant comparative method was also employed: each text assigned to a category was compared with each of those already assigned to that category. Systematic comparison not only helps in understanding the theoretical properties of the category, but is also useful in making differences between categories apparent (Zhang & Wildemuth, 2009). Because the analysis of this section of the semi-structured interview was only to test the validity of the analysis of the concept listing data, an inter-coder reliability check was not performed.

Data generated by the second section of the interviews were procedural steps used by subjects to find information for a hypothetical task. Information search strategies, particularly general searching (strategy A), an efficient path to access information in MedlinePlus (strategy B), and general browsing (strategy C) were coded based on subjects' descriptions. If a subject reported using multiple strategies, multiple codes were assigned to that subject.

4.7.3 Mental models construction

The process of users constructing mental models of the MedlinePlus website was demonstrated by subjects' mental models at four points in time, as shown in Figure 4: (1) subjects' mental models of general information-rich web spaces before they saw MedlinePlus (T0), (2) subjects' mental models of MedlinePlus after 5 minutes of free exploration (T1), (3) subjects' mental models of MedlinePlus after the first search session (T2), and (4) subjects' mental models of MedlinePlus after the second search session (T3).

Subjects' initial models (T0) of information-rich web spaces before they saw MedlinePlus were derived from content analysis of the description that subjects provided in the demographic questionnaire about their impressions of and opinions about general information rich web spaces and their strategies for using such sites.

Subjects' mental models of MedlinePlus at T1, T2, and T3 were constructed through both bottom-up and top-down approaches. On one hand, dimensions of the mental models emerged from the content analysis of the concepts that subjects contributed in the concept listing protocols. The analysis of the first section of the semi-structured interviews was used to verify the structure derived from the concept listing data. On the other hand, mental models theory suggests that mental models encompass procedural knowledge of how to perform tasks using a system. This behavioral dimension of mental models was induced based on the data from the second section of the semi-structured interviews, where subjects described steps for solving a hypothetical question.

The comparison of the mental models at T1, T2, and T3 was performed from two perspectives: (1) paired-samples *t* tests were performed to test whether the number of concepts that subjects dedicated to particular parts of the mental models changed over time; and (2) qualitative analysis was conducted to determine whether the content or substance of the concepts changed over time, in other words, whether subjects' understanding of the system changed over time. These analyses directly addressed the first research question.

4.7.4 Evaluating the impact of task complexity

The second research question was concerned with the impact of task complexity on subjects' mental models. For this analysis, the models expressed at T2 and T3 were of interest, with a focus on comparing the models of the two groups. At T2, the two groups had just completed different sets of tasks: one group had completed only simple tasks (the simple task group) and the other group had completed only complex tasks (the complex task group). At T3, each group also had completed a common set of tasks that included both simple and complex tasks. To address the second research question, the two groups' mental models were compared at the conclusion of each phase both quantitatively and qualitatively.

Chapter 5: Results and discussion: Subjects, tasks, and user perceptions of the system

Mental models are not constructed in a vacuum; they are always developed and constructed in a particular context. This chapter reports data concerning subjects and their perceptions of the tasks in the study, as well as the time they spent completing the tasks. In addition, subjects' perceptions of the information-rich web space, MedlinePlus, are reported. These data on subject characteristics, subjects' perceptions of the assigned search tasks, the length of the search sessions, and subjects' perceptions of the MedlinePlus system will set the context for the discussion of the construction of mental models, taken up in Chapter 6.

5.1 Subjects

Previous research suggested that mental models of IR systems are affected by individual differences, such as gender, domain knowledge, spatial visualization ability, and computer experience (e.g, Mayer, 2002; Thatcher & Greyling, 1998; Vicente, Hayes, & Williges, 1987; Zhang, 2008b). In this section, the following characteristics of the subjects in the study are reported and discussed: demographic data, including gender, age, and major fields of study, subjects' spatial ability, subjects' internet experience and their experience with searching medical information.

5.1.1 Demographic data

A total of 39 (19 males, 20 females) undergraduate students participated in the study. One subject was excluded from the data analysis due to his extremely low spatial ability, because spatial ability is related to navigation in hyper-link based systems (e.g., Campagnoni & Ehrlich, 1989). All the subsequent data analysis was based on 38 subjects (18 males, 20 females).

Of the 38 participants, 3 (7.9%) were freshmen, 3 (7.9%) were sophomores, 3 (7.9%) were juniors, and 29 (76.3%) were seniors. The ages of the participants ranged from 18 to 22 years. Subjects were randomly assigned to two groups, the simple task group and the complex task group, within gender, as they arrived for the experiment. Table 1 lists the demographic information for the two task groups. The two groups were not significantly different in their demographic characteristics.

Table 1. Demographic information for the simple and complex task groups

	N	Female	Male	Mean age in years (SD)
Simple task group	19	9	10	20.95 (0.85)
Complex task group	19	11	8	20.37 (1.21)
Total	38	20	18	20.66 (1.07)

To control the impact of domain knowledge on mental model construction, students who majored in a medical related field, such as nursing, medicine, pharmacy, public health, and biology, were excluded from the recruitment. As a result, participants in the study were from 22 different non-medical related majors. One subject had not decided on a major area of study. Table 2 shows a summary of subjects' major areas of study.

Table 2. Subjects' major areas of study

Major	No. of subjects
Art history	2
Business	3
Communication studies	2
Economics	2
Elementary education	1
English	2
English/Journalism	1
English/Math	1
Geography	2
History/Spanish	1
History/Peace, war, and defense	1
History/Political science	1
International studies	1
International studies/French	1
International studies/History	1
International studies/Political science	1
Journalism	2
Latin/French	1
Mathematics/Linguistics	1
Peace, war, and defense	1
Physics	1
Psychology	4
Psychology/History	1
Religious studies	1
Sociology/Japanese	1
Spanish	1
Undecided	1

5.1.2 Spatial visualization ability

In the study, the complete version of VZ-2 was administered, which includes 20 questions. The VZ-2 test was scored based on the instructions from Ekstrom et al. (1976). The instructions state that the score on this test will be "the number marked correctly minus a fraction of the number marked incorrectly”. Therefore, in this study, subjects’ scores were calculated using the formula: C-IC/5, where C is the number of correct answers out of the 20 items and IC is the number of incorrect answers out of the 20 items. This formula is adjusted for guessing (Alonso, 1998). Table 3 shows the spatial ability measurements for both the simple and the complex groups. No statistically significant difference was found between the groups. The result suggests that possible differences in the construction of mental models between the two groups should not be attributed to the difference in spatial visualization ability.

Table 3. Subjects’ spatial visualization ability

	N	Mean (SD)	Max.	Min.
Simple task group	19	12.99 (3.56)	18.8	8.0
Complex task group	19	12.42 (4.04)	18.8	6.8

5.1.3 Experience with the web and medical information searching

Subjects came to MedlinePlus with different pre-existing models and assumptions about the system. Their past experience with the internet, where hypertext-based systems are hosted, might have an impact on the construction of their mental models of MedlinePlus. In this study, subjects had used the web for 6 to 13 years. The simple task group, on average, had 9.89 years of experience (*SD* = 1.76) with the web, and the complex task group had 9.84 years of experience (*SD* = 2.03). The two groups were not significantly different, which indicates that experience with the web couldn’t be used to account for possible differences between the groups in the construction of mental models.

Exploring subjects' experience with searching for medical information online and the use of medical information sources could shed light on how people perceive MedlinePlus. Thus, subjects' experience with medical information searching is reported. Among the 38 subjects, 35 (92.1%) claimed that they had used the web to look for medical information. The frequency of usage of the web for medical information ranged from once per year to 3 times per week. Table 4 shows the frequency of the subjects' searching for medical information by group. As shown in the table, two categories were defined: at least monthly, and yearly or never. A Fisher's exact test showed that the two groups were not significantly different in their experience with using the web for medical information (p = 0.74).

Table 4. Frequency of searching for medical information online

	At least monthly	Yearly or never
Simple task group	9 (47.4%)	10 (52.6%)
Complex task group	7 (36.8%)	12 (63.2%)
Total	16 (42.1%)	22 (57.9%)

In the demographic questionnaire, subjects were also asked to report the information sources that they referred to for medical information. Table 5 shows the information sources that the subjects had used for medical information.

Table 5. Information sources for medical information

Information Sources	Simple task group	Complex task group	Total
General search engines	17	15	32 (84.2%)
Family & friends	16	14	30 (78.9%)
Doctors	14	16	30 (78.9%)
WebMD	13	11	24 (63.2%)
Wikipedia	9	9	18 (47.4%)
UNC online health sources	6	5	11 (28.9%)
Books	4	1	5 (13.2%)

As shown in the table, three main types of resources for medical information were reported by the subjects: people, websites, and books. About 79% of the subjects reported that they had asked for medical information from family, friends, and doctors. One subject mentioned that she had sought medical advice from her athletic trainer. It is clear that consulting with people is a major way to get medical information or advice.

Among subjects who used online sources for medical information, the most popular sources were general search engines (84.2% of subjects had used), followed by WebMD (63.2% of subjects), and Wikipedia (47.4% of subjects). Close to 29% of the subjects also had used the online health sources provided by UNC (i.e., health information under the domain of unc.edu). Only 13.2% of subjects reported using books for medical information. A series of Fisher's exact tests showed that the two groups did not have significant differences in their use (and non-use) of the listed medical sources. Other health sources used by only a couple of subjects included Yahoo Health, PubMed, and the People's Pharmacy.

The analysis of subjects' experience with medical information searching suggested that they were not heavy users of medical information. Over half of them only seek medical information on a yearly basis. Meanwhile, the majority of the subjects in the study recognized and used general web search engines; WebMD, a commercial consumer health information resource; and Wikipedia, a user-contributed online encyclopedia. The low frequency of seeking medical information and the heavy use of online resources might be typical for young people whose medical needs are limited and who grew up with the web. The two groups were comparable in their experience with searching medical information online.

5.2 Users' assessments of tasks

Investigating the impact of task complexity on the construction of mental models is one of the main goals of the study. Thus, task complexity is an important variable in the study. It is important to ensure that the construct was successfully operationalized. The validation was achieved by examining the agreement between the subjects' and medical information professionals' judgments on task complexity. Recall that the complexity of each assigned task was reviewed and agreed upon by two medical information professionals before the study. This section reports the subjects' assessments of the tasks.

After finishing each task, subjects were asked to rate task difficulty, mental effort to solve the problem, and satisfaction with their own performance, on a 5-point Likert scale. Table 6 lists subjects' self assessments of tasks and their performance on the first set of tasks, where the simple task group performed 12 simple tasks, and the complex task group performed 3 complex tasks. The number in the table is the average of ratings across tasks in each task group.

Table 6. Subjects' assessments of task difficulty, mental effort, and satisfaction with performance for task set 1 (mean ratings)

Group	N	Difficulty [a]	Mental Effort [b]	Satisfaction [c]
Simple	19	2.07	2.11	4.04
Complex	19	2.86	2.93	3.46

[a] 1 – Very easy, 5 – very difficult
[b] 1 – Very small amount, 5 – very large amount
[c] 1 – Very disappointed, 5 – very satisfied

All the differences between the two groups' ratings were statistically significant. As shown in the table, the complex task group rated the complex tasks as more difficult ($t(36) = 5.49$, $p < .001$) and requiring more mental effort ($t(36) = 5.99$, $p < .001$) than the simple task group rated the simple tasks. The complex task group also felt less satisfied with their performance on their tasks ($t(36) = 4.36$, $p < .001$). The statistically significant differences between the subjects' assessments of the tasks that they performed indicate that subjects can easily differentiate the complexity of the tasks. Subjects' judgments of task complexity matched the judgments made by two medical information professionals, who classified tasks into simple and complex groups before the study. This agreement suggests that the construct of task complexity as a between-subject variable was operationalized reliably.

Table 7 lists subjects' assessments of tasks and their performance on the second set of tasks, a common set of tasks performed by both groups. This task set includes 4 simple tasks and 2 complex tasks.

Table 7. Subjects' assessments of task difficulty, mental effort, and satisfaction with performance for task set 2

Group	N	Assessments of simple tasks in the task set			Assessments of complex tasks in the task set		
		Difficulty [a]	Men. Eff. [b]	Satisfaction [c]	Difficulty [a]	Men. Eff. [b]	Satisfaction [c]
Simple	19	1.74 (.40)	1.79 (.44)	4.29 (.58)	3.29 (.67)	3.24 (.65)	3.03 (.90)
Complex	19	1.49 (.35)	1.82 (.61)	4.53 (.49)	3.29 (.65)	2.97 (.46)	3.08 (.53)

[a] 1 – Very easy, 5- very difficult
[b] 1 – Very small amount, 5 – very large amount
[c] 1 – Very disappointed, 5 – very satisfied

Test of Homogeneity of Variances was performed and the two groups did not show significantly different variances. A series of *t* tests were then run to examine whether the two groups were different in their assessments of task difficulty, mental effort, and satisfaction for both simple and complex tasks in the common task set. The results showed that there were no statistically significant differences between the groups in assessing task difficulty, mental effort and satisfaction with performance for both the simple and complex tasks.

Paired-samples *t*-tests were performed to test whether subjects were different in their assessments of the task difficulty, mental efforts, and satisfaction for simple and for complex tasks. The test results showed that, within this common set of tasks, subjects in both groups reported that the complex tasks were more difficult ($t(37) = 14.58$, $p < .001$) and required higher mental effort ($t(37) = 11.97$, $p < .001$) than the simple tasks. And they were less satisfied with their performance on complex tasks than they were on simple tasks ($t(37) = 10.94$, $p < .001$). The results suggest that subjects were able to distinguish the differences between the two types of tasks and the manipulation of the variable of task complexity was successful in this second set of tasks.

5.3 Session length

In the study, subjects were allowed to spend as much time as they want on each set of tasks. Table 8 shows the time that both groups spent on the two search sessions.

Table 8. Time for completing two sets of tasks

	Session 1 (mins)	Session 2 (mins)
Simple task group	20.11 (4.68)	11.72 (4.08)
Complex task group	19.86 (7.20)	12.98 (3.51)

As shown in the table, both groups spent about 20 minutes performing the first set of tasks and about 12 to 13 minutes on the second set of tasks. There were no significant differences between the two groups in the amount of time they spent finishing both set of tasks, which suggests that the length of time that subjects spent with the system would not be a factor affecting the construction of mental models.

Another perspective on the time spent searching is to think about the amount of time required to complete each task as an indicator of its complexity. For the first set of tasks, the simple task group spent, on average, 1.68 minutes on each task ($SD = 0.39$), and the complex task group spent, on average, 6.62 minutes on each task ($SD = 2.4$). A t-test shows that the complex task group spent significantly more time on each task ($t(36) = 8.86$, $p < .001$). For the second set of tasks, across both groups, the average amount of time spent on each simple task was 1.05 minutes ($SD = 0.51$) and the average amount of time spent on each complex task was 3.46 minutes ($SD = 1.21$). A Paired-samples t-test comparing these two means indicates that the difference was statistically significant ($t(37) = 12.21$, $p < .001$).

This result provides further evidence that the implementation of the construct of task complexity was a valid one.

5.4 Users' perceptions of the system

Users' perceptions of the system were measured twice by a user experience questionnaire, once after they performed the first set of tasks (the simple task group performed simple tasks and the complex task group performed complex tasks), and the other after they performed the second set of tasks (the same set of tasks for both groups).

The user experience questionnaire (Appendix E) consists of a series of statements; subjects rated each statement on a 5-point Likert scale (1 = strongly disagree, 3 = neutral, and 5 = strongly agree). At the end of the questionnaire, subjects rated their overall experience of using MedlinePlus (satisfied/dissatisfied, pleased/displeased, contented/frustrated, delighted/disappointed) on a 7-point Likert scale (1 = extremely dissatisfied/displeased/frustrated/disappointed, 4 = neither, 7= extremely satisfied/pleased/contented/delighted). Table 9 summarizes these ratings after each session.

Table 9. Users' perceptions of the system after session 1 and after session 2 (means and standard deviations in each column)

User experience with the system		Simple task group		Complex task group	
		After session1	After Session2	After session1	After Session2
Easy to learn		3.73 (.87)	3.63 (.76)	3.53 (.70)	3.79 (.54)
Easy of use		4.03 (.67)	4.05 (.54)	3.65 (.68)	3.73 (.59)
Usefulness of the system		3.86 (.68)	3.96 (.61)	3.44 (.81)	3.57 (.57)
Understanding of the website's working mechanisms		**3.58 (.49)**	3.75 (.47)	**3.39 (.55)**	3.57 (.59)
Satisfaction with the content		**4.07 (.58)**	**4.19 (.58)**	4.03 (.38)	4.12 (.42)
Interface design		3.95 (.76)	3.89 (.68)	3.82 (.65)	3.61 (.72)
Enjoyment of the website		3.32 (.75)	3.37 (.76)	3.11 (.74)	3.21 (.71)
Engagement of the website		3.37 (1.12)	3.16 (.90)	3.31 (.89)	3.37 (1.01)
Intention to use the website in the future		3.92 (.71)	3.97 (.87)	**3.45 (1.03)**	**3.63 (.97)**
Overall experience	Satisfied	**5.63 (1.01)**	***5.79 (.79)***	**4.68 (1.42)**	***5.05 (1.27)***
	Pleased	5.16 (1.01)	5.32 (1.06)	4.42 (1.35)	4.63 (1.30)
	Contented	**5.26 (1.15)**	5.05 (1.47)	**4.11 (1.33)**	4.42 (1.22)
	Delighted	4.37 (1.21)	***4.68 (1.20)***	4.00 (.75)	***3.95 (.71)***

We will first turn our attention to comparisons across groups after session 1, where the simple task group performed simple tasks and the complex task group performed complex tasks. At this point, the simple task group reported significantly better understanding of the website's working mechanisms ($t(36) = 2.28$, p=.029). They also felt more satisfied ($t(36) = 2.37$, p=.023) and more contented ($t(36) = 2.88$. p=.007) with their overall experience with the system. The differences between the two groups suggest that task complexity might have an impact on users' mental models of MedlinePlus. This possibility will be analyzed and discussed in Chapter 6. Next, we will examine user perceptions after session 2, where both groups performed the same set of tasks. Also at this point, the simple task group tended to have a better experience with the system than the complex group. For example, overall, the simple task group felt more satisfied ($t(36) = 2.15$, p=.038) and more delighted with their experience ($t(36) = 2.30$, p=.027), but the differences were not as dramatic as they were at the end of the session 1.

Finally, we will examine the differences between the two time points, within each group. In general, the simple and complex task groups showed similar developmental paths in their experience with the system. Particularly, for the simple task group, satisfaction with the content increased significantly ($t(18) = 2.34$, p=.031). The complex task group's intention to use the system in the future increased ($t(18) = 2.69$, p=.015).

The development of users' perceptions of MedlinePlus as they had more interaction with the system suggests that subjects' mental models of the system might change over time. Such changes will be analyzed and discussed in detail in Chapter 6.

5.5 Summary

This chapter reported results concerning the characteristics of subjects, tasks, and subjects' perceptions of the system. The discussion of these variables is necessary for providing context for a discussion of mental model construction. The data analysis showed that task complexity, as defined by three factors (clarity of the answer, location of the answer, and requirement for high-level cognitive activities), is a valid construct and has been successfully operationalized and manipulated in this study. Subjects in the study were not heavy users of medical information. General web search engines, WebMD, and Wikipedia were the main online resources that they referred to for medical information.

Variances in individual difference variables that have potential impact on people's mental models, specifically the age, gender, web experience, and spatial visualization ability of the two experimental groups (simple task group and complex task group), were well controlled.

Users' perceptions of MedlinePlus were affected by both task complexity and time: subjects' perceptions changed over time and the two task groups showed distinctions in some measurements. The change of users' perceptions of the system over time suggests a dynamic construction process of mental models, to a certain degree. Detailed analysis of subjects' construction of mental models of MedlinePlus is reported in the next chapter.

Chapter 6: Results and discussion: Construction of mental models

Chapter 5 described the context within which mental models were constructed. This chapter focuses on describing and discussing the process through which subjects constructed mental models of MedlinePlus. The process of mental models construction was examined by capturing and analyzing subjects' mental models of MedlinePlus (or information-rich web spaces more generally) at four time points:

T0: Subjects' initial mental models of information-rich web spaces before they saw MedlinePlus.

T1: Subjects' mental models of MedlinePlus after freely using the system for 5 minutes.

T2: Subjects' mental models of MedlinePlus after the first search session, where the simple task group performed a set of simple tasks and the complex task group performed a set of complex tasks. The session lasted about 20 minutes for each group.

T3: Subjects' mental models of MedlinePlus after the second search session, where both groups performed the same set of tasks. This common set of tasks included both simple and complex tasks. The second session lasted about 12 minutes for each group.

This chapter starts with an introduction of subjects' initial (T0) models of general information-rich web spaces. Then, subjects' mental models of MedlinePlus at T1, T2, and T3 are presented, followed by an overview of the construction of each of the three dimensions of the mental models: 1) structure, 2) evaluation and emotion, and 3) behavior. In the following several sections, the construction of the components embedded within each dimension are discussed. During the discussions, the impact of the tasks on mental model construction is examined and reported. Finally, developmental processes that emerged from the analysis of the construction of mental models are discussed.

6.1 Mental models at T0: Initial models of information-rich web spaces

As reported in the previous chapter, subjects in the study had about 10 years experience with the internet, on average. Thus, it is natural for them to have pre-existing models or expectations of various web-based systems, including information-rich web spaces. Subjects' pre-existing models of information-rich web spaces may serve as a foundation or starting point for subjects to construct their mental models of MedlinePlus, a typical information-rich web space. Thus, understanding subjects' initial models of information-rich web spaces, generally, is necessary for studying the construction process of the subjects' mental models of MedlinePlus.

In the demographic questionnaire administered at the beginning of the study protocol before they saw MedlinePlus, subjects were asked to provide written answers to two requests. The first asked them to describe characteristics of information-rich web spaces based on their experience and understanding. Two dimensions of subjects' perceptions of information-rich web spaces emerged from a content analysis of their responses to this first request: a structure dimension, in which subjects perceived various parts of the system, and an emotion and evaluation dimension, in which subjects expressed their evaluations or emotions about some of the parts that they perceived. The second request was to describe their general approaches to using this type of website. Mental models theory suggests that mental models, as a rich mental structure, encompass procedural knowledge of how to use a system to solve problems (Carroll & Olson, 1988). Therefore, behaviors or strategies when using information-rich web spaces was added as a third dimension of mental models to the two dimensions that emerged from subjects' descriptions of their perceptions and opinions of information-rich web spaces. Subjects' responses to the second request were analyzed to represent this third dimension of their mental models of information-rich web spaces.

Thus, the analysis of subjects' questionnaire responses produced a three-dimensional representation of their initial mental models of information-rich web spaces. The three dimensions are introduced separately in the following subsections.

6.1.1 The structure dimension of mental models

As has been mentioned, when representing information-rich web spaces, subjects represented various objects, parts, attributes, or mechanisms of the system. These representations of the system clustered into four components: system, content, information organization, and interface.

The *system component* was subjects' representation of the overall system, rather than any individual part of it. Subjects represented the system component of information-rich web spaces in relation to two aspects: the structure of the systems and the audience for such web spaces. For example, in talking about the structure, one subject suggested that:

> Based on my own experience and understanding, information-rich websites are defined as those that compile any amount of databases into what could be considered a data warehouse. Utilizing the warehouse search capabilities, users are directed to the individual databases, which are hyperlinked onto the site.

Another subject described information-rich web spaces as a collection of other sites.

In talking about the audience, subjects said that some information-rich websites serve novices in the field of their interest and some are directed more specifically toward experts. A couple of subjects suggested that information-rich web spaces contain a lot of useful information that can be used by a wide variety of people.

The *content component* was subjects' representations of the information contained in information-rich web spaces. Generally, subjects perceived this component as having three aspects: subject, type of information, and format of information. For the subject of the information, some subjects described information-rich web spaces as covering one particular topic, while some said they cover various topics. Some stated that the topics are general, while others said the topics are specialized. Subjects also mentioned several types of information, specifically, Q&A, help, and advertisements. In information-rich web spaces, subjects expect to see various formats of information, including images, graphics, text, videos, multimedia, and audios.

The *information organization component* was subjects' understanding of the ways in which information is organized. Generally, subjects recognized that, in information-rich web spaces, information is interlinked, often with links leading to outside sources. Sometimes, a table of contents is used to help organize information. In their initial model, instead of having a specific information organization schema, such as alphabetical or hierarchical organization, subjects tended to have only a very general idea that information was categorized in some way in these sites. For example, one subject commented that:

> Information is usually organized into categories of some kind on these websites, to make it more accessible for users. […]. The categories can be manually looked through usually.

The *interface component* was subjects' representation of the site interface. In this initial model, subjects thought about common interface elements and navigation tools, including menus, side bars, tabs, the homepage, hyperlinks, subheadings, and bold font. They also mentioned functionality made available though the interface, including search, advanced search, suggesting other pages of interest, and frequently searched links and topics.

6.1.2 The evaluation/emotion dimension of mental models

In representing objects, parts, attributes, or mechanisms of information-rich web spaces, subjects often simultaneously expressed their evaluations of or emotions about the aspects of the system that they perceived and represented. These evaluations and emotions corresponded to the four components in the structure dimension of the mental model, that is, subjects evaluated information-rich web spaces as integrated systems, in regard to content, in regard to information organization, and in regard to the interface of the systems.

Subjects' evaluations of the four components, as reported in the initial questionnaire, are listed below:

- *System*: convenient, accessible
- *Content*: a lot of info, useful, helpful, reliable, readable, updated frequently, clear, informative, unreliable, overwhelming
- *Information organization*: rigid organization
- *Interface*: easy to navigate, cluttered, distracting

6.1.3 The (expected) behavior dimension of mental models

The behavior dimension of mental models was informed by mental model theory and constructed from subjects' responses to the request (in the initial questionnaire) to describe their general approaches or strategies to using information-rich web spaces. At this point, no behavioral data were gathered, so these study results are based solely on subjects' self-reports of the behaviors they expected to exhibit when interacting with such systems. Thus, this dimension was named the expected behavior dimension. The behaviors and strategies reported by subjects included:

- Google to reach the site
- Search
- Browse
- Read citation pages
- Click random links
- Avoid watching videos

♦ Glance for graphs, pictures, and headings

As a summary, Figure 5 illustrates subjects' initial mental models of information-rich web spaces.

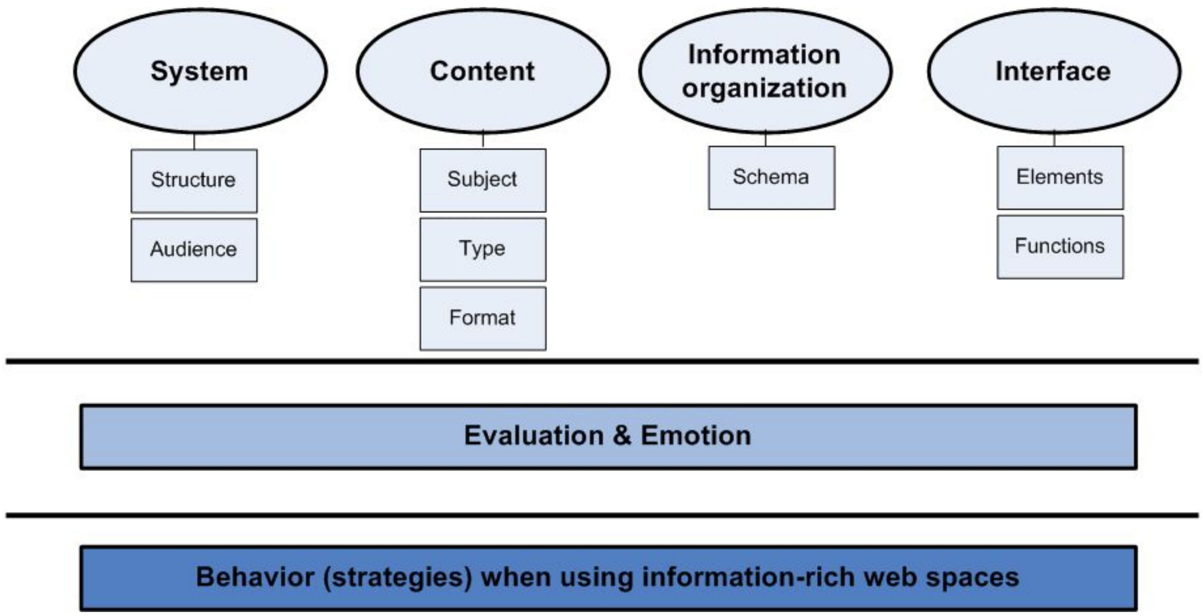


Figure 5. Overview of subjects' initial models of information-rich web spaces

As shown in the figure, subjects' initial models of information-rich web spaces had three dimensions: structure (represented in the ovals and rectangles at the top of the figure), evaluation/emotion (represented in the middle of the figure), and (expected) behaviors and strategies when using such systems. The first two dimensions emerged from a content analysis of the descriptions provided by the subjects, and the third dimension was informed by the mental model theory and constructed from subjects' descriptions of their strategies of using information-rich web spaces. In the structure dimension, subjects' representations of the system were clustered around four components (as represented by the four ovals): system, content, information organization, and interface. Different aspects of each of the four components were also described by the subjects and are shown by the rectangles underneath each oval.

6.1.4 Characteristics of the initial (T0) model

The analysis of subjects' descriptions of information-rich websites revealed that subjects held an initial model of general information-rich web spaces, before they encountered MedlinePlus, the platform used in this study. This initial model consisted of three dimensions: structure, evaluation and emotion, and (expected) behaviors when using such systems. The structure dimension consisted of four components (system, content, information organization, and interface), and attributes or aspects of those components. The evaluation/emotion dimension expressed subjects' opinions of the four components in the structure dimension. The expected behavior dimension was concerned with subjects' general approach to using such web spaces (constructed based on subjects' descriptions of their strategies of using information-rich web spaces, rather than their actual behaviors of performing tasks).

Although subjects had a mental model of general information-rich web spaces, this mental representation was general, primitive, and superficial. A typical example was that subjects' description of information organization was limited to general categorization. They were not able to articulate specific information organization schemas. The initial model also was not complete: few subjects represented search mechanisms or presentation and ranking of results. However, being general and incomplete is not necessarily a drawback for the initial model. If the initial model is a foundation for constructing mental models of a particular system, constructing mental models based on a general and primitive initial model could be easier than based on a detailed and inflexible initial model. In the next section, subjects' mental models of MedlinePlus at different time points when they interacted with the system are examined and the effects of the initial model on subjects' mental models of MedlinePlus are explored.

6.2 Mental models at T1, T2, and T3

6.2.1 Data used to create mental models at T1, T2, and T3

As introduced in Chapter 4, three methods were employed to measure mental models: concept listing protocols, semi-structured interviews, and drawings. Only data produced by the first two methods were analyzed in this study. Those two data sources are described here.

6.2.1.1 Concepts listed by subjects

Unlike the initial model that was created from descriptions provided by subjects in response to the questions in the demographic questionnaire, mental models at T1, T2, and T3 were derived from a content analysis of the concept listing data. Table 10 shows the descriptive statistics related to the concept listing protocol, including number of concepts contributed by subjects in each group and number of meaningful concepts. The total number of concepts is the number of concepts contributed by subjects in the concept listing protocols at the three different time points. Some concepts, such as "online", "technical", "organize", and "words", are too general and lack enough contextual information for reliable interpretation. These concepts were excluded from data analysis. Concepts that were included in the data analysis were considered meaningful concepts, and their frequency is shown in the third row of data.

Table 10. Descriptive statistics for concept listing protocols

	Simple task group			Complex task group		
	T1	T2	T3	T1	T2	T3
Total No. of concepts	437	451	440	435	394	394
No. of concepts/subject (SD)	23.0 (7.82)	23.7 (9.06)	23.1 (7.10)	22.8 (8.63)	20.7 (7.28)	20.6 (8.00)
Meaningful concepts	413	437	424	416	378	378

As shown in the table, at T1, the simple task group listed 437 concepts with an average of 23.0 concepts per subject, and the complex task group listed 435 concepts with an average of 22.8 concepts per subject. Plots of the number of concepts contributed by the subjects and their expected cumulative probability show that the sample data are likely to come from a normally distributed population. A *t*-test showed that the two groups were not significantly different in contributing concepts in the T1 concept listing protocol. Both groups listed about the same number of meaningful concepts. The result indicates that the two groups were comparable in articulating their thoughts about the system using the concept listing method and the computer program developed for performing the concept listing protocol.

At T2 and T3, the complex task group appeared to contribute fewer concepts than the simple task group, but the differences were not statistically significant. The concept listing protocols not only generated quantitative information as reported above, but also provided qualitative information, such as concepts and associations between concepts (indicated by the time elapsed between the generation of two concepts), which allows researchers to examine how subjects make sense of and represent a system. Subjects' mental models of MedlinePlus were induced from a semantic analysis of the concepts contributed by subjects in the concept listing protocols.

The coding unit for the content analysis was a concept. In the concept listing protocols, subjects listed single terms, phrases, and sometimes sentences. Single terms often express one concept (e.g., database), as do some phrases (e.g., alphabetical listing). However, sentences and some phrases (e.g., medical journals and search bar very useful) may contain multiple concepts. These cases were treated as multiple concepts, with one code assigned for each concept represented in the sentences or phrases.

The coding scheme was developed by exploring the meanings of a subset of the concepts. The coding scheme contained the names of the categories, definitions of the categories, and examples for each category. The scheme was then applied to code the rest of the concepts. A second coder was trained and coded about 10% of the concepts contributed by subjects. During the training, the author explained the coding scheme to the second coder and the coder familiarized herself with the MedlinePlus system before she started coding the subset of concepts.

Inter-coder agreement was calculated using Krippendorff's alpha, $\alpha = 1 - D_0/D_e$, where D_0 is a measure of the observed disagreement and D_e is a measure of the disagreement that can be expected when chance prevails (Krippendorff, 2004). The agreement (α) reached 78.4% in the first round of coding. A review session was conducted by the two coders to examine the disagreements. Most of the differences occurred in concepts referring to specific elements or content in MedlinePlus. The second coder miscoded them due to her unfamiliarity with the system. For example, in the first round coding, the second coder coded “NIH” as “system: similar sites”. After the discussion, she agreed with the author and changed the code to “system: agencies involved” because she came to know that NIH contributed content to the website in the review. After the review session, the majority of the disagreements were resolved and the inter-coder reliability (Krippendorff’s alpha) reached 95.7%.

The complete coding scheme is included in Appendix G. Here are some coding examples:

Concepts listed by subjects		**Assigned codes**
lung cancer	→	content: specific
alphabetical	→	information organization: schema
well organized	→	information organization: evaluation
tabs	→	interface: element
easy interface	→	interface: evaluation
data-rich	→	system: evaluation
CDC	→	system: agencies involved

In the semi-structured interviews that followed the concept listing protocols, subjects were asked a series of questions about their understanding of the MedlinePlus system and its working mechanisms. The interviews were transcribed and analyzed using content analysis. Because the analysis of the interviews was used only to verify the analysis of the concepts contributed by subjects in the concept listing protocols, an inter-coder reliability check was not performed. The rationale for validating the analysis of the concept listings using interview data is that semi-structured interviews have been widely used in prior research to elicit and represent people's mental models. The results of the content analysis of the interview transcripts will not be reported here. However, quotes are extracted from the transcripts to help interpret some concepts or to support some inferences.

6.2.1.2 Descriptions of procedures to solve a hypothetical task

As has been mentioned, mental model theory suggests that people's mental models encompass procedural knowledge of how to use a system to solve problems. Due to the limited ability of the concept listing protocol in eliciting subjects' procedural knowledge of how to perform tasks using a system, additional data were gathered. During the semi-structured interviews, subjects were asked to describe the steps they would use to solve a task using MedlinePlus. The same hypothetical task was used in the interviews at all three times: T1, T2, and T3. The descriptions were used to represent subjects' procedural knowledge of solving problems using MedlinePlus.

6.2.2 Subjects' mental models of MedlinePlus at T1

When subjects were first introduced to MedlinePlus, they were asked to take 5 minutes to explore the system in the way that they normally would when encountering a new website. After the 5-minute free exploration, they were asked to finish the first set of mental model measurements (a concept listing protocol, a semi-structured interview, and drawing). The mental models at T1 were the models that subjects had about MedlinePlus after freely using the system for 5 minutes. At T1, subjects in the two groups had completed the same tasks (spatial ability, demographic questionnaire, and 5 minutes of free exploration of the system), and they showed equal ability to articulate concepts about MedlinePlus and to use the concept listing tool (reflected by the fact that the two groups contributed similar numbers of concepts in the concept listing protocols). Therefore, the concepts contributed by subjects across the two groups in the concept listing protocol were combined to create their mental model of MedlinePlus at T1.

The analysis of the concepts that subjects contributed in the concept listing protocol at T1 generated a two-dimension and four-component framework that is similar to subjects' initial models of information-rich web spaces. The two dimensions are the structure dimension and the evaluation/emotion dimension. The four components within each dimension are system, content, information organization, and interface. Subjects' descriptions of the steps they would use to solve a hypothetical task using MedlinePlus, elicited in the semi-structured interview, were used to represent the subjects' procedural knowledge of using the system to solve problems. Informed by mental model theory, subjects' procedural knowledge was added as a third dimension, the (expected) behavior dimension, to the two dimensions that emerged from the concept listing data. Figure 6 illustrates subjects' mental models of MedlinePlus at T1.

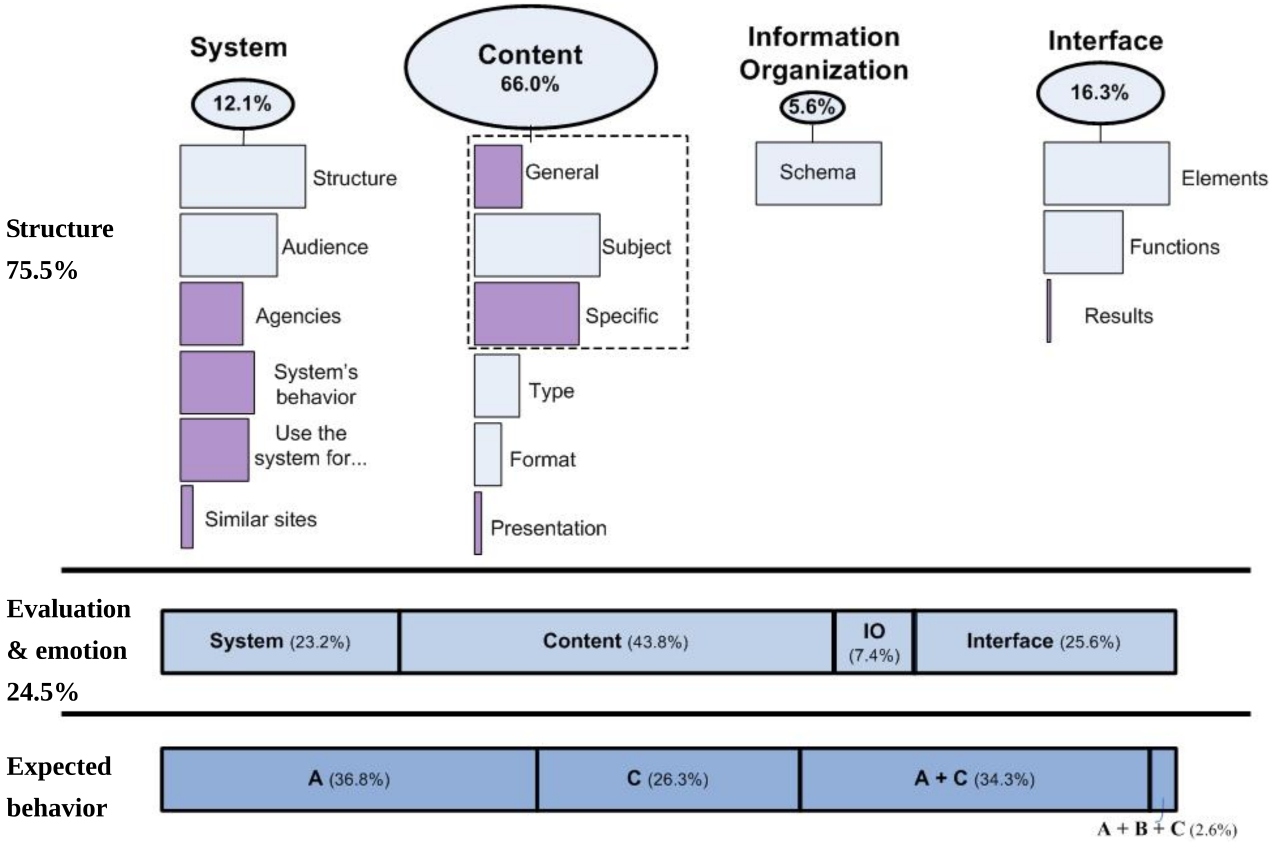


Figure 6. Overview of subjects' mental models of MedlinePlus at T1

As shown in the figure, 75.5% of the concepts (626 concepts) contributed by subjects were related to the structure dimension of the mental model. This dimension encompassed four components: system, content, information organization, and interface. They are depicted as four ovals at the top of the figure. The four components were consistent with subjects' initial model of general information-rich web spaces. However, they were not equally represented. Subjects dedicated the majority of the concepts (413 concepts, 66.0%) to represent the content component, followed by the interface (102 concepts, 16.3%) and the system (76 concepts, 12.1%) components. The information organization component was least represented (35 concepts, 5.6%).

Subjects represented each component from various aspects. The aspects are depicted as bars under each oval. Subjects inherited some aspects from the initial model (denoted by bars in light blue). They also came up with new aspects (denoted by bars in purple) of some components. The lengths of the bars represent the relative frequency of appearance of the aspects in each component. The first three bars of the content component are surrounded by a dotted-line box because they represent one aspect, the subject of the content of MedlinePlus, at three different levels. The specific aspects of each component will be discussed in more detail in sections 6.4-6.7.

About 24.5% of the concepts (203 concepts) contributed by subjects were related to the evaluation/emotion dimension. The evaluations and emotions were also clustered around the four components found in the structure dimension. As shown in the figure, subjects evaluated the content the most, followed by the interface and system components. They evaluated the information organization component the least. *T*-tests showed that the two groups are not significantly different in the number of concepts that they used to represent the two dimensions (structure and evaluation/emotion) and four components (system, content, information organization, and interface) in the dimensions. The results further validate that at T1, the two groups could be combined for the analysis.

The (expected) behavior dimension of the mental model was informed by mental model theory and constructed from the strategies that subjects reported they would use to solve a hypothetical task (rather than subjects' actual behavior of performing the task using

MedlinePlus). After interacting with MedlinePlus for 5 minutes, subjects came up with three distinct strategies (A, B, and C) for finding information in the system. Strategy A is a general search strategy, where subjects type queries into the search bar and review the results returned by the system. Strategy C is a general browsing strategy, where subjects follow links on the homepage to health topic pages, where all the information pertaining to a particular disease, condition, or drug is listed. Strategy B is a combination of searching and browsing. Instead of following a chain of links to reach the health topic page, subjects type a search query in the search bar, click on the top result to reach the health topic page pertaining to the disease or condition that is of concern, and then browse by following links on the health topic page. Because strategy B is more closely aligned to the capabilities of MedlinePlus than the more general strategies A and C, it can be considered a more sophisticated method for accessing information in MedlinePlus.

As reflected in the figure, the majority of the subjects used the general search strategy (strategy A), general browsing strategy (strategy C), or the combination of the two. Only one subject planned to use the more sophisticated method (strategy B) to access information. The results showed that, after a brief interaction with MedlinePlus, the majority of the subjects expected to employ general search and browsing strategies to solve problems in this website. Few of them were able to identify more sophisticated ways to access information in the system. Fisher’s exact test shows that the two groups are also not significantly different in the strategies that they would use to solve the hypothetical task.

Compared to subjects' initial mental models of information-rich web spaces, subjects' mental models at T1 became more specific, incorporating aspects or attributes that relate specifically to MedlinePlus, such as agencies that contribute to the system and the behavior of the system (e.g., pop-up windows). Subjects also began to form opinions of what the system can be used for (e.g., for background information). Meanwhile, they used similar sites (e.g., WebMD) to help them understand the system. After a brief interaction with the system, the strategies that users planned to use to solve problems using the system remained general, primarily general search and browsing strategies. It is clear that in order to reveal the level of understanding subjects had about MedlinePlus, the content of the aspects of each component and subjects' evaluation and emotions about them need to be examined. A detailed analysis will be reported in sections 6.4-6.7.

6.2.3 Subjects' mental models of MedlinePlus at T2

After exploring the system for 5 minutes and finishing the first set of mental model measurements (a concept listing protocol, a semi-structured interviews, and drawing) (T1), subjects were randomly assigned to two groups, the simple task group and the complex task group. The simple task group performed 12 simple tasks and the complex task group performed 3 complex tasks. Each group spent about 20 minutes finishing the corresponding task set. After finishing the tasks, subjects were asked to take the same mental model measurements (a concept listing protocol, a semi-structured interviews, and drawing) for a second time (T2).

Similar to the mental model representation at T1, an overview of subjects' mental models of MedlinePlus at T2 was constructed using both bottom-up and top-down approaches. The structure and evaluation/emotion dimensions were induced from the analysis of concept listing data. The (expected) behavior dimension was informed by mental model theory and constructed from subjects' descriptions of steps that they would take to solve a hypothetical task (rather than subjects' actual behavior of search for the task). Figure 7 and Figure 8 depict, respectively, the simple and complex task groups' mental models of MedlinePlus at T2.

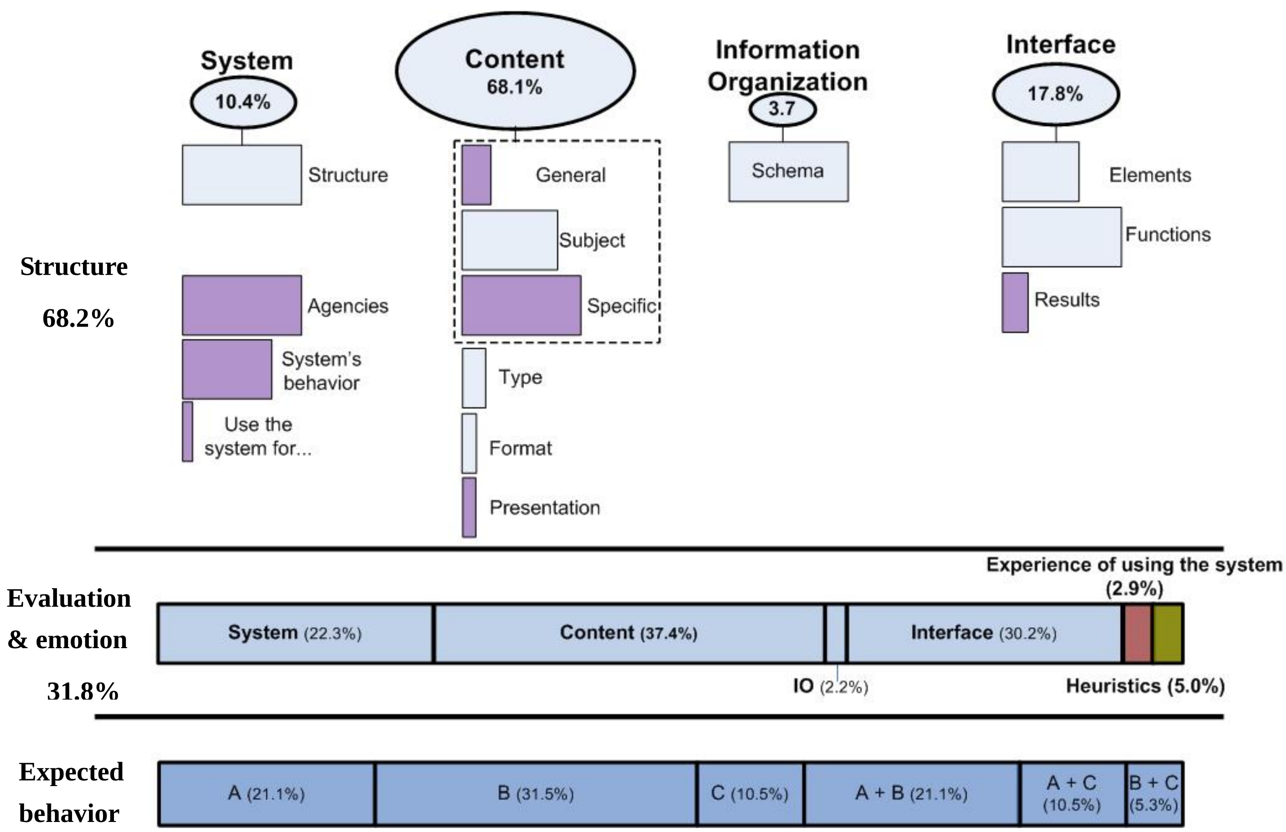


Figure 7. Overview of the simple task group's mental models of MedlinePlus after Session 1 (T2)

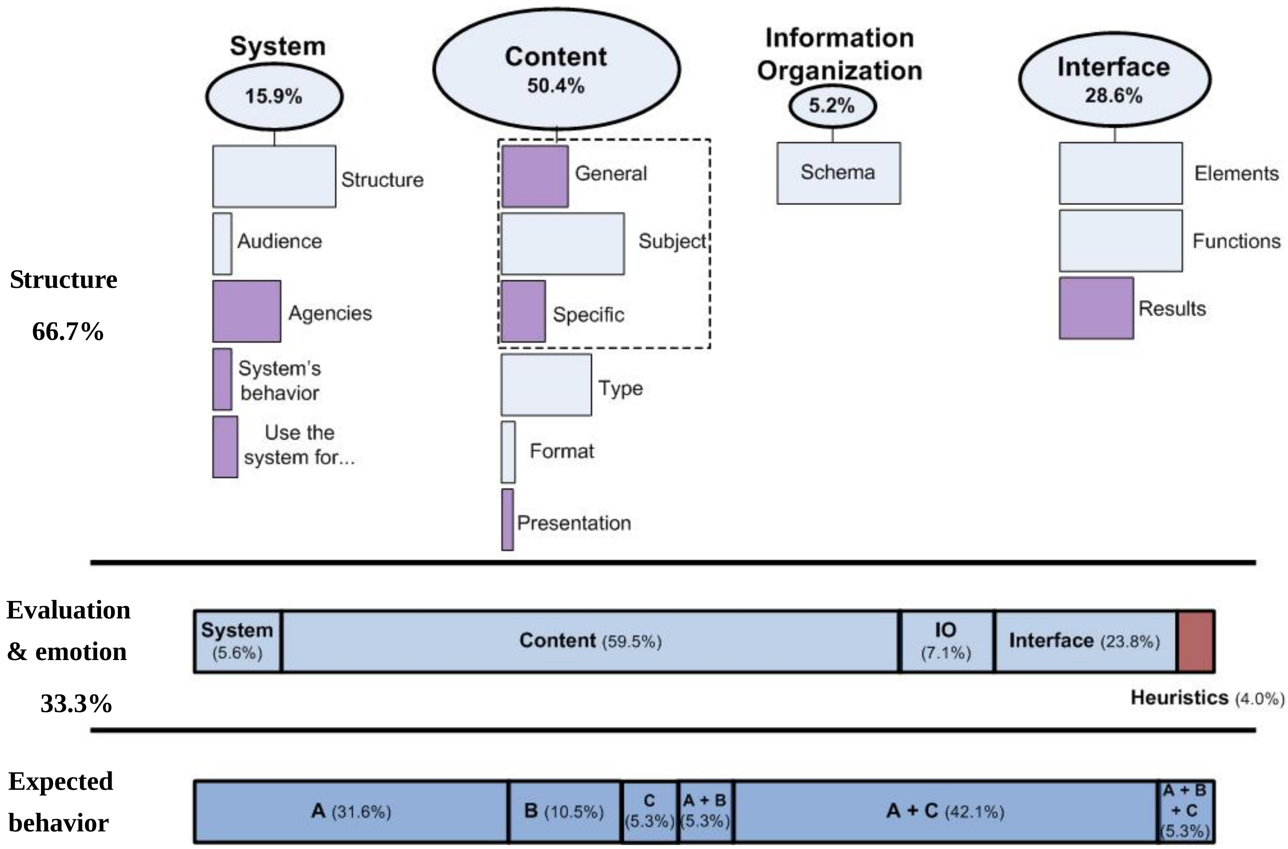


Figure 8. Overview of the complex task group's mental models of MedlinePlus after Session 1 (T2)

As shown in the two figures, at T2, after subjects had used MedlinePlus to solve a set of tasks, their mental models demonstrated the same dimensions as were found at T0 and T1, with structure, evaluation and emotion, and (expected) behavior dimensions. The simple task group contributed 68.2% of their concepts (298 concepts) and the complex task group contributed 66.7% of their concepts (252 concepts) to the structure dimension. The rest were contributed to the evaluation/emotion dimension. *T*-tests showed that the two groups were not significantly different from each other in the number of concepts contributed to the two dimensions.

Within the structure dimension, subjects' representations of MedlinePlus continued to cluster around the four components: system, content, information organization, and interface. Consistent with T1, both groups dedicated more than half of the structure-related concepts to the content of MedlinePlus, followed by the interface, system, and information organization components. Subjects continued to represent each component from various aspects. Light blue bars are aspects that appeared in the subjects' initial models of information-rich web spaces. Purple bars are aspects that did not appear in the subjects' initial models (i.e., they appeared for the first time at T1). The lengths of the bars represent the relative frequency of appearance of the aspects in each component. The first three bars of the content component are surrounded by a dotted-line box because they represent one aspect, the subject of the content of MedlinePlus, at three different levels. The changes in subjects' representations of the structure dimension between T1 and T2 will be discussed in section 6.3. The changes in the representations of each aspect of the four components within the structure dimension between T1 and T2 will be discussed in detail in sections 6.4-6.7.

Within the evaluation/emotion dimension, subjects continued to evaluate or express emotions about the four components in the structure dimension. However, it is worth noting that, subjects in both groups began to develop heuristics for using the system (shown as the red bar in the evaluation/emotion dimension). Heuristics are subjects' perceptions concerning what is good to do, what is not, what is easy, and what is difficult to do with the system. For example, one subject commented that "typing in a question does not work well". Meanwhile, subjects in the simple task group also began to reflect on their experience of using the system (shown as the yellow bar in the evaluation/emotion dimension). For example, a subject described experiencing "a little frustration" when using the system. The changes in subjects' representations of the evaluation/emotion dimension between T1 and T2 will be discussed in section 6.3. The changes in the representations of each component within the evaluation/emotion dimension between T1 and T2 will be discussed in detail in sections 6.4-6.8.

Within the (expected) behavior dimension, subjects continued to use general searching (strategy A), general browsing (strategy C), and searching then following the first result to the health topic page (strategy B). For both groups, more subjects -- 11 (57.9%) subjects in the simple task group and 4 (21.1%) in the complex task group -- adopted the more sophisticated strategy B. In terms of the use of combined strategies, 10 subjects (52.7%) in the complex task group planned to use combined methods, compared to 7 (36.8%) in the simple task group. The change of subjects' strategies of solving the hypothetical task (expected behaviors) between T1 and T2 will be discussed in section 6.3.

6.2.4 Subjects' mental models of MedlinePlus at T3

After finishing their corresponding tasks and the mental model measurements, both simple and complex task groups were asked to complete a common set of tasks, which included 4 simple tasks and 2 complex tasks. Both groups took approximately the same time to finish the tasks (simple task group: 11.7 minutes, complex task group: 13.0

minutes). After finishing this common set of tasks, subjects were asked to take the same mental model measurements (a concept listing protocol, a semi-structured interview, and drawing) a third time. Figure 9 and Figure 10 depict, respectively, the simple and complex task groups' mental models of MedlinePlus at T3.

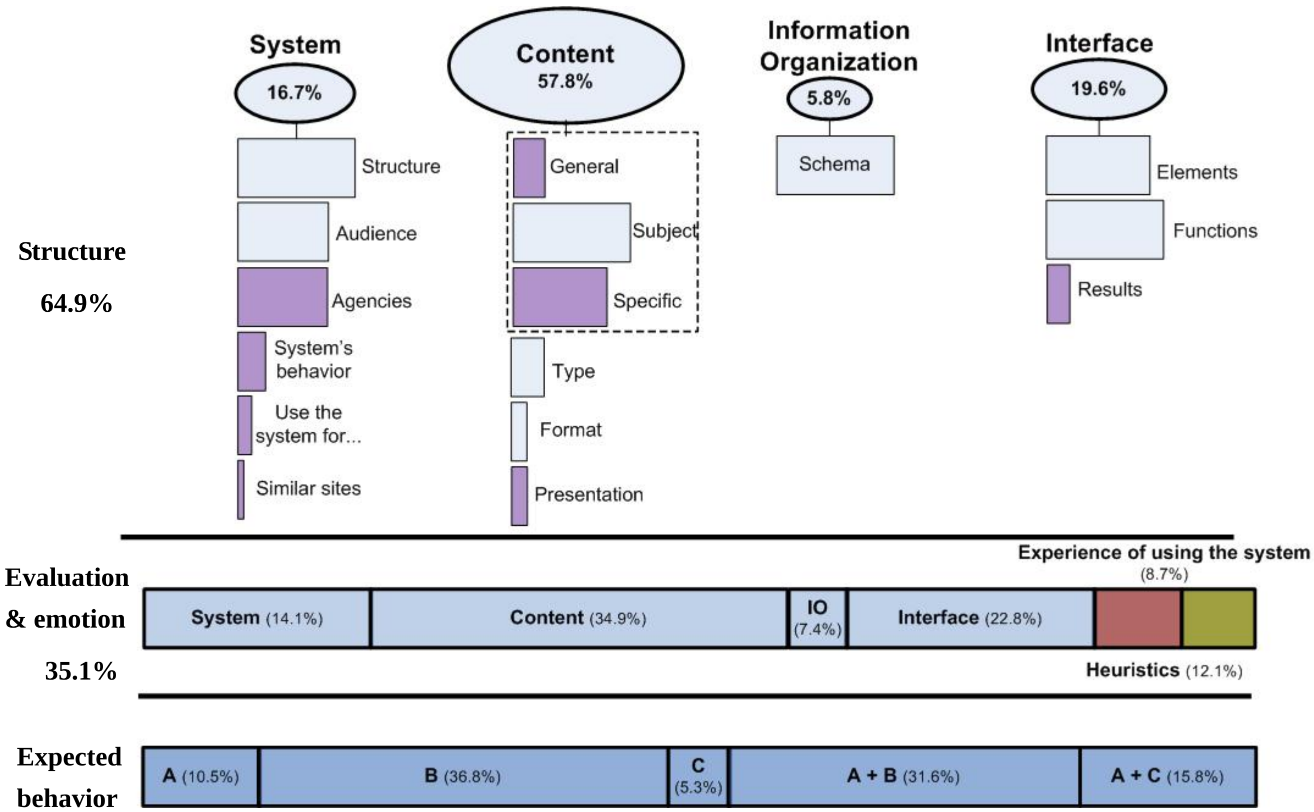


Figure 9. Overview of the simple task group's mental models of MedlinePlus after Session 2 (T3)

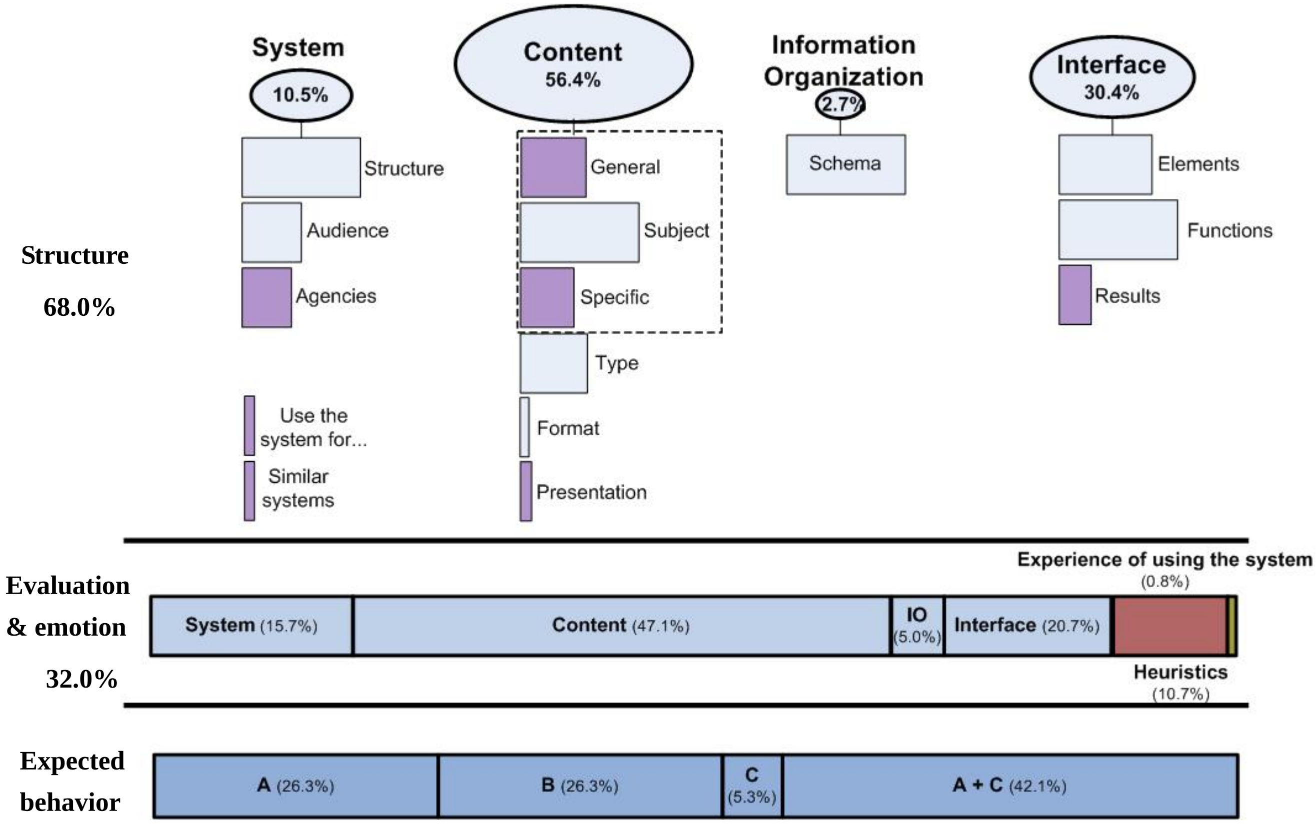


Figure 10. Overview of the complex task group's mental models of MedlinePlus after Session 2 (T3)

As shown in Figure 9 and Figure 10, both groups' mental models of MedlinePlus at T3 still consisted of three primary dimensions: structure, evaluation/emotion, and behaviors. The first two dimensions of the models were induced from the concept listing data at T3. The (expected) behavioral dimension was informed by mental model theory and constructed from subjects' descriptions of the steps that they said they would take to solve the same hypothetical task (rather than their actual behaviors of performing the task).

At T3, the simple task group contributed 64.9% of their concepts (275 concepts) and the complex task group contributed 68.0% of their concepts (257 concepts) to the structure dimension. The rest of the concepts were dedicated to the evaluation/emotion dimension. *T*-tests show that the two groups were not significantly different from each other in the number of concepts contributed to the two dimensions.

Within the structure dimension, similar to T1 and T2, subjects' representations of the system clustered around four components: system, content, information organization, and interface. Also consistent with T1 and T2, both groups dedicated more than half of the concepts to the content of MedlinePlus, followed by the interface and system. Information organization remained the least represented component. Subjects continued to represent each component from various aspects. Light blue bars are aspects that appeared in the subjects' initial (T0) model of information-rich web spaces. Purple bars are aspects that did not appear in subjects' initial (T0) model of information-rich web spaces. The lengths of the bars represent the relative frequency of appearance of the aspects in each component. The first three bars of the content component are surrounded by a dotted-line box because they represent one aspect, the subject of the content of MedlinePlus, at three different levels. The changes in subjects' representations of the structure dimension between T2 and T3 will be discussed in section 6.3. The changes in the representations of each aspect of the four components between T2 and T3 will be discussed in detail in sections 6.4-6.7.

Within the evaluation/emotion dimension, subjects continued to evaluate and express emotions about the four components in the structure dimension. Similar to T2, they also represented heuristics for using the system (shown as the red bar) and their experience of using the system (shown as the yellow bar). The changes in subjects' representations of the evaluation/emotion dimension between T2 and T3 will be discussed in section 6.3. The changes in the representations of each component within the dimension between T2 and T3 will be discussed in detail in sections 6.4-6.8.

Within the (expected) behavior dimension, subjects continued to employ general searching (strategy A), general browsing (strategy C), and searching then following the first result to the health topic page (strategy B). Compared to T2, more subjects in both groups planned to use the more sophisticated strategy B: 13 (68.4%) subjects in the simple task group and 5 (26.3%) in the complex task group proposed to use it. In terms of the use of combined strategies, 9 subjects (47.4%) in the complex task group planned to use combined methods, compared to 8 (42.1%) in the simple task group. The change of subjects' strategies (expected behaviors) for solving the hypothetical task between T2 and T3 will be discussed in section 6.3.

6.3 The construction of the mental models' three dimensions

In the previous two sections, subjects' initial model of information-rich web spaces and their mental models of MedlinePlus at three time points – after 5 minutes of free exploration (T1), after the first search session (T2), and after the second search session (T3) – were introduced. This section focuses on providing an overview of the changes that subjects' mental models experienced at the dimension level as those mental models were constructed over time. The effects of task complexity on mental model construction at the dimension level are also discussed. Details of the changes and development of each component within the first two dimensions will be reported and discussed in detail in sections 6.4-6.8.

6.3.1 The construction of the structure dimension of mental models

At T1, subjects contributed 626 concepts (75.5% of all the concepts listed by subjects at T1) to describe the structure dimension of subjects' mental models of MedlinePlus. At T2, both groups contributed fewer concepts to the structure dimension (simple task group: 298 concepts, 68.2% of concepts that the group listed; complex task group: 252 concepts, 66.7%). The reduction of the complex task group between T1 and T2 was statistically significant ($t(18)= 2.74$, $p=0.013$). At T3, the number of concepts that contributed to the structure dimension was consistent with what each group contributed at T2.

In this study, the subjects were novice users of MedlinePlus. Thus, it is not surprising that, at the beginning of the interaction, they tried to understand the system and contributed the majority of the concepts (75.5%) to represent various aspects of the system. As they used the system to solve problems and became familiar with the system, it might have become less necessary for them to represent certain elements, features, or functions of the system. However, it is possible that, after subjects gained a certain amount of experience with the system, their attention to and representations of the structure dimension would stabilize. In future research, it would be interesting to study how subjects represent a system with which they are familiar and examine the number of concepts dedicated to the structure dimension of users' mental models of the system.

The structure dimension of subjects' mental models at different times (T1, T2, and T3) all encompassed four components: system, content, information organization, and interface. Table 11 shows the composition of the structure dimension of subjects' mental models of MedlinePlus in terms of its four components, at different time points. The numbers in the columns are the number of concepts that subjects used to represent each component and the numbers in parentheses are the percentages of each component in the structure dimension of the subjects' mental models.

Table 11. Number of concepts in each component of the structure dimension of subjects' mental models of MedlinePlus over time

	T1	T2		T3	
		Simple	Complex	Simple	Complex
System	76 (12.1%)	**31 (10.4%)**	40 (15.9%)	**46 (16.7%)**	27 (10.5%)
Content	**413 (66.0%)**	203 (68.1%)	**127 (50.4%)**	159 (57.8%)	145 (56.4%)
IO	35 (5.6%)	11 (3.7%)	13 (5.2%)	16 (5.8%)	7 (2.7%)
Interface	102 (16.3%)	53 (17.8%)	72 (28.6%)	54 (19.6%)	78 (30.4%)
Total	**626**	298	**252**	275	257

Bold face indicates that differences across time (compared with the previous time period) were statistically significant ($p<.05$).

As shown in the table, the four components were not equally represented across time. The content component was consistently the most heavily-weighted component in the mental models across the three points in time. At T1, after subjects freely used the system for 5 minutes, 66.0% of the structure dimension of their mental model was related to content. At T2, after the first search session, where the simple task group performed 12 simple tasks and the complex task group performed 3 complex tasks, the simple task group gave equal emphasis to the content as at T1, while the complex task group significantly decreased the number of concepts to represent the content component ($t(18) = 3.12$, $p =$

0.006). The different developmental trends for the two task groups might be because tasks (or task scenarios) are a prominent source for concepts concerning the content component. The simple task group received more tasks than the complex task group, and therefore they were able to come up with more concepts concerning the content of MedlinePlus. At T3, after both groups performed the same set of tasks, they produced similar numbers of concepts representing the content component, which supports the speculation that tasks could affect subjects' mental models of MedlinePlus. A close look at the meaning of the concepts representing the content component (to be discussed in section 6.5) will provide more insights into the possible impact of task complexity.

The next most heavily-weighted component in the structure dimension of the mental models is the interface. At T1, 16.3% of the structure dimension of the mental model was related to the interface. Both groups incrementally gave more weight to the interface component at both T2 and T3. It might be natural for subjects to increase their emphasis on the interface over the two search sessions, because they were in the process of learning about the interface. However, the differences in representing the interface component across time and between groups were not statistically significant. The qualitative features of the interface components and their development over time will be discussed in section 6.7.

The system component is the third emphasized component in the structure dimension of subjects' mental models of MedlinePlus. For this component, both the simple and the complex task groups did not significantly change their emphasis from T1 to T2. While at T3, the simple task group contributed significantly more concepts to represent the component than at T2 ($t(18) = 2.51$, $p = 0.022$). The concepts that were employed to represent the system component and the development of the system component will be analyzed in more detail in section 6.4.

For the information organization component, the two groups did not show significant differences in terms of the number of concepts used to represent the component across time (T1, T2, and T3). Between groups, there were also no significant differences at each time. It is worth pointing out that, across time, information organization was consistently the least represented component in the structure dimension of subjects' mental models of MedlinePlus. This pattern could be due to two reasons. First, it is natural for people to pay little attention to the information organization of an information retrieval (IR) system (no matter how well or poorly the information is organized). Second, information in MedlinePlus is well organized and subjects did not have encounter great difficulty in moving around in its information space. Thus, they did not need to represent information organization intensively in their mental models. If the information organization in MedlinePlus were poorly designed and hindered effective retrieval, subjects might have paid more attention to the component. Again, further inspection of the content of the information organization component, in section 6.6, will provide further insights.

From the discussion of the construction of the structure dimension of subjects' mental models of MedlinePlus, it is clear that in order to understand the development of subjects' mental models over time and the differences between groups, it is necessary to examine the changes in each component, not only the changes of emphases that subjects put on each component, but also the changes in the semantic content of the components. The figures above (Figures 6-10) illustrate that, at each point in time, subjects represented each component from different aspects. Some aspects were phased out and some were added in at different times. The analysis of the changes of subjects' mental models at the component level will be performed and reported in the next several sections (section 6.4-6.8).

6.3.2. The construction of the evaluation/emotion dimension of mental models

At T1, subjects contributed 203 concepts (24.5% of all their concept listings) to evaluate or express emotions about various aspects of the system. At T2, both groups increased their emphases on this dimension (simple task group: 139 concepts, 31.8%; complex task group: 126 concepts, 33.3%). The increase seen in the simple task group was statistically significant ($t(18) = 2.38$, p=0.028). The increased representation of evaluations of and emotions about the system might be because people are more able to evaluate a system when they have learned to use it. At T3, the number of concepts contributed to this dimension for both groups did not show significant differences from what they were at T2.

The evaluation/emotion dimension of subjects' mental models of MedlinePlus at T1 encompassed four components (corresponding to the structure dimension): system, content, information organization, and interface. At T2, subjects also began to represent heuristics for using the system and experience with the system. Table 12 shows the distribution of concepts for each component at different points in time. The numbers in the columns represent the numbers of concepts that subjects used to describe each component and the numbers in parentheses are the percentages of each component in the evaluation/emotion dimension of the subjects' mental models.

Table 12. Number of concepts in each component of the evaluation/emotion dimension of subjects' mental models of MedlinePlus over time

	T1	T2		T3	
		Simple	Complex	Simple	Complex
System	47 (23.4%)	31 (22.3%)	**7 (5.6%)**	21 (14.1%)	**19 (15.7%)**
Content	**89 (43.8%)**	52 (37.4%)	**75 (59.5%)**	52 (34.9%)	57 (47.1%)
Information organization	15 (7.4%)	3 (2.2%)	9 (7.1%)	11 (7.4%)	6 (5.0%)
Interface	52 (25.6%)	42 (30.2%)	30 (23.8%)	34 (22.8%)	25 (20.7%)
Heuristics		7 (5.0%)	5 (4.0%)	18 (12.1%)	13 (10.7%)
Experience		4 (2.9%)		13 (8.7%)	1 (0.8%)
Total	**203**	**139**	126	149	121

Bold face indicates that differences across time (compared with the previous time period) were statistically significant (p<.05).

As shown in the table, the six components in the dimension were not equally represented. Consistent with the fact that the content component was the most represented component in the structure dimension of the mental models across time, it was also the most evaluated component across time by both groups. The result is not surprising because it is reasonable to expect that subjects cared the most about the content of a system.

Subjects' evaluations of the content of MedlinePlus were affected by task complexity. As shown in the table, at T2, the complex task group devoted significantly more concepts to evaluate the content component than at T1 ($t(18) = 2.98$, $p = 0.008$), while the number of concepts produced by the simple task group remained similar to T1. The different development paths of the two groups might be related to the level of difficulty in finding information for tasks. For the complex tasks, subjects needed to find information from multiple places in MedlinePlus and integrate the information to form answers. The challenges imposed by this requirement made it natural for subjects in the complex task group to express more feelings and evaluations about the content. In contrast, for the simple tasks, the answers were located at one place and easy to recognize. Therefore, subjects in the simple task group produced fewer evaluations of the content. Examination of the actual concepts that subjects used to evaluate content could shed more light on the impact of tasks (see section 6.5).

The second most evaluated component in the evaluation/emotion dimension was the interface component. The two groups did not show significant differences in representing this component over time and between groups.

For the system component, at T2, the two groups were significantly different in evaluating the system component, with the simple task group devoting significantly more concepts to evaluating the system ($t(36) = 2.72$, $p = 0.029$). It is possible that the complex task group's focused attention on the content component diluted their attention to the system component. In other words, the difficulty associated with finding answers to the complex tasks led subjects in the complex task group to focus intensively on evaluating task-related elements: content and information organization in the system (content and information organization components), rather than MedlinePlus as an integrated system (the system component).

At T3, after performing the second set of tasks, the complex task group significantly increased their evaluations of the system component ($t(18) = 2.47$, $p = 0.024$). This result supports the speculation that subjects' attention to the system component is related to difficulty in finding answers for tasks: as the tasks became easier for the complex task group, subjects in the group increased their representations of the system component.

The information organization component was consistently the least evaluated component (except in the complex task group's mental models at T2) among the four components in the evaluation/emotion dimension. This result might be directly related to the fact that this component was the least represented component in the structure dimension of the mental models. The two groups did not show significant differences across time or between groups in evaluating this component.

At T2, after they had performed a set of tasks (either simple or complex), subjects augmented the evaluation/emotion dimension by incorporating heuristics of using the system and their self-reflections on their experience with the system. The number of concept listings related to heuristics increased slightly from T2 to T3 for both groups, but the increase is not statistically significant. Between groups, at both T2 and T3, the two groups listed about the same number of heuristics.

In expressing self-reflections on their experience with using the system, compared to T2, subjects in the simple task group contributed more concepts to this aspect (from 4 concepts at T2 to 13 concepts at T3). Between groups, at T3, it seemed that subjects in the simple task group were more likely to evaluate their experience with using the system than the subjects in the complex task group.

It is clear that, in order to make sense of the construction of the evaluation/emotion dimension over time and the differences in the construction of this dimension between the groups, it is necessary to inspect the semantic development of evaluations of each component, including heuristics and subjects' self-reflections on their experience with using the system. The analysis will be performed and reported in the next several sections (section 6.4-6.8), along with the analysis of the construction of the components in the structure dimension.

6.3.3 The construction of the (expected) behavior dimension of mental models

This dimension of mental models was derived from mental model theory and constructed based on subjects' descriptions of steps that they would take to solve a hypothetical task using MedlinePlus (not their actual behavior of solving the task using MedlinePlus). Three search strategies were identified based on subjects' descriptions:

1) Strategy A: Search → read results
2) Strategy B: Search → top result → health topic page → review content and links
3) Strategy C: Browse: Health issues → health topic page → review content and links

As has been mentioned, strategy A is a general search strategy. Strategy B is a combination of search and browsing. It is more aligned to the capabilities of MedlinePlus, so it is considered a more sophisticated way to access information in MedlinePlus. Strategy C is a general browsing strategy. Table 13 lists search strategies that both groups planned to use to look for information for the same task at different times. The numbers are the number of subjects who planned to use each of the strategies.

Table 13. Search strategies to be used for solving the hypothetical task for the two groups at T1, T2, and T3

Strategies	Simple task group			Complex task group		
	T1	T2	T3	T1	T2	T3
A	9	4	2	5	6	5
B		6	7		2	5
C	5	2	1	5	1	1
A+B		4	6		1	
A+C	5	2	3	8	8	8
B+C		1				
A+B+C				1	1	

As shown in the table, at T1, subjects in both groups planned to use general search strategy A, general browsing strategy C, or the combination of A and C to solve the hypothetical task. Only one subject in the complex task group planned to use the combined strategy of A, B, and C. Fisher's exact test shows that, at T1, the two groups did not show significant differences in planning use of strategies to solve the hypothetical task. The result suggests that in the first 5-minute interaction with the MedlinePlus system, subjects had difficulties in identifying the more sophisticated strategy B to access information.

In order to investigate the use of individual specific strategies, subjects' overall expected uses of search strategies A, B, and C are listed in Table 14. In this table, if a subject expected to use multiple strategies, the subject is counted multiple times. Use of each of these three strategies will be discussed in turn.

Table 14. Expected use of search strategies A, B and C for the two groups at T1, T2, and T3

Strategies	Simple task group			Complex task group		
	T1	T2	T3	T1	T2	T3
A	14	10	11	14	16	13
B		11	13	1	4	5
C	10	5	4	14	10	9

At T1, 14 (73.7%) subjects in each group planned to use general search strategy A to approach the hypothetical task. At T2, the two groups showed significant differences in the use of general search strategy A. A Fisher's exact test shows that subjects in the complex group were more likely to use strategy A than those in the simple task group (p=0.039). The differences in subjects' use of strategy A at T2 might be due to the different tasks that they had performed. The paths to answers for simple tasks tended to help subjects in the simple task group to recognize strategy B. Therefore, at T2 many subjects in the simple task group diverted from strategy A to strategy B to solve the hypothetical task. Because complex tasks required subjects to synthesize information from different places, they tended to encourage subjects in the complex task group to continue to use the general search strategy A. Thus, subjects in the complex task group were more likely to use strategy A at T2 to solve the hypothetical task. At T3, the two groups did not show significant differences in the use of strategy A.

The second row of the table describes subjects' expected use of strategy B. Across time, both groups were more and more likely to use strategy B. At T1, only one subject planned to use strategy B. At T2, 11 subjects in the simple task group and 4 in the complex task group planned to use the strategy. At T3, slightly more subjects in both groups planned to adopt strategy B. The data suggest that subjects in both groups tended to accept strategy B over time, but the acceptance rate was different. Fisher's exact test shows that, at both T2 and T3, the simple task group was more likely to use the more sophisticated strategy B to solve the hypothetical task than the complex task group (T2: p=0.022; T3: p=0.011). Prior to this study, it was hypothesized that subjects who performed complex tasks during the first search session would develop better mental models and thus would come up with better strategies for solving problems using the system. However, the results show that the simple task group was more likely to use the more sophisticated strategy. As mentioned earlier, the paths to the answers for simple tasks help users recognize the more system-specific strategy B. Therefore, it is reasonable to speculate that it is not the complexity of the task that impacts the adoption of a more sophisticated strategy; rather, it is the ability of the tasks to help users reveal the structure of information organization in the system that has a significant impact on the choice of strategies.

The third row in the table shows subjects' use of the general browsing strategy C. At T1, the majority of the subjects (10 in the simple task group and 14 in the complex task group) planned to use the general browsing strategy C. At T2, 5 subjects in the simple task group and 10 in the complex task group planned to use the strategy. At T3, fewer subjects planned to adopt the strategy. The two groups did not show significant differences in the adoption of strategy C at any of the three times. It is worth pointing out that the design of the interface of MedlinePlus encouraged the use of a general browsing strategy. However, over time, when they had more experience using the system to solve problems, subjects in both groups became less and less likely to use the strategy. The reasons why subjects became less favorable to browsing over time need more investigation.

6.3.4 Summary and discussion

This section illustrated the construction of subjects' mental models of MedlinePlus and the changes that the mental models experienced over time at the dimension level. The results showed that subjects constructed their mental models of MedlinePlus based on their initial model of general information-rich web spaces. Subjects' mental models of MedlinePlus consistently have three dimensions: structure, evaluation/emotion, and (expected) behavior. The first two dimensions emerged from the subjects' concept listings and the third dimension was informed by mental model theory and constructed from subjects' descriptions of steps they planned to use to solve a hypothetical task (rather than their actual behaviors of solving the task).

The structure dimension of the subjects' mental models encompassed four main components: system, content, information organization, and interface. The content component was consistently the most represented component over time, followed by the interface and system components. Information organization was consistently the least represented. The evaluation/emotion dimension of mental models encompassed subjects' evaluations of the same four components identified in the structure dimension (system, content, information organization, and interface). At T2, after using the system to complete several tasks, subjects began to represent heuristics for using the system and to reflect on their experience with using the system. As subjects' experience with MedlinePlus increased, they decreased their representations of the structure dimension and correspondingly increased their emphasis on the evaluation/emotion dimension. Subjects proposed three strategies to look for information in MedlinePlus for a hypothetical task: general search strategy A, general browsing strategy C, and a system-specific strategy B. Over time, as subjects used the system more, they planned to use more sophisticated system-specific strategy B to access information.

Tasks might have an impact on the development of subjects' mental models at the dimension level. At T2, after the simple task group performed a set of simple tasks and the complex task group performed a set of complex tasks, the complex task group significantly reduced their representations of the content component in the structure dimension while increasing the content component in the evaluation/emotion dimension. However, for the simple task group, the comparative weights of the content component in both dimensions remained similar to T1.

Tasks might also have an effect on subjects' strategies for using the system to solve problems. At T2, the complex task group was more likely to plan to adopt the general search strategy A than the simple task group, while the simple task group was more likely to plan to adopt the more system-specific strategy B than the complex task group. These differences between the two groups in planning strategies for solving a task might be because the paths to the answers of the simple tasks assigned in the first search session helped reveal the structure of the site, therefore encouraging the use of strategy B, whereas the need to synthesize information from different places to answer the complex tasks encouraged the use of the general search strategy.

Discussions of the changes of mental models at the dimension level only provide a general overview of the subjects' construction of mental models of MedlinePlus. To understand subjects' mental model construction during their interaction with MedlinePlus at a more in-depth level, it is necessary to examine the changes that mental models experienced at the component level, that is, how various aspects of the components in each dimension change over time. In sections 6.4-6.7, the subjects' representations of each of the four components, system, content, information organization, and interface, in the structure and evaluation/emotion dimensions, and the changes of these representations over time will be discussed. The two additional components in the evaluation/emotion dimension, heuristics and subjects' self-reflections on their experience of using the system, and their changes over time will be discussed in section 6.8.

6.4 The construction of the system component of the structure and evaluation/emotion dimensions

The system component of the subjects' mental models represents MedlinePlus as an integrated system, rather than any particular part or section of the system. This section starts by describing the aspects of the system component in the structure dimension of subjects' mental models of MedlinePlus at three different points in time. The descriptions demonstrate the changes that subjects' representations of MedlinePlus as an integrated system experienced over time. The second part of the section describes changes in subjects' evaluative and emotional responses to the system component over time, which demonstrates the changes that subjects' evaluations of the system experienced over time. At the end of each part, a discussion of the development of the system component in the corresponding dimension will be provided.

6.4.1 The construction of the system component of the structure dimension

In order to understand how subjects' representations of the system component changed over time, it is necessary to examine the content of the representations at each data collection point. In this section, subjects' representations of the system component are analyzed at the semantic level.

6.4.1.1 Representations of the system component at T1

At T1, the subjects listed 76 concepts related to the system component. This number represented 12.1% of all the concepts associated with the structure dimension at T1. Table 15 lists some of the concepts that subjects contributed to represent MedlinePlus as an integrated system at T1. Each row of the table represents a different aspect of the system component. The numbers associated with each aspect are the number and percentage of the concepts used to describe that aspect among those concepts describing the system component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 15. Subjects' representations of MedlinePlus as an integrated system at T1

System structure (22, 28.9%)	
- Database	- Compilation
- Links to outside info; links to other medical sites	- Inside information
- Linked pages	- The site seems to work primarily based on links rather than search features.
- Storage tank	
Audience (17, 22.4%)	
- All genders/ages	- Patients
- Laypeople	- Parents
- Everyday users	- Novices
- Juvenile	- Older adults
- Someone with an interest in a particular condition	- Adults
- Professionals	
The usage of the site (11, 14.5%)	
- Learning how to be healthy	- Counseling, consult
- Consult a doctor	- Explore
- Starting point	- Self-diagnosing
- Information gathering	- Used to see if any major health issues
- Useful to research minor twinges	- Would also seek info from doc
Agencies involved (11, 14.5%)	
- General: government, agencies, partners, medical associations	- Specific: NIH
System behavior (13, 17.1%)	
- Pop-up windows	- Pooling information
- Redirection	- Search results yield snippets
- Processes search information	- Filter
- Click to read more on specific topic	- Aggregate
- Updated daily; Recently added info	- Spam
Similar sites (2, 2.6%)	
- WebMD	- PubMed

As shown in the table, subjects represented the MedlinePlus system as an integrated system from the following aspects: system structure, the audience, the possible usage of the site, agencies involved in the site, system behavior, and similar sites.

In representing the structure of MedlinePlus, subjects expressed the idea that MedlinePlus was a large database, storage tank, or compilation of information. It has inside information, but also linked out to external resources and information. The site worked primarily based on links rather than search features. Within the site, the pages were linked.

Meanwhile, subjects started to incorporate information specific to MedlinePlus into their mental representations. They began to form an idea of the user population for which MedlinePlus is suitable. Some indicated that MedlinePlus was for everyday users, lay people, parents, adults, juveniles, and someone with an interest in a particular condition. Some also believed MedlinePlus was for novice users, while others thought that it could be useful for medical professionals or curious professionals and that the site was simple for doctors.

Along with forming opinions about the audience for MedlinePlus, subjects started pondering their potential usage of the site. After interacting with the site for 5 minutes, subjects formed the ideas that the site was a starting point for research and that it could be used for information gathering, consulting, or counseling. Specially, they believed that it would be useful for self-diagnosis or research on specific topics like minor twinges. It could also be used to learn how to be healthy and to see whether there are any major health issues. Meanwhile, a couple of subjects said that they would still need to consult doctors and seek information from them.

Subjects also paid attention to another factor specific to MedlinePlus: agencies that create, contribute to, or are involved in the site in any way. MedlinePlus organizes links to information from the NIH and other government agencies and links to information from carefully evaluated professional and voluntary health organizations (Marill, Miller, & Kitendaugh, 2006). In the short 5-minute exploration, subjects noticed that the site was associated with the government in general and NIH specifically. They understood that various agencies, partners, and medical associations were involved.

During the initial interaction, subjects also observed the system's behavior. They speculated that MedlinePlus aggregated, pooled, and filtered information from other medical websites. They believed that the system was updated daily and that some information was added recently. They also found out that clicking on links on the site redirected the user to another page.

In attempting to understand MedlinePlus, a couple of subjects recalled sites with which they had experience, particularly WebMD and PubMed. These connections are apparently built on the subject similarity between MedlinePlus and the two named sites. This empirical observation suggests that, when beginning to interact with a new system, some users employ metaphors to assist them in making sense of the new system. The metaphors that they employ at this early stage of interaction are likely to share the subject matter with the new system being explored.

6.4.1.2 Representations of the system component at T2

At T2, the simple task group used 31 concepts (10.4% of all the concepts associated with the structure dimension) and the complex task group used 40 concepts (15.9%) to represent the system component. Table 16 lists the concepts that subjects contributed to each aspect of the system component at T2. The numbers associated with each aspect are the number of the concepts used to describe that aspect and the percentage of the concepts among all the concepts describing the system component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 16. Subjects' representations of MedlinePlus as an integrated system at T2

Simple task group	Complex task group
System structure (12, 38.7%) - Google-like - Linear - Links to outside official sources; Outside links to other medical websites	**System structure** (20, 50%) - Connections to outside sources; External article sources; Partner sites - The combination of articles from the site and from sites that it links to provides a large quantity of information overall.
Audience (0)	**Audience** (3, 7.5%) - Layman - Some doctors
The usage of the site (1, 3.2%) - Seek medical help	**The usage of the site** (3, 7.5%) - Background research - Educating households - Starting point
Agencies involved (12, 38.7%) - General: major organizations, health agencies, government - Specific: Mayo clinic, CDC, NIH, US Health Department	**Agencies involved** (11, 27.5%) - General: government, organizations - Specific: CDC, NIH, FDA, Google
System behavior (4, 12.9%) - Aggregate - Pop up windows/separate screens - Pre-formulated search responses	**System's behavior** (3, 7.5%) - Some links did not work - Search bar directs to links - The search function tends to bring up several links to the same article (on different sites).
Similar sites (2, 6.5%) - Google-like - Harder to use than Google	**Similar sites (0)**

Compared to T1, both groups slightly increased their representation of the structure aspect of the system component at T2, though the increases were not statistically significant. Between the two groups, the complex task group showed more in-depth understanding of the structure of the system. As can be seen from the table, both groups recognized that MedlinePlus pulled information from outside medical information websites (as they did at T1), while subjects in the complex task group also pointed out that MedlinePlus had its own information. As one subject commented:

> Much of the useful information comes from sites linked to by MedlinePlus rather than articles that are a part of the site itself.

Another subject from the complex task group commented that:

> The combination of articles from the site and from sites that it links to provides a large quantity of information overall.

At T2, unlike at T1, the simple task group did not generate any concepts related to the audience for the system. The complex task group did, but their representations of the audience of MedlinePlus were much fewer in number and much less specific than at T1. They mentioned only two general types of audiences: laymen and doctors.

Similar to T1, subjects represented agencies involved in MedlinePlus at two levels: general agencies and specific institutions. However, compared to T1, both groups increased their representation of the agencies involved (from 11 concepts at T1 to 23 concepts for the two groups, combined, at T2). Both groups identified more specific agencies involved in MedlinePlus, such as the Mayo Clinic, CDC, and FDA.

At T2, both groups reduced their representations of system behavior. The two groups showed differences in representing system behaviors. The complex task group focused more on specific instances that they experienced on the site; for example, one subject noticed that “some links did not work” and another commented that “the search function tends to bring up several links to the same article (on different sites).” The simple task group was more focused on the general behavior of the system, such as pop-up windows, just as they were at T1.

As shown in the table, only one subject in the simple task group employed similar websites to make sense of MedlinePlus at T2. Instead of bringing up websites with similar subject matter, like subjects did at T1, the subject brought up the site, Google, which shares only some basic structural similarity with MedlinePlus.

6.4.1.3 Representations of the system component at T3

At T3, the simple task group significantly increased their representation of the system component of MedlinePlus from 31 concepts (10.4%) to 46 concepts (16.7%) ($t(18) = 2.51$, $p = 0.022$). The complex task group did not show significant difference from their system component representations at T2, but the data show that the group decreased their representation of the system component as whole from 40 concepts (15.9%) to 27 concepts (10.5%). Table 17 lists concepts concerning the system component at T3. The numbers in parentheses are the number of concepts contributed by the subjects to a particular aspect and the percentage of the concepts among all the concepts describing the system component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 17. Subjects' representations of MedlinePlus as an integrated system at T3

Simple task group	Complex task group
System structure (17, 37.0%) - Centralized; Central location of information - Links to outside sources; Links to other medical sites - Specific details on linked websites, the actual medline plus website seemed to just be a summary of everything	**System structure** (12, 44.4%) - Algorithm - Database - External links; Outside sources - For finding information about treatment options, it helps to look both at the information provided by the site and at the other sites that it links to.
Audience (8, 17.4%) - Adults - Patients - Different people - Medical professionals - Not good for physicians - Teenagers	**Audience** (6, 22.2%) - Everyone - Older adults - Expert - Layman - Parents - User
The usage of the site (1, 2.2%) - Self diagnosis	**The usage of the site** (1, 3.7%) - First-step in research
Agencies involved (13, 28.3%) - General: foundations, health authorities, other medical sites, government program, government - Specific: NIH, CDC, FDA, Jewish health center, NLM, Mayo Clinic	**Agencies involved** (6, 22.2%) - General: Government, health organizations - Specific: CDC, NIH, Mayo Clinic
System behavior (4, 8.7%) - Aggregate - Getting an error when I clicked a link - Pop up screens; Separate screens	**System behavior** (0)
Similar sites (3, 6.5%) - WebMD - Wikipedia - Yahoo like	**Similar sites** (2, 7.4%) - Like a medical Google - Like an encyclopedia

Compared to T2, both groups used about the same number of concepts to represent the structure aspect of the system component. The simple task group used 12 concepts (38.7% of the system component) at T2 and 17 concepts (37.0%) at T3; the complex task group used 20 concepts (50%) at T2 and 12 concepts (44.4%) at T3. In addition, both groups recognized that MedlinePlus consists of information from both inside and outside of the website. For example, one subject in the simple task group commented that:

> Specific details on linked websites, the actual medline plus website seemed to just be a summary of everything.

One subject in the complex task group commented that:

> For finding information about treatment options, it helps to look both at the information provided by the site and at the other sites that it links to.

In contrast to T2, when subjects' representation of the audience decreased dramatically, both groups increased their representations of the audience of MedlinePlus. The representations from the two groups at T3 had a similar weight in the system component (the simple task group: 8 concepts, 17.4%; the complex task group: 6 concepts, 22.2%) and were as specific as they were at T1.

The representations of the usage of the site were infrequent at T3, as they had been at T2. Even after two sessions of searching activity, the simple task group and the complex task group each generated only one usage-related concept.

Both groups also continued to represent the agencies involved in MedlinePlus (with 13 concepts, 28.3%, in the simple task group, and 6 concepts, 22.2%, in the complex task group). Although the two groups still represented the agencies at general and specific levels, the simple task group was able to recognize more specific institutions related to MedlinePlus, such as the National Library of Medicine (NLM), the FDA, and the Jewish Health Center. The complex task group was able to recognize one new agency: the Mayo

Clinic. Thus, overall, despite the fact that the proportion of concepts associated with agencies decreased slightly, the specificity of the representations improved slightly.

Representations of system behaviors continued to be minimal, and were completely absent from the complex group's concept listings. The simple task group generated 4 concepts (8.7% of its system component) at T3, including pop-up windows, aggregating information, and an instance of using the system. These representations were at the same level of specificity as the representations at T2.

At T3, both groups did compare MedlinePlus with other similar sites. Unlike at T1 when they only mentioned WebMD and Wikipedia, or at T2 when they only mentioned Google, at T3, subjects in both groups came up with sites that share subject matters with MedlinePlus (WebMD and Wikipedia) and sites that share structural similarities with MedlinePlus (Yahoo and encyclopedia). In addition, one subject brought these two aspects together, by suggesting a medical Google.

6.4.1.4 Discussion of the construction of the system component of the structure dimension

Table 18 illustrates the compositions of the system component of the structure dimension at each data collection point for both groups, in terms of the number of concepts listed. Though it is included in the discussion, T0 is not included in the table since it was based on subjects' written answers to a question in the demographic questionnaire rather than concept listing protocols.

Table 18. Subjects' representations of the system component of the structure dimension at T1, T2, and T3

	T1	T2		T3	
		Simple	Complex	Simple	Complex
Structure	22 (28.9%)	12 (38.7%)	20 (50.0%)	17 (37.0%)	12 (44.4%)
Audience	17 (22.4%)		3 (7.5%)	8 (17.4%)	6 (22.2%)
Usage of the system	11 (14.5%)	1 (3.2%)	3 (7.5%)	1 (2.2%)	1 (3.7%)
Agencies involved	11 (14.5%)	12 (38.7%)	11 (27.5%)	13 (28.3%)	6 (22.2%)
System behavior	13 (17.1%)	4 (12.9%)	3 (7.5%)	4 (8.7%)	
Similar sites	2 (2.6%)	2 (6.5%)		3 (6.5%)	2 (7.4%)

At T0, in subjects' initial mental models of general information-rich web spaces, subjects represented the system component in relation to two aspects: structure and audience. At T1, after subjects had used the MedlinePlus system for 5 minutes, they tended to make sense of and represent MedlinePlus as an integrated system in relation to the following aspects: structure, audience, usage of the system, agencies involved in the system (creator or contributors, e.g., NIH and CDC), system behavior, and similar systems. The structure of the system refers to subjects' understanding of the general information architecture of the system. For example, subjects pointed out that the site links to outside information. The audience of the system is subjects' understanding of the population that MedlinePlus serves, such as novices, adults, and patients. The usage of the system is subjects' understanding of the potential usage of the system, such as for information gathering, and for researching minor twinges. System behavior refers to system actions such as aggregating, filtering, updating, and pop-up windows. It is clear that, when subjects started to explore a new system, they tried to represent various aspects of the system. Such representations suggest that, when a novice user interacts with a new system, he/she does not focus on only one or two aspects; rather, he/she may try to understand the system in relation to multiple different aspects.

From the table, it is clear that system structure was the most represented aspect of the system component across time. One quarter to one half of the concepts listed at each data collection point by each group were associated with this aspect of the system. In the initial (T0) model, subjects represented the structure of general information-rich web spaces as databases, collections of other sites, and links to outside websites; subjects' representations of the structure of MedlinePlus after using it for 5 minutes did not go too much beyond their initial model, which might be due to their brief exposure to the system. At T2 and T3, as subjects' representations of the system structure increased in quantity, their understanding of the structure also improved. Subjects were able to recognize and articulate that MedlinePlus had its own information and also linked to outside sources.

In the initial (T0) model, subjects thought the audience for information-rich web spaces could be novices, experts, or a wide variety of people. At T1, subjects paid a good amount of attention (17 concepts and 22.4% of the system component) to the audience aspect of the system component. Subjects thought that MedlinePlus would be useful for many different populations, such as laypeople, juveniles, professionals, novices, and patients. Thoughts about the intended audience are important because they indicate that subjects began to think about whether the system fits them. In real use cases, these judgments might help users to decide whether they are going to continue interacting with the site or not. Compared to the initial model, where subjects represented the audience as novices, experts, and various types of people, the representations of the audience for MedlinePlus at T1 were more specific.

At T2, the simple task group's representations of audience for MedlinePlus completely disappeared; the complex task group retained this aspect, but reduced it to 7.5% of the system component (3 concepts) and the representations were fairly general (laymen and doctors). At T3, both groups increased their representations of this aspect (simple task group: 8 concepts, 17.4%; complex task group: 6 concepts, 22.2%). The representations also became more specific, with a level of specificity that is similar to T1 (e.g, adults, patients, medical professionals, and teenagers). It is clear that the representations of the MedlinePlus audience developed from specific to general and then returned back to specific. It is also worth pointing out that, at T3, subjects started to develop ideas of whom the system is not suitable for. A subject in the simple task group pointed out that MedlinePlus was not good for physicians.

At T1, 14.5% (11 concepts) of the system component was about the usage of the site. After the brief 5 minute interaction, subjects thought that MedlinePlus would be potentially useful for information gathering, consulting, and learning to be healthy. At T2 and T3, the representations of this aspect were fewer for both groups. This trend of decrease in the representation of potential usage of the site suggests that, as users' experience with the system increases, instead of thinking about possible uses of the site, users are more likely to focus on what the system is actually doing, represented as other aspects of the system.

The agencies involved in MedlinePlus were another highly represented aspect of the system component. Subjects' representations of the agencies involved in MedlinePlus became more specific over time. Initially, subjects represented the agencies at a general level, pointing out that government and medical associations were related to the site. At T2 and T3, subjects increased the representation of this aspect and they were able to recall more specific institutions that contribute content to MedlinePlus, such as the NIH and the CDC.

The system behavior aspect was emphasized most at T1 (13 concepts, 17.1% of the system component). At T1, subjects represented various actions of the system at a general level, such as pop-up windows and information aggregating, pooling, and updating. This aspect of the system component was less emphasized at T2 and T3. The complex task group did not mention system behavior at T3. Over time, the representations of this aspect became more specific. For example, one subject in the complex task group commented at T2 that "the search function tends to bring up several links to the same article", and a subject in the simple task group pointed out at T3 that he/she got an error when clicking a link. The developmental trend suggests that, when users first encounter the system, they observe actions of the system and represent these actions at a general level; when their experience with the system increases, they gradually reduce the emphasis on this aspect and the representations become more specific.

When interacting with a new system, some people employ metaphors to assist them in making sense of the new system. In the study, at T1, subjects mentioned WebMD and PubMed. At T2, the simple task group mentioned Google, while the complex task group did not include any mention of similar sites. At T3, both groups increased the mention of similar sites, such as Wikipedia, an encyclopedia, Medical Google, and Yahoo. This result suggests that, in the process of learning to use a new system, users think about similar systems. At the beginning of the interaction, users are more likely to think about sites that share similar subject matter with the current system. When users' experience with the system increases, they start thinking about sites that have a structure similar to the current system. Similar systems that are mentioned at different times might serve different purposes. Both topic- and structure-similar sites mentioned by users at early stages of interaction are likely to be employed to help them understand the system. When users have more experience with the system, the topic-similar sites are more likely to be alternative sites that they would use for tasks they want to perform.

6.4.1.5 Summary of the construction of the system component of the structure dimension

As has been discussed, the subjects' representations of the system component of the structure dimension experienced changes in the quantity and quality of the concepts dedicated to each aspect over time. The changes of quantities are reflected in Table 18. The qualitative development of each aspect of the system component was described and discussed in the previous section.

In summary, subjects' understanding of the structure aspect improved over time. At T1, due to the short interaction with the system, subjects did not go beyond their mental models of the general information-rich web space, representing the structure of MedlinePlus as a database or a collection of other sites. But at T2 and T3, after using the system to solve assigned tasks, subjects were able to articulate that MedlinePlus owned its own information and also extensively linked out to other websites.

Subjects' representations of the agencies involved and the system behaviors aspects of the system component both developed from general to specific. At T1, subjects recognized the involvement of government websites and medical associations. At T2 and T3, they were able to point out specific institutions such as the CDC and Mayo Clinic. At T1, the behavior of the system was described at a general level such as pop-up windows. At T2 and T3, subjects described system behaviors at a more specific level, such as getting an error when clicking on a link.

Subjects' representations of the audience aspect moved from specific to general and back to specific. At T1, subjects pointed out that MedlinePlus could be used by various specific populations, such as juveniles, patients, and parents. At T2, both groups reduced the emphasis on this aspect and only two general populations were mentioned: laymen and doctors. At T3, both groups increased the emphasis on this aspect and the representations became as specific as they were at T1.

Subjects' emphasis on the usage of the system decreased over time and the representations became less specific. The developmental trend suggests that as users' experience with the system increases, they emphasize what the system is actually doing more than what the system could potentially do.

Subjects employed metaphors (similar sites) to help them make sense of MedlinePlus. The function of these metaphors in facilitating users' understanding of the system changed over time. At the beginning, similar sites were employed to help subjects understand the subject matter of the site. As subjects' experience with the system increased, similar sites were employed to help understand the structure of the site, or both the structure and the subject matter.

6.4.2 The construction of the system component of the evaluation/emotion dimension

When attempting to construct their mental models of MedlinePlus by representing various aspects of the system, subjects simultaneously evaluated or expressed emotions about each individual component in their mental models. This section reports on the construction of subjects' evaluations and emotions about the system component in their mental models of MedlinePlus over time.

6.4.2.1 Evaluations of and emotions about the system component at T1

About 23.2% of the concepts associated with the evaluation/emotion dimension of subjects' mental models of MedlinePlus at T1 were about the system component (47 concepts across both groups). Specific aspects of subjects' evaluations of and emotions about the system component at T1 are shown in Table 19. The numbers associated with each aspect are the number of the concepts used to describe that aspect and the percentage of the concepts among all the concepts evaluating the system component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 19. Evaluations and emotions about the system component of subjects' mental models of MedlinePlus at T1

Attributes of the system (12, 25.5%)	
- Portal - Search oriented - Searchable - Free - Similar to other information-rich sites	- Large database - Data-rich, Resourceful - Quite large in scope - Big - Updated daily
Usefulness (10, 21.3%)	
- Helpful - Wide array of use - Less need for doctors	- Good place to seek initial advice - Good for avoiding doctor - Not self sufficient
Usability of the system (24, 51.1%)	
- Accessible - Quick access - Easy to use - User friendly - Good - No glitches	- Fast paced - Prompt - Fast - Not reactive - Not clear how they pick links
Public awareness (1, 2.1%)	
- Not well known to the public	

As reflected in the table, subjects evaluated MedlinePlus as an integrated system in relation to four different aspects: attributes of the system, usefulness, usability of the system, and public awareness. Attributes of the system refers to subjects' understanding of the characteristics of MedlinePlus as an integrated system. In talking about the system's attributes, subjects described MedlinePlus as a search oriented portal site with frequent updates. Similar to other information-rich websites, it was considered resourceful and free to access.

Usefulness of the system refers to whether subjects think the system is useful. At T1, after a short interaction with the system, subjects recognized that MedlinePlus was helpful, serving as a good place for initial medical advice and helping people avoid having to see the doctor. Only one subject pointed out that the system was not self-sufficient.

At T1, over half of the evaluations and emotions (24 concepts, 51.1%) expressed toward the system component were about the usability of the system. The usability of the system concerns the ease of use of MedlinePlus as an integrated system, rather than any single function or element in the system. System usability is achieved by many aspects of the system design, such as system structure, response speed, content, and interface. In line with acknowledging the usefulness of the system, subjects largely agreed that MedlinePlus was usable. The system was considered fast to access, easy to use, friendly to users, and having no noticeable glitches. Only a couple of subjects criticized the system, saying that it was not reactive and it was not clear how MedlinePlus selected the links. Meanwhile, one subject pointed out that MedlinePlus was not well known to the public. In summary, it is clear that, after exploring the system for 5 minutes, subjects had a positive view of MedlinePlus as an integrated system.

6.4.2.2 Evaluations of and emotions about the system component at T2

At T2, both groups reduced their representations of evaluations of and emotions about the system component in their mental models. In combination, the two groups generated 38 concepts to evaluate the system component, compared to 47 at T1. At T2, the two groups were also significantly different in the emphasis they placed on the evaluations and emotions related to the system component ($t(36) = 2.27$, $p = 0.029$): among the 38 concepts, 31 were from the simple task group and 7 were from the complex task group.

Table 20 shows both groups' evaluations of and emotions about the system component at T2. The numbers associated with each aspect are the number of the concepts used to describe that aspect and the percentage of the concepts among all the concepts evaluating the system component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 20. Evaluations and emotions about the system component of subjects' mental models of MedlinePlus at T2

Simple task group	Complex task group
Attributes of the system (5, 16.7%) - Information-rich, Resourceful - Legitimate website because its governmental - Search-heavy - Personalized	**Attributes of the system** (2, 28.6%) - Complex - Complicated
Usefulness (3, 10.0%) - Helpful	**Usefulness** (1, 14.3%) - Helpful
Usability of the system (22, 73.3%) - Quick, Fast - Responsive - Easy to use - Direct, Straight-forward - Easy access to information - Logical - Not stressful to use - Not intrusive - User friendly - Convenient - Harder to use than Google - Difficult - Not good for browsing	**Usability of the system** (4, 57.1%) - Accessible - Easy - Simple to use - User friendly

As shown in the table, at T2, subjects evaluated MedlinePlus as an integrated system in relation to three aspects: attributes of the system, the usefulness of the system, and the usability of the system. None of the concepts listed by the subjects were related to public awareness of the system.

The two groups showed some differences in evaluating the system component. When talking about the attributes of MedlinePlus, the complex task group focused on the system's complexity, while the simple task group was able to represent various other attributes of the system, such as personalization and the legitimacy of the site. When talking about usability of the system, subjects in both groups generally agreed that MedlinePlus was usable, though some subjects in the simple task group felt that the system was difficult to use and harder than Google.

6.4.2.3 Evaluations of and emotions about the system component at T3

At T3, after performing the same set of tasks, the simple task group slightly decreased the representation of the evaluations and emotions about the system component, from 31 concepts (22.3% of the evaluation/emotion dimension) at T2 to 21 concepts (14.1%) at T3, but the decrease is not statistically significant. The complex task group significantly increased their representation of their evaluation of and emotions about the system component, from 7 concepts (5.6%) of the evaluation/emotion dimension at T2 to 19 concepts (15.7%) at T3 ($t(18) = 2.47$, $p = 0.024$).

Table 21 shows both groups' evaluations and emotions about the system component at T3. The number in the parentheses is the number of concepts that were contributed by the subjects; the percentage of each group's concepts associated with a particular aspect or attribute is also given. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 21. Evaluations and emotions about the system component of subjects' mental models of MedlinePlus at T3

Simple task group	Complex task group
Attributes of the system (3, 14.3%) Pre web 2.0 Searchable Good for searching	**Attributes of the system** (1, 5.3%) Text-based
Usefulness (7, 33.3%) Helpful One stop shop Would probably use Wikipedia I would use webmd Not sure medlineplus is necessary Not sure its uniqueness is effective enough	**Usefulness** (4, 21.1%) Helpful Good for medical research Need foundational knowledge
Usability of the system (11, 52.4%) Easy to use, Easy Straight-forward Easy access Fast Immediate Responsive Relevant info easy to find Like the separate screens User friendly Good for searching	**Usability of the system** (14, 73.7%) Accessible to everyone User friendly Easy fact-finding Easy to find; Easy to use Quick; Fast Responsive Good access to outside information Good system The information is there, the trick is finding it.

Similar to T2, subjects evaluated MedlinePlus as an integrated system in terms of three aspects: the system's attributes, its usefulness, and its usability. The two groups differed in the focus of their evaluations/emotions. As shown in the table, the simple task group paid attention to both usefulness (7 concepts, 33.3%) and usability (11 concepts, 52.4%) of the system, while the complex task group focused more on the usability (14 concepts, 73.7%) of the system. Very few of the concepts listed by the subjects were related to the system attributes.

When talking about the usefulness of the system, both groups acknowledged that MedlinePlus was a good and helpful system. However, the simple task group was more critical of the usefulness of the system: they questioned the effectiveness and necessity of MedlinePlus and said that they would have switched to WebMD or Wikipedia if they were allowed.

At T3, both groups tended to remain positive about the system's usability. Subjects' evaluations of the system's usability were still focused on it being easy to use and the quick responses of the system, and the concepts they listed were about the same as at T1 and T2. However, subjects in the complex task group also commented that sometimes it was tricky to find the information needed in the system.

6.4.2.4 Discussion of the construction of the system component of the evaluation/emotion dimension

Table 22 illustrates the distribution of subjects' evaluations of and emotions about the system component of subjects' mental models of MedlinePlus at the three data collection points for both groups, in terms of the number of concepts listed. Though it is included in the discussion, T0 is not included in the table since it was based on subjects' written answers to a question in the demographic questionnaire rather than concept listing protocols.

Table 22. Subjects' evaluations of and emotions about the system component at T1, T2, and T3

	T1	T2		T3	
		Simple	Complex	Simple	Complex
Attributes	12 (25.5%)	5 (16.7%)	2 (28.6%)	3 (14.3%)	1 (5.3%)
Usefulness	10 (21.3%)	3 (10.0%)	1 (14.3%)	7 (33.3%)	4 (21.1%)
Usability	24 (51.1%)	22 (73.3%)	4 (57.1%)	11 (52.4%)	14 (73.7%)
Public awareness	1 (2.1%)				

In their initial model (T0), subjects generally mentioned the usability of information-rich web space as convenient and accessible. When they saw MedlinePlus, as shown in the table, subjects tended to evaluate the system component in relation to three aspects: the attributes of the system, its usefulness, and its usability.

The attributes are characteristics of MedlinePlus. Subjects' representations of the attributes aspect tended to decrease over time, from 12 concepts at T1 to 7 concepts at T2 to 4 concepts at T3 (the two groups combined). The diminishing representation of attributes over time is reasonable because attributes were more related to understanding the system than to the use of the system. When subjects first encounter a system, it is natural for them to try to understand characteristics of the system; but when users become more focused on how to use the system to solve real tasks, it is natural for them to think less about the characteristics of the system and pay more attention to aspects related to the use of the system, particularly the usability of the system.

Despite the diminishing weight, subjects' representations of the attributes of the system developed over time. At T1, after a short interaction with MedlinePlus subjects focused more on the superficial characteristics of the system, such as it being searchable, free, big, data-rich, and resourceful. At T2 and T3, subjects still listed superficial attributes of the system, such as resourceful and searchable, but they also started representing attributes that resulted from using the system to solve particular tasks. At T2, subjects listed system attributes, such as search-heavy, complex, and complicated, and at T3, pre web 2.0 and good for searching.

The number of concepts related to the usefulness of the system experienced a drop at T2, right after the first set of tasks, then increased at T3. At T1, subjects showed overly positive feelings about MedlinePlus, thinking the system was a good place to seek initial advice and good for avoiding a doctor visit. Only one subject pointed out that the system was not self sufficient. At T2, the two task groups did not show notable differences in their evaluation of the usefulness of the system. Generally, they felt that MedlinePlus was helpful, which failed to support the speculation that the simple task group would believe the system to be more useful than the complex task group due to the fact that it was easier to find answers for simple tasks than for complex tasks in MedlinePlus.

At T3, after they performed the second set of tasks and had more experience with the system, subjects, especially those in the simple task group, started questioning whether MedlinePlus was necessary and effective. Meanwhile, they showed signs of wanting to switch to alternative systems, particularly WebMD and Wikipedia, for the tasks. The simple task group's change of attitude toward the usefulness of the system might be due to the changes in the tasks that they performed. In the first search session, the simple task group performed a set of simple tasks, for which the answers were easy to find in MedlinePlus. Thus, it was natural for subjects to feel that the system was useful at T2. However, the second search session included two complex tasks. MedlinePlus was not as effective in answering the complex tasks. Therefore, subjects began to question the usefulness of the system.

At all three times, usability was consistently the most emphasized aspect of the system component in the evaluation/emotion dimension. At T1, over half of the evaluations of and emotions about the system component were related to usability. At T2, the weight of the usability aspect increased for both groups, with the concepts contributed by the simple task group outnumbering the complex task group (22 concepts vs. 4 concepts). At T3, the complex task group increased the number of usability evaluations (from 4 to 14 concepts), while the simple task group decreased it (from 22 to 11 concepts), with the former slightly outnumbering the latter. These findings illustrate that the increase in the usability evaluation of the system accompanied the performing of simple tasks. It is possible that the path for simple tasks helps reveal the information architecture of MedlinePlus, which encouraged subjects to express their evaluations about the usability of the system. Future studies could be designed to explore which particular characteristics (e.g., task complexity, paths to answers) of tasks affect subjects' evaluation of the usability of an IR system.

Despite the changes of the quantity of the evaluations of system usability at different times, subjects' attitudes and feelings about the system's usability were consistently positive over time: 94.3% of the evaluations of the usability were positive across the three data collection points. At all three times, subjects thought that the system was fast, accessible, easy to use, and user-friendly. Nevertheless, subjects also pointed out usability problems of the system. At T1, they pointed out that the system was not reactive and was not clear about how the system picks links. At T2 and T3, after gaining experience with using the system to solve particular tasks, subjects focused on usability problems more related to the use of the system, such as it being hard to use and tricky to find answers.

6.4.2.5 Summary of the construction of the evaluations of and emotions about the system component

Subjects evaluated MedlinePlus as an integrated system in relation to three aspects: the attributes of the system, the usefulness of the system, and the usability of the system. Subjects' emphasis on the attributes of MedlinePlus decreased over time and the representations developed from superficial evaluation, such as big, data-rich, and free to more specific evaluations, such as complex and search-heavy, as a result of using the system to solve assigned tasks.

Subjects' evaluations of the usefulness of the system experienced a drop from T1 to T2 and an increase from T2 to T3. At T1 and T2, subjects in both groups thought MedlinePlus was useful. At T3, subjects in the simple task group started questioning the usefulness of the system and showed a tendency to switch to alternative systems, such as WebMD. It is worth noting that subjects in the complex task group did not feel the system was less useful because it was harder to find answers for the complex tasks.

The usability of the system was the most evaluated aspect over time. Subjects expressed very positive feelings about the usability of the system, considering the system fast, easy to access, and user-friendly. The number of concepts that subjects contributed to evaluating the usability of the system was related to the tasks that subjects performed. After performing simple tasks, subjects tended to express more evaluations of the usability of the system. Future research is needed to explore how tasks affect users' evaluations of an IR system's usability.

6.5 The construction of the content component of the structure and evaluation/emotion dimensions

This section starts by describing subjects' representations of the content component in the structure dimension of mental models at different times, to demonstrate how subjects' representations of the content of MedlinePlus changed over time (section 6.5.1). Subjects' evaluations of and emotions about the content of MedlinePlus, as well as their changes over time, are discussed in section 6.5.2.

6.5.1 The construction of the content component of the structure dimension

6.5.1.1 Representations of the content component at T1

The content component of the mental model is subjects' representations of the information contained in MedlinePlus. At T1, the content component comprised 66.0% of the structure dimension of subjects' mental model of MedlinePlus (413 concepts).

The content component of subjects' mental models at T1 illustrates the six different aspects included in subjects' representations of the content of MedlinePlus (see Figure 11). Three of these (general, topical, and specific) were related to the subject of the content; the other three (type, format, and presentation) were related to the form of the information provided in MedlinePlus.

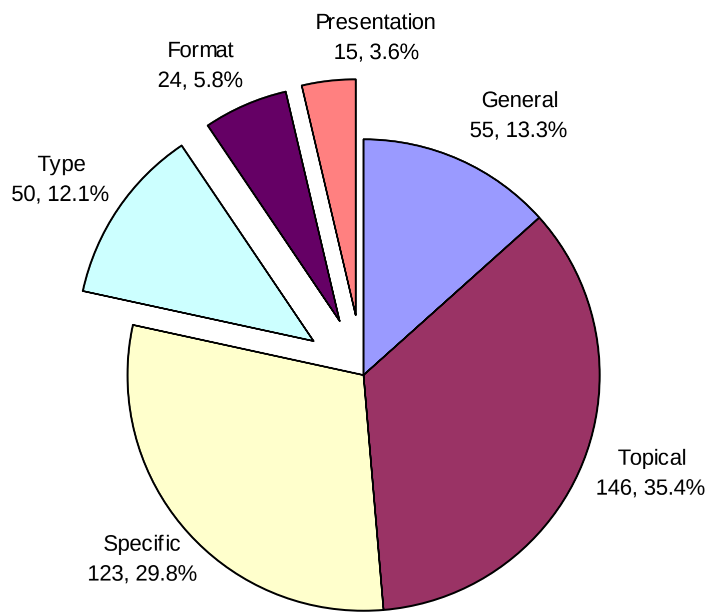


Figure 11. The content component of subjects' mental models at T1

6.5.1.1.1 Subject of the content: general, topical, and specific

At T1, 324 concepts (78.5% of the content component) were employed to describe the subject of the MedlinePlus content. Subjects represented the subject of the information in MedlinePlus at three different levels: general, topical, and specific.

Concepts classified as general refer to the MedlinePlus content in general. Concepts at this level did not reveal the subject matter of the site and would not differentiate MedlinePlus from other websites. Examples of such concepts included information, data, advice, articles, definition, issues, literature, research, analysis, problems, questions, inquiry, and sources. The general representations of the content included 55 concepts

(13.3% of the content component).

Concepts classified at the topical level described or named high level topics covered in MedlinePlus. Examples of such concepts included medical information, disease, symptoms, treatments, alternative medicine, health information, diagnosis, supplements, drugs, disease, lifestyle, and prevention. The topical representations of the content included 146 concepts (35.4% of the content component).

Concepts classified at the specific level were more specific than the content at the topical level. They described or named specific diseases, treatments, supplements, or drugs described in MedlinePlus. Examples included diabetic foot, bioterrorism, herbal medicine, diabetes, exercise routines, healthy eating, finger pricks, eye problems, cardiology, endocrine system, black widow, bed bug, anthrax vaccine, bee, ear infection, tuberculosis, lungs, x-rays, and insulin. Specific content listed by subjects at T1 either came from their personal information needs or the content that impressed them when they interacted with MedlinePlus. The specific representations of the content included 123 concepts (29.8% of the content component).

It is clear that, at T1, subjects emphasized the subject of the content in MedlinePlus. In total, 269 concepts (65.1% of the content component) were about the subject of the information, either at the topical or the specific level.

6.5.1.1.2 Type, format, and presentation of information

In addition to the subject matter of the system, subjects also listed concepts related to the type, format, and presentation of the system's content. Table 23 shows subjects' representations of the information type, format, and presentation at T1. The numbers in the parentheses are the number of instances and the percentage of that aspect within the content component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 23. Subjects' representations of information type, format, and presentation in MedlinePlus at T1

Type (50, 12.1%)	
- Dictionary	- Tutorials
- Directories	- Presentations
- Encyclopedia	- Scholarly articles
- FAQs	- Medical sites
- Glossary	- Magazine
- Journals	- Q&A
- News	
- References for medicine	
Format (24, 5.8%)	
- Text	- Photos
- Movies	- Videos
- Images	- Flash
- Pictures	- Multimedia
	- PDF
Presentation (15, 3.6%)	
- Demonstrations	- Short articles
- Details	- Several sections
- Overview	- Summaries
- Diagrams	

After a 5-minute brief exposure to MedlinePlus, subjects had developed a fairly comprehensive understanding of the types and formats of information in the system. Subjects recognized that MedlinePlus contained various types of medical information, including a dictionary, directories, an encyclopedia, FAQs, a glossary of diseases, journals, medical news, Q&A, references for medicine, and tutorials. The information was manifested in different formats, including text, movies, images/pictures/photos, videos, flash, and multimedia.

Meanwhile, subjects paid attention to different ways of presenting information in MedlinePlus. They found that some articles (or webpages) were divided into several sections. Some information was presented as overviews, summaries, demonstrations, diagrams or short articles. Detailed information was also presented in MedlinePlus.

6.5.1.2 Representations of the content component at T2

At T2, the content component was still the most represented component in subjects' mental models of MedlinePlus. For the simple task group, the content component included 203 concepts (68.1% of the structure dimension of their mental model); for the complex task group, it was 127 concepts (50.4%). However, the difference between the groups is not statistically significant. Consistent with T1, subjects represented the content of MedlinePlus from the following aspects: subject (general, topical, and specific), type, format, and presentation of information.

Figure 12 shows the distribution of different aspects of the content component for the two groups.

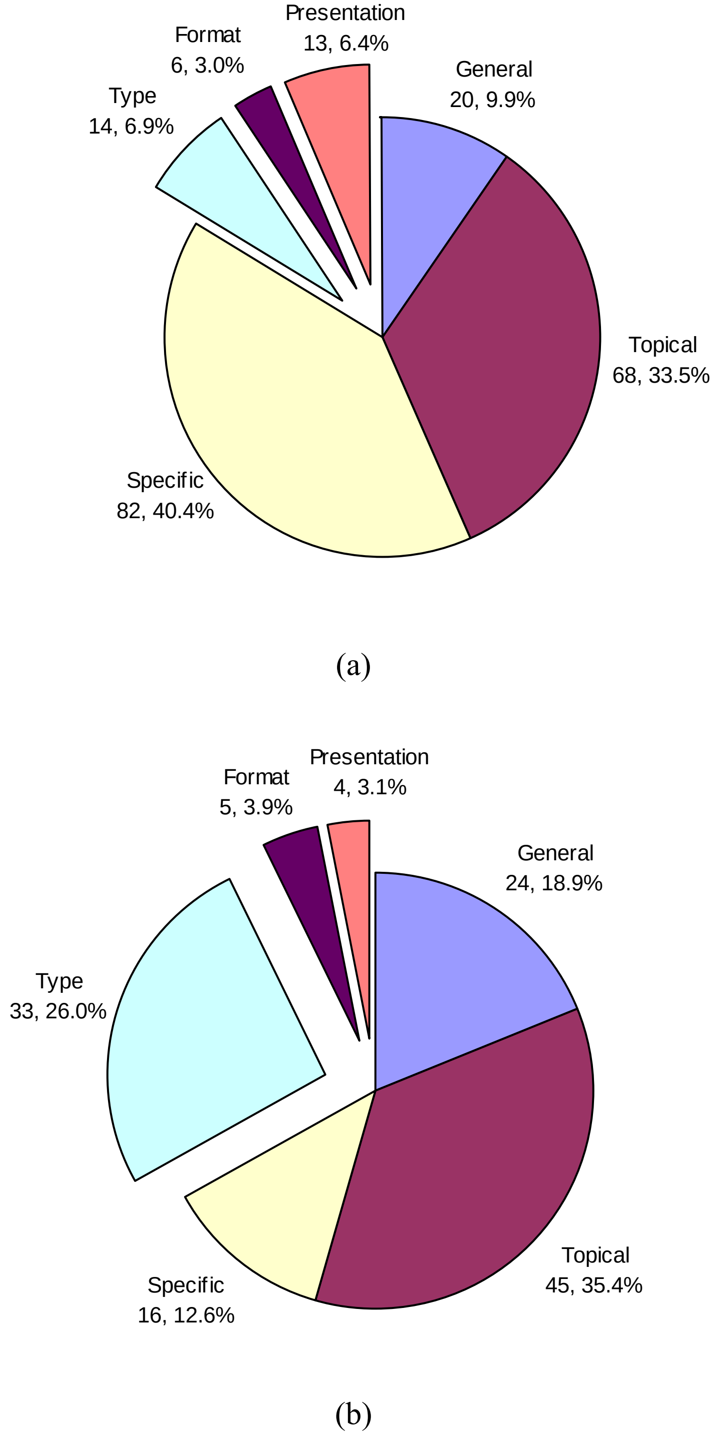


Figure 12. (a) The content component of the simple task group's mental models at T2; (b) The content component of the complex task group's mental models at T2

6.5.1.2.1 *Subject of the content: general, topical, and specific*

For both groups, the biggest chunk of representation of the content still was the content at three different subject levels: general, topical, and specific. The simple task group dedicated 170 concepts (83.8% of the content component), and the complex task group dedicated 85 concepts (66.9% of the content component) to represent it. The difference is not statistically signficant.

At T2, subjects in both groups contributed about the same number of concepts to describe the content of MedlinePlus at a general level. The simple task group contributed 20 concepts (9.9% of the content component), and the complex task group contributed 24 concepts (18.9% of the content component). The concepts were similar to those that appeared in the concept listing at T1, such as information, advice, resources, studies, terms, articles, concepts, suggestion, answers, literature, research, definition, and description.

The two groups also did not differ in representing the content at the topical level. The simple task group contributed 68 concepts (33.5% of the content component) and the complex task group contriubted 45 concepts (35.4% of the content component). The listed concepts were also similar to those that appeared at T1, such as drugs, diseases, information about health, diagnosis, medical information, medicine, treatments, symptoms, surgery, preventions, nutrition, and clinical.

In representing content at the specific level, the simple task group contributed 82 concepts (40.4% of the content component), while the complex task group contributed 16 concepts (12.6% of the content component). The difference is statistically significant ($t(35) = 2.14$, $p = 0.04$). Examples of the specific level content include: health insurance, protein, Lou gehrig, STD, STI, AIDS, blood pressure, diet, dosage, essentials, exercise, HIV, kidney, virus, vitamins, heart disease, side effects, smoking, smoking prevention, diabetes, insulin, liver problems, hypertension, and home diagnosis.

An inspection of the concepts provided by the subjects in the concept listing protocols revealed that subjects tended to recall terms that appeared in the descriptions of the assigned tasks. For example, the simple task group listed terms such as low blood pressure, HIV, heart diseases, protein, vaccinations, and vitamins. The complex task group listed terms such as diabetes, insulin, and hypertension. (Note that both the simple and complex tasks are attached in Appendix C.)

6.5.1.2.2 Type, format, and presentation of information

At T2, subjects continued to represent the type, format, and presentation of the information in MedlinePlus (Table 24). The numbers in the parentheses are the number of instances and the percentage of that aspect within the content component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 24. Subjects' representations of information type, format, and presentation in MedlinePlus at T2

Simple task group	Complex task group
Type (14, 6.9%) - Glossary - News - Fact sheet - FAQs - Dictionary - Journals - Reports - Statistics - What-to-do articles	**Type** (33, 26.0%) - Journals - Scholarly articles; Academic articles - News - Dictionary - Glossary - Clinical trials - Encyclopedia - Statistics - Tutorials
Format (6, 3.0%) - Pictures - Text - PDF	**Format** (5, 3.9%) - Pictures, Graphics, Image - PDF - Videos
Presentation (13, 6.4%) - Tables - Summaries - Descriptions - Overviews - Specifics - Diagram - Figures	**Presentation** (4, 3.1%) - Overviews

As shown in the table, the two groups differed slightly in representing the type of information in MedlinePlus. The complex task group listed 33 concepts (26.0% of the content component) and the simple task group listed 14 concepts (6.9% of the content component)

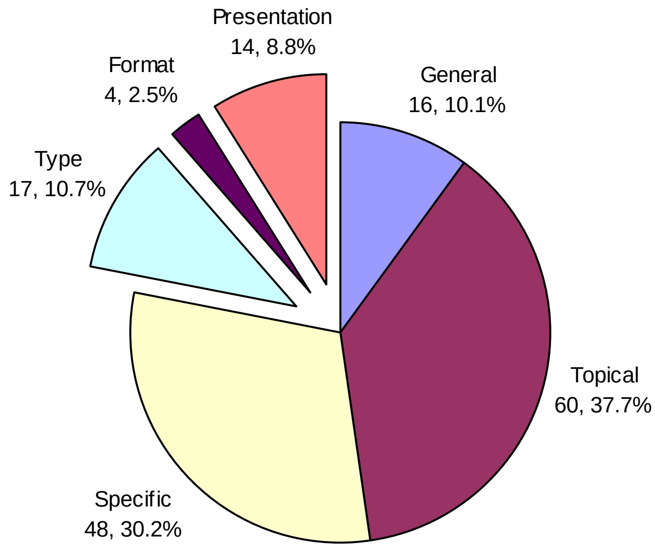
Inspecting the specific types of information that subjects listed also suggested the impact of the nature of the tasks. Both groups included basic information types in MedlinePlus, such as dictionary, encyclopedia, news, glossary, journals, and statistics. However, the representations of the rest of the information types were closely associated with the tasks assigned. The simple task group identified fact sheets, FAQs, reports, and what-to-do articles. These information types were useful for answering the simple tasks, which required mostly factual information. The complex group identified scholarly and academic articles, clinical trials, and tutorials. These information types were useful for answering the complex tasks, which required more in-depth information and required subjects to synthesize information from different sources.

At T2, both groups reduced their representations of the format of information in MedlinePlus: the simple task group listed 6 concepts and the complex task group listed 5 (compared to the total of 24 at T1). However, there were no significant differences in the magnitude and content of these representations from what they were at T1. All the formats that were mentioned at T2 had already appeared at T1.

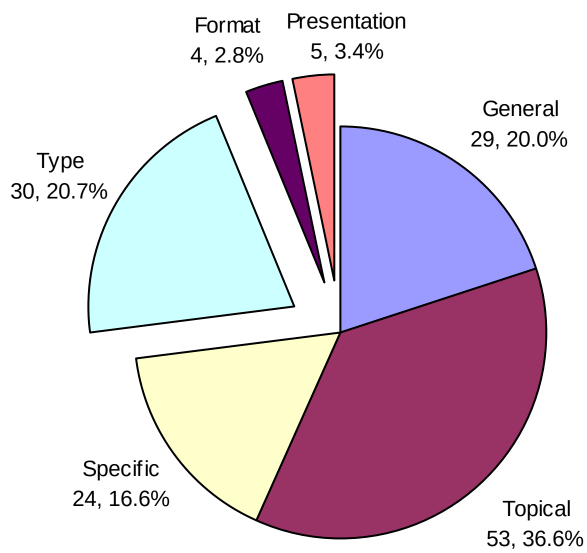
The two groups differed slightly in representing the presentation of information in MedlinePlus at T2, with the simple task group giving more emphasis to this aspect (the simple task group: 13 concepts, 6.4% of the content component; the complex task group: 4 concepts, 3.1% of the content component). The simple task group also identified more ways of presenting information in MedlinePlus, such as tables and figures.

6.5.1.3 Representations of the content component at T3

Unlike at T2, when the two groups had some dramatic differences in their emphases on the content component, at T3 both groups gave this component about the same attention. The simple task group devoted 159 concepts (57.8% of the structure dimension) and the complex task group devoted 145 concepts (56.4% of the structure dimension) to this component. Consistent with T1 and T2, subjects represented the content of MedlinePlus from the following aspects: subject (general, topical, and specific), type, format, and presentation of information. Figure 13 shows the distribution of different aspects in the content component for the two groups.



(a)



(b)

Figure 13. (a) The content component of the simple task group's mental models at T3; (b) The content component of the complex task group's mental models at T3

6.5.1.3.1 Subject of the content: general, topical, and specific

At T3, the simple task group contributed 16 concepts (10.1% of its content component) and the complex task group contributed 29 concepts (20.0%) to the general level content of the content component, emphases that were comparable to what they were at T2. The concepts that subjects listed at T3 were about the same as T2, including information, advice, articles, description, data, materials, definitions, analysis, literature, problems, and questions.

The two groups were similar in representing the content at the topical level. The simple task group listed 60 concepts (37.8% of the content component) and the complex task group listed 53 concepts (36.6%). Both groups did not change the emphasis on this aspect of the content component much from their representations at T2. The content of the concepts also did not change. The topical concepts that subjects contributed at T3 included medical, clinical, treatments, diseases, diagnosis, drugs, health, healthcare, treatment, preventions, medications, nutrition, medicine, prescription, symptoms, wellness, prevention, and supplements.

The simple task group listed 48 concepts (30.2% of the content component) to represent the content of MedlinePlus at the specific level, a decrease from 82 concepts (40.4%) at T2. The complex task group listed 24 concepts (16.6%) to represent this aspect of the content, an increase from 16 concepts (12.6%) at T2. Unlike at T2, the difference between the two groups was not statistically significant. The two groups also did not differ in the content of the concepts. Examples of the concepts at the specific level that subjects listed at T3 include: asthma, blood pressure, bronchitis, exercise, heart disease, humidifier, medical procedures, side effects, vaccination, vaccine, vitamins, what to expect for certain procedures, LASIK, opthamology, eyeglasses, CT scan, body functions, eye care, diabetes, and alternative medicine. Among the concepts, some were related to the assigned tasks that subjects performed in the second search session, such as LASIK, eye glasses, opthamology, CT scan, blood pressure, vaccine, exercise, asthma, bronchitis, and humidifier. (Tasks for session 2 are attached in Appendix C.)

6.5.1.3.2 Type, format, and presentation of information

Table 25 shows subjects' representations of the information type, format, and presentation at T3. The numbers in the parentheses are the number of instances and the percentage of the aspect in the content component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 25. Subjects' representations of information type, format, and presentation in MedlinePlus at T3

Simple task group	Complex task group
Type (17, 10.7%)	**Type** (30, 20.7%)
- Clinical trials	- Encyclopedia
- Journals	- Tutorials
- Encyclopedia	- Dictionary
- Glossary	- Glossary
- Guide; What-to-do guides	- News
- News	- Journals
- Tutorials	- Statistics
- Reference	- Step-by-step guides
	- Directory
Format (4, 2.5%)	**Format** (4, 2.8%)
- PDF	- Pictures
- Pictures, Images	- Videos
	- PDF
Presentation (14, 8.8%)	**Presentation** (5, 3.4%)
- Overviews	- Overviews
- Summaries	
- Outline format	
- Sectioned	
- Chart	
- Diagrams	

At T3, the complex task group (30 concepts, 20.7% of the content component) gave slightly more emphasis to the type of information than the simple task group (17 concepts, 10.7%), just as it had at T2. Looking at the specific types of information listed by subjects revealed that all the types of information mentioned at this time appeared at T2, except reference. The specific types of information listed by the two groups were closer to each other than at T2.

The format of the information continued to receive little emphasis in subjects' mental models. Both groups pointed out the same information formats as they did at T2: pictures, videos, and PDF, except that at this time, no one in the simple task group included the format of text.

At T3, the simple task group was able to represent new information presentation forms (not mentioned at T2), specifically, chart, sectioned, and outlined formats. Subjects in the complex task group identified only one type of presentation form: overview.

6.5.1.4 *Discussion of the construction of the content component of the structure dimension*

Table 26 shows an overview of the relative magnitude of subjects' representations of various aspects of the content component at T1, T2, and T3. The corresponding data from T0, subjects' mental models of information-rich web spaces, are included in the discussion but not in the table, since they were based on subjects' written answers to a question in the demographic questionnaire rather than the concept listing protocols.

Table 26. Subjects' representations of the content component at T1, T2, and T3

	T1	T2		T3	
		Simple	Complex	Simple	Complex
General	55 (13.3%)	20 (9.9%)	24 (18.9%)	16 (10.1%)	29 (20.0%)
Topical	146 (35.4%)	68 (33.5%)	45 (35.4%)	60 (37.8%)	53 (36.6%)
Specific	123 (29.8%)	82 (40.4%)	16 (12.6%)	48 (30.2%)	24 (16.6%)
Type	50 (12.1%)	14 (6.9%)	33 (26.0%)	17 (10.7%)	30 (20.7%)
Format	24 (5.8%)	6 (3.0%)	5 (3.9%)	4 (2.5%)	4 (2.8%)
Presentation	15 (3.6%)	13 (6.4%)	4 (3.1%)	14 (8.8%)	5 (3.4%)

In the initial models of information-rich web spaces, subjects represented the content component in relation to three aspects: topic, type of information, and format of information. After subjects had experience with MedlinePlus, they started to represent the content of MedlinePlus in relation to the following aspects: topic (general, topical, and specific), type, format, and presentation of information.

From the table, it is clear that the majority of the content component was dedicated to represent the topic aspect (at general, topical, and specific levels). In the initial model (T0), instead of suggesting any particular topic, subjects mentioned that information-rich web spaces often cover one or multiple topics. After their first 5-minute interaction with MedlinePlus, subjects elaborated their ideas about topicality by representing the subject of the content in MedlinePlus at three different levels: general, topical, and specific. At the general level, content was represented by concepts such as information, advice, and sources. Such concepts did not reflect the topical domain of the information in MedlinePlus, thus could not differentiate MedlinePlus from other information-rich web spaces. At the topical level, content was represented by concepts such as medicine, drugs, symptoms, and treatments. Such concepts were associated with the specific domain covered by MedlinePlus. At the specific level, content was represented by concepts such as diabetes, high blood pressure, and vitamin A. Such concepts were very specific topics that are included in the website. They are often instances of the concepts at the topical level.

At T2, after the simple task group performed a set of simple tasks and the complex task group performed a set of complex tasks, both groups' representations of the general and topical level of content were similar to T1 in terms of the number of concepts and the content of the concepts contributed to each aspect. However, the two groups differed significantly in their representations of specific level content. The simple task group contributed significantly more concepts to this aspect than the complex task group ($t(35) =$

2.14, $p = 0.04$). The difference might be due to the number of tasks that subjects performed in the first search session. The simple task group had performed 12 simple tasks before the concept listing protocol. These tasks provided a wide range of medical scenarios and terms for subjects and exposed the subjects to many different specific topics in MedlinePlus. However, the complex task group only performed 3 complex tasks, which provided a comparatively limited number of scenarios and terms for the subjects.

At T3, after the second search session where all subjects performed the same set of tasks containing both simple and complex tasks, both groups' representations of general level and topical level content remained similar to T1 and T2 in terms of the number of concepts and the substance of the concepts. However, in representing the specific level content, the simple task group reduced their representations (T2: 82 concepts, 40.4%; T3: 48 concepts, 30.2%), while the complex task group maintained a similar level of attention to this aspect as at T2 (T2: 16 concepts, 12.6%; T3: 24, 16.6%). Therefore, the difference between the groups was not as big as at T2. One reason could be that, at T3, the two groups had completed the same set of tasks and the impact of tasks on their representations of the content at the specific level was lessened.

In the initial model (T0), subjects represented three types of information: Q&A, help, and advertisements. At T1, subjects' representations of information types in MedlinePlus had broader coverage. They were able to identify most basic information types such as dictionary, encyclopedia, news, tutorials, and directories. At T2, the complex task group emphasized information types slightly more than the simple task group (33 concepts, 26.0%, vs. 14 concepts, 6.9% of the content component). The difference in magnitude might be because the difficulties associated with solving complex tasks might have exposed the subjects in that group to more types of information and prompted subjects to represent those different types of information. Task complexity also affected the content of the representations. The inspection of the specific information types listed by the subjects revealed that subjects in the simple task group tended to represent information types useful for answering simple tasks, such as FAQs, and fact sheets; while subjects in the complex task group were more likely to represent information types such as scholarly articles and clinical trials.

At T3, the complex task group continued to give slightly more emphasis to this aspect, but the difference between the groups was reduced from T2. Also reduced was the difference between the two groups in their mentions of specific types of information. The reduced gap between the groups might be attributable to the fact that both task groups performed the same set of tasks.

In the initial model (T0), subjects provided a list of general information formats, including images, graphics, text, videos, multimedia, and audio. At T1, after using MedlinePlus for 5 minutes, subjects augmented their representations with more formats, including movies, photos, pictures, flash, and PDF. At T2, subjects paid less attention to this aspect and all the formats that were mentioned at T2 had already appeared at T1. At T3, for both groups, the format of the information continued to receive little emphasis in subjects' mental models. Both groups pointed out the same information formats as they did at T2.

At T1, subjects developed a new aspect to represent information in MedlinePlus: presentation. During the short interaction with the system, they started forming views of how information is presented in the system. They noticed that information in MedlinePlus is presented in the form of overviews and summaries. They observed that there are some short articles, but details are also provided at times. They also noted that diagrams and demonstrations are used to present information as well.

At T2, the simple task group gave a little more emphasis to the presentation of information in MedlinePlus than the complex task group (13 concepts, 6.4% vs. 4 concepts, 3.1% of the content component). The simple task group also identified more ways of presenting information in the system, such as tables and figures. The slight difference between the groups might be due to the nature of the tasks: it is possible that simple tasks encouraged subjects to pay more attention to different information presentation forms.

At T3, the magnitudes of both groups' representations of the presentation aspect of the information in MedlinePlus remained similar to what they were at T2, with the simple task group giving slightly more emphasis to this aspect (14 concepts, 8.8%, vs. 5 concepts, 3.4%). The simple task group was also able to list new information presentation forms (not mentioned at T2), while subjects in the complex task group identified only one type of presentation form. At T3, the two groups performed the same set of tasks, thus the differences between the two groups could not be attributed to the differences in tasks. Further investigation of the impact of tasks on users' representation of information presentation forms is needed.

The model construction process outlined above suggests that users' initial models of general information-rich web spaces served as a basis for subjects to construct their mental models of the content in MedlinePlus. Meanwhile, during subjects' interactions with the system, they incorporated MedlinePlus-specific information into the model. Both the top-down and bottom-up processes contributed to the construction of subjects' mental representations of the content of MedlinePlus.

6.5.1.5 Summary of the construction of the content component of the structure dimension

Subjects represented the content of MedlinePlus according to multiple aspects: the subject of the content (at general, topical, and specific levels), type, format, and presentation of information. Subjects' representations (the number of concepts dedicated to the aspect and the content of the concepts) of the general and topical levels of the subject aspect remained constant over time for both groups. However, subjects' representations of the specific level content were affected by the tasks that they performed. At T2, the simple task group contributed significantly more concepts to this aspect. Many of the concepts were related to the assigned tasks that they had performed. At T3, the concepts contributed by the simple task group continued to outnumber those of the complex task group, but the difference was not statistically significant and the content of the concepts from the two groups was more homogenous than it was at T2.

The type of information that subjects represented was affected by the tasks. At T2, the complex task group gave more emphasis to the type of information than the simple task group. In addition, the specific types of information that each group presented are closely related to the answers required by the two types of tasks. At T3, the specific types of information listed by the two groups became more similar to each other. Tasks might also have an impact on subjects' emphasis on the information presentation forms. At T2 and T3, the simple task group not only contributed more concepts to describe this aspect, but also listed more specific information presentation forms in MedlinePlus.

The format of information was consistently the least represented aspect of the content of MedlinePlus. However, at T1, subjects were able to provide a comprehensive list of information formats available in MedlinePlus. At T2 and T3, their representations of this aspect of the content were reduced, and the specific formats mentioned at T2 and T3 were a subset of formats subjects represented at T1.

6.5.2 The construction of the content component of the evaluation/emotion dimension

This section reports subjects' evaluations of and emotions about the content in MedlinePlus, as well as their changes over time.

6.5.2.1 Evaluations of and emotions about the content component at T1

At T1, a large portion of the evaluation concepts (89 concepts, 43.8% of all the evaluation/emotional concepts) was dedicated to content. Examination of the concepts contributed by subjects revealed that subjects evaluated or expressed emotions about the content of MedlinePlus in relation to four aspects: quantity, attributes, quality, and utility. Table 27 lists concepts that subjects used to evaluate the content of MedlinePlus from these four aspects. The numbers in the parentheses are the number of instances and the percentage of the aspect in the content component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 27. Subjects' evaluations of and emotions about the content of MedlinePlus at T1

Quantity of information (11, 12.4%)

- Good array of information
- Lots of information
- Information on a large number of topics, including lesser-known conditions
- Overload of information
- Access to biggest database
- Overwhelming amounts of information
- Not many articles
- Limited
- Additional content is needed

Attributes of information (42, 47.2%)

Comprehensiveness

- Comprehensive
- Diverse
- Not restrictive to one area of medicine
- National and local information
- Information on a large number of topics, including lesser-known conditions

Currency

- Current
- Contemporary
- Recent
- Up-to-date

Objectivity

- Facts

Depth of information

- Basic
- General
- Common
- Broad, breadth not depth
- Specific
- Quick information

Language

- Available in different languages
- Words were mostly monochromatic

Others

- Popular topics
- Concise

Quality of information (13, 14.6%)

- Academic, Scientific, Well-researched
- Clear
- Consistent
- Thorough
- Reputable
- Reliable, trustworthy
- Authoritative
- Question of credibility

Utility of information (21, 23.6%)

- Helpful
- Useful
- Interesting
- Relevant
- Additional content other than medical data might make it more interesting, such as an article about health insurance
- Informative
- Easy to read
- Self-explanatory
- Understandable
- No information on brown recluse bite

Evaluation of specific sections of content (2, 2.2%)

- The general information on bodily systems and functions supplements and provides a foundation for the information on diseases.
- More description on black widow bite

The subjects had differing opinions about the quantity of information in MedlinePlus. Some subjects thought that MedlinePlus contained a good amount of information on various medical topics, including less well-known topics, while other subjects believed that there were not many articles in the system and that additional content, such as information about health insurance and more descriptions on certain topics, was needed.

Meanwhile, subjects expressed opinions about the characteristics of the information in MedlinePlus by presenting various attributes of the information: including comprehensiveness, currency, objectivity, depth, language, and others. Subjects generally agreed that MedlinePlus provided basic and common factual information. They indicated that the information covered a broad spectrum of medical topics and was regularly updated. They also recognized that some information in MedlinePlus was available in multiple languages.

At the time when they had interacted with the system for only 5 minutes, subjects had few criticisms of the quality of the information in MedlinePlus. They felt that the content of the system was not only comprehensive and thorough, but also clear, consistent, reputable, and reliable. Only one subject questioned the credibility of the information.

The utility of information refers to whether the information is useful and usable for a particular group of users. It is an important criterion when people select and use a piece of information or an information resource. Subjects felt that the information in MedlinePlus was relevant, useful, and informative; furthermore, they indicated that the information was self-explanatory and easy to read.

Although subjects had had only a brief interaction with the system, some of them evaluated specific sections of content in MedlinePlus. For example, one subject suggested that the system should provide “more description on black widow bite.” Another subject commented that:

> The general information on bodily systems and functions supplements and provides a foundation for the information on diseases.

The analysis of subjects’ evaluations of the content of MedlinePlus at T1 revealed that, when users encounter a new IR system, they are able to quickly form opinions about the information in the system. They not only develop an initial impression about the sheer amount of information in the system and the attributes (or characteristics) of the information, but also develop judgments about the quality and utility of the information. Furthermore, they start expressing their thoughts on particular sections or types of content in the system.

6.5.2.2 Evaluations of and emotions about the content component at T2

At T2, the simple task group dedicated 52 concepts (37.4% of the evaluation/emotion dimension) to evaluate the content of MedlinePlus. The complex task group dedicated 75 concepts (59.5% of the evaluation/emotion dimension), significantly more concepts than at T1 ($t(18) = 2.98$, $p = 0.008$), to this component in the evaluation/emotion dimension. Table 28 lists concepts contributed by subjects in the two groups at T2. The numbers in the parentheses are the number of instances and the percentage of the aspect in the content component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 28. Subjects' evaluations of and emotions about the content of MedlinePlus at T2

Simple task group	Complex task group
Quantity of information (11, 21.2%) - Lots of articles, Lots of information, Lots of text - Overwhelming - Expansive - Ismetric - Large volume of knowledge - Wealth of knowledge - Numerous articles on topics	**Quantity of information** (16, 21.3%) - Tons of information, Lots of data, Lots of information, lots of outside sources, A lot of information to dig through - Vast - A large quantity of information - Copious - Plenty of info - Limited, Limited external sources, Limited general info - Exclusive Medline plus information is scant - Few scholarly articles - Not a lot of simple information
Attributes of information (17, 32.7%) ***Comprehensiveness*** - Lots of variety - Comprehensive - Many different topics - Many subjects - Everything you need ***Currency*** - Contemporary - Latest ideas ***Depth of information*** - Common questions - General information - Sometimes too specific ***Other attributes*** - Official - Detailed information - Free Information - Concise	**Attributes of information** (32, 42.7%) ***Comprehensiveness*** - Comprehensive - Not comprehensive - Diverse information and sources in links - Breadth - Broad spectrum of information - A little something of everything ***Currency*** - Current - Updated - Recent ***Depth of information*** - Simple - Superficial - Good background information - General - Broad articles ***Objectivity*** - Government bias - Info pre-screened - Un-varying opinions - Facts - Filtered ***Language*** - Accessibility to Spanish-speakers - Different languages ***Other attributes*** - Many government links - Interactive - Repetitive - Indexed information

Simple task group	Complex task group
Quality of information (10, 19.2%) - Well-researched - Thorough - Good source of medical information - Consistent - Expert - Knowledgeable - Professional opinion - Supported by research - Accurate information - Reputable	**Quality of information** (9, 12.0%) - Trusted information - Accurate information - Believable resources - Reputable sources - Reliable - Well-established research - Scientific information - Prominent articles
Utility of information (8, 15.4%) - Language is easy to understand - Easily understood - Further research sometimes still needed - Informative - Interesting	**Utility of information** (10, 13.3%) - Educational - Informational articles easy to read - Informative - Understandable - Difficult to understand - Interesting - Need to know about health to navigate health topics
Evaluation of specific sections of content (6, 11.5%) - Drug summaries easy to navigate and especially good - Good summary of what articles would be about - Overview of diseases useful - Overview tabs useful - Sometimes too general in description - More information on some topics than others (lots on AIDS, less on low blood pressure)	**Evaluation of specific sections of content** (8, 10.7%) - Good overview by medline - (Encyclopedia) similar information to health issues - Encyclopedia function is useful as a starting point for searches - Tutorial may be very helpful - Anatomy page limited - Anatomy links vary in usefulness - It would be helpful if the encyclopedia included more articles based on general descriptions of organs or systems, rather than consisting almost entirely of articles on disorders. - Additional content for the encyclopedia would be nice

As shown in the table, after performing a set of assigned tasks (the simple task group performed simple tasks and the complex task group performed complex tasks), subjects in both groups evaluated the content of MedlinePlus in a similar manner as T1. They evaluated the content in relation to the following aspects: quantity, attributes, quality, utility, and specific sections of content in MedlinePlus.

Consistent with their evaluations at T1, subjects in both groups believed that MedlinePlus contained an overwhelmingly large amount of information. However, subjects in the complex group were more critical of the quantity of information. Some of them pointed out that there were few scholarly articles in the system, the external sources were limited, exclusive MedlinePlus information was scant, and there was not a lot of simple information.

Subjects also attempted to represent a number of attributes of the information in MedlinePlus. Consistent with T1, the complex task group represented the characteristics of MedlinePlus information as having the following attributes: comprehensiveness, currency, depth, objectivity, language, and others. The simple task group did not comment on the objectivity and language of the content in their representations. Generally, subjects in both groups still believed that the content of MedlinePlus was superficial, but comprehensive, current, and factual. The two groups showed somewhat different attitudes to the comprehensiveness of the information in MedlinePlus. Subjects in the simple task group agreed that the information in MedlinePlus was comprehensive and had lots of varieties; while a few subjects in the complex task group pointed out that MedlinePlus was not comprehensive.

At T2, both groups largely inherited their positive view of the quality of information in MedlinePlus from the views that they presented at T1. They agreed that information in

MedlinePlus was accurate, reputable, consistent, and well researched.

Both groups also inherited positive feelings about the utility of the MedlinePlus system. Subjects still felt that the content was informative, interesting, and could be easily understood. However, after using the system to solve real problems, subjects in both groups realized the difficulties in using MedlinePlus. One subject in the simple task group commented that “further research sometimes [is] still needed”. Subjects in the complex task group suggested that the information in MedlinePlus was not all that easy to understand; sometimes, users “need to know about health to navigate health topics”.

In evaluating specific sections of content in MedlinePlus, the two groups showed different focuses. The simple task group focused on evaluating summaries, descriptions, and overviews, while the complex task group focused more on evaluating the encyclopedia and tutorials. Subjects in both groups also tended to provide insights or propose means to improve the content of MedlinePlus when they evaluated specific sections in the system. For example, one subject in the simple task group commented that,

> [There is] more information on some topics than others (lots on AIDS, less on low blood pressure).

One subject in the complex task group suggested that,

> It would be helpful if the encyclopedia included more articles based on general descriptions of organs or systems, rather than consisting almost entirely of articles on disorders.

6.5.2.3 Evaluations of and emotions about the content component at T3

At T3, the simple task group listed about the same number of concepts to evaluate or express emotions about the content component as T2 (T3: 52 concepts, 34.9% of the evaluation/emotion dimension; T2: 52 concepts, 37.4%). The complex task group used 57 concepts (47.1%) to represent the content of MedlinePlus, decreased from T2 (75 concepts, 59.5%). Table 29 lists concepts that the two groups used to evaluate the content of MedlinePlus. The numbers in the parentheses are the number of instances and the percentage of the aspect in the content component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 29. Subjects' evaluations of and emotions about the content of MedlinePlus at T3

Simple task group	Complex task group
Quantity of information (4, 7.7%) - Abundant - Extensive volume of information - Lots of articles; Many articles	**Quantity of information** (2, 3.5%) - Large amount of information - So much information - Fewer scholarly articles
Attributes of information (24, 46.2%) ***Comprehensiveness*** - Limited scope - Broad - All medically related topics covered - Varied topics ***Depth of information*** - General - Common health issues - Basic concepts - Easy Information - Not as good for physician - Non-specialist information - Specialized sources ***Currency*** - Current research articles ***Objectivity*** - Factual; Fact based, Truth - Uncontroversial information ***Presentation*** - Brevity - Concise information in articles ***Other attributes*** - Specific - Non-comparative - Can be specific	**Attributes of information** (33, 57.9%) **Comprehensiveness** - There is a good variety in the information available concerning treatment options. - Not comprehensive - Varied information - Not good source for complex information - Different information from health issues **Depth of information** - General, General health information; General medical information; Generalized information - In-depth - Broad topics - Good source of basic information - Basic information - Background information - Information is mostly about patients, not doctors **Currency** - Current health beliefs - Recent findings **Objectivity** - Less research, more factual - Not opinion - Factual; Fun health facts - All concrete data **Language** - Language options - In everyday language
Quality of information (9, 17.3%) - Accurate - Reputable sources - Consistent - Professional - Preciseness - Verified - Reliable sources - Scientific terms - Good explanations	**Quality of information** (4, 7.0%) - Consistent - Reliable - Reputable sources

Simple task group	Complex task group
Utility of information (9, 17.3%) - Useful images - Informative - Simple language - Helpful information - Useful	**Utility of information** (8, 14.0%) - Easy to read - Valuable - Comprehensive - Informative - Helpful info - Relevant articles - External links most helpful - External links to websites and academic articles are useful - Great for specific problems
Evaluation of specific sections of content (6, 11.5%) - Good diagrams - Good overview - Descriptive titles - Overviews with broad details - Wrong terms - Good with treatments, symptoms	**Evaluation of specific sections of content** (10, 17.5%) - Encyclopedia limited - General overviews - Specific category is best - Helpful tutorials - Why does encyclopedia and topics do not share contents? - Not as many news items as originally thought - Ailment pages have wide range of information - Health issues pages comprehensive - General entries on diseases, organs

As shown in the table, after performing the second set of tasks (the two groups performed the same tasks), subjects in both groups still evaluated the content of MedlinePlus from the following aspects: quantity, attributes, quality, and utility of information, and specific sections of content in MedlinePlus.

When evaluating the quantity of information, subjects still agreed that MedlinePlus has an extensive volume of information. However, the number and proportion of concept listings that represented the quantity of information were reduced for both groups, compared to T2.

Subjects continued to represent characteristics of the information in MedlinePlus in relation to the following attributes: comprehensiveness, currency, depth, objectivity, language, and others. The complex task group generated concept listings for all these attributes. The opinions of the group on these attributes remained similar to what they were at T2: the information in MedlinePlus was general, but comprehensive, current, and objective. Different language options were offered by the system. Several subjects in the complex task group also had the opinion that the content was not comprehensive.

Different from the complex task group, the simple task group kept leaving out the language attribute from their representations, but added another attribute, presentation of information, to the evaluation. Subjects in this group thought that the information was up-to-date general factual knowledge on varied topics, and that the information was presented in a concise and brief manner. But in contrast to T2, some subjects in this group started to question the comprehensiveness of the information, pointing out that the content of the system had a limited scope.

Similar to T1 and T2, subjects in both groups held positive views of the quality of information in MedlinePlus. They still represented the information as accurate, reputable, consistent, and well researched. When talking about the utility of information, subjects still thought that the information in MedlinePlus was useful and usable. They did not bring up any concerns or difficulties with using the information in the system.

When evaluating specific sections of content, subjects continued to provide insights or comments for specific content in MedlinePlus. Similar to T2, the two groups also focused on different aspects of evaluation: the simple task group focused more on overviews, titles, and diagrams, while the complex task group focused more on a higher level of content, such as tutorials, encyclopedia, and news. For example, one subject in the complex task group discovered that there were "not as many news items as originally thought." Another subject wondered "why does encyclopedia and topics do not share contents."

6.5.2.4 Discussion of the construction of the content component of the evaluation/emotion dimension

Subjects evaluated the content of MedlinePlus in relation to the following aspects: quantity, attributes, quality, utility, and specific sections of content. Table 30 shows the distribution of their evaluations of and emotions about content according to these different aspects across time. T0 is discussed in the text, but is not included in the table since the data on which it is based are subjects' written answers to a question in the demographic questionnaire, rather than the concept listing protocols used at T1, T2, and T3. The numbers in the columns are the number of concepts that subjects used to describe each specific aspect and the percentage of that particular aspect within the content component evaluation.

Table 30. Subjects' evaluations of and emotions about the content component at T1, T2, and T3

	T1	T2		T3	
		Simple	Complex	Simple	Complex
Quantity	11 (12.4%)	11 (21.2%)	16 (21.3%)	4 (7.7%)	2 (3.5%)
Attributes	42 (47.2%)	17 (32.7%)	32 (42.7%)	24 (46.2%)	33 (57.9%)
Quality	13 (14.6%)	10 (19.2%)	9 (12.0%)	9 (17.3%)	4 (7.0%)
Utility	21 (23.6%)	8 (15.4%)	10 (13.3%)	9 (17.3%)	8 (14.0%)
Specific content	2 (2.2%)	6 (11.5%)	8 (10.7%)	6 (11.5%)	10 (17.5%)

In subjects' initial models, they evaluated the content of information-rich web spaces from four aspects: quantity, attributes, quality, and utility. After they had experience with

MedlinePlus, as shown in the table, subjects evaluated the content component in relation to five aspects: quantity, attributes, quality, utility, and specific sections of the content.

At T1, when subjects only had 5 minutes of experience with MedlinePlus, subjects dedicated 11 concepts (12.4% of the content component in the evaluation/emotion dimension) to evaluate the quantity of information in MedlinePlus. They believed that MedlinePlus contained a large amount of information. At T2, both groups increased slightly their emphases on this aspect. Subjects in the complex task group pointed out that the information in MedlinePlus was limited, while those in the simple task group did not. At T3, both groups reduced the emphasis on the quantity of information, but still recognized that the site contained abundant information.

The attributes aspect was consistently the most represented aspect over time. At T1, subjects dedicated about half of the concepts (42 concepts, 47.2%) to this aspect and pointed out a few attributes of the information: comprehensiveness, currency, depth of information, objectivity, and language. They thought that the information in MedlinePlus was comprehensive, current, and objective common knowledge that was available in different languages. At T2, the groups showed some differences in this aspect: the complex task group represented more attributes of the information than the simple task group and were more critical of the comprehensiveness of the information. The criticisms might be associated with the difficulty that the complex task group experienced in finding information to answer complex tasks.

At T3, subjects in both groups still recognized various attributes of the information, such as comprehensiveness, currency, depth, objectivity, and language. The simple task group added another attribute, information presentation, to the evaluation. Compared to T2, the two groups also became closer to each other in their opinions of the attributes of information. In addition to keeping their positive opinions at T2, both groups recognized the limited scope of the information in the system.

At T1, subjects agreed that the information in MedlinePlus was of good quality, clear, consistent, and reputable. Only one subject questioned the credibility of the information in the system. At T2, subjects' attitudes toward the quality of information remained the same as at T1. They still believed that information in MedlinePlus was accurate and reputable. At T3, subjects inherited the positive feelings toward the quality of MedlinePlus content from T2 and did not bring up particular concerns about this aspect of MedlinePlus.

Utility of information has two sides: whether the information is useful, and whether the information is usable. Useful information does not have much utility if it is not readily usable. At T1, after a brief interaction with the system, subjects believed that the content in MedlinePlus was useful and usable, though they also pointed out that some additional information on certain topics was needed. At T2, similar to T1, subjects in both groups still thought that the content in MedlinePlus was useful and easy to understand. However, subjects in both groups started realizing the difficulties in using the content, which might be directly associated with their experience of using the system to find information for the first set of assigned tasks. At T3, subjects inherited positive feelings toward the utility of MedlinePlus content from T2 and did not bring up particular concerns about this aspect. At T1, subjects also started paying attention to and evaluating specific sections of content in the system, but there were only two instances of this very specific type of evaluative comment. At T2, both groups increased their evaluation of specific sections of content in the system, which could be a natural trend, given their increased exposure to that content. The evaluations of the two groups had different emphases. The simple task group focused on evaluating summaries, descriptions, and overviews, while the complex task group focused more on evaluating the encyclopedia and tutorials. The differences in focus might be due to the type of information that was required to complete the simple and the complex tasks. Subjects often could find answers for the simple tasks in summaries and overviews of an article, while subjects needed to scan through different types of sources, such as encyclopedias and tutorials, and integrate information from these sources to form answers for the complex tasks. At T3, the two groups maintained their corresponding T2 emphases in evaluating the specific sections in the system: the simple task group focused on summaries, overviews, and titles, and the complex task group focused on tutorials and the encyclopedia.

The analysis showed that subjects evaluated the information in MedlinePlus from various aspects. The multiple perspectives that subjects employed to evaluate the information suggest that users evaluate information in an IR system using criteria far beyond relevance. Research on users' information evaluation in information retrieval needs to extend from relevance judgment to roles that other criteria play in people's use of information. Such research is urgently needed, especially in the open-web environment where finding information is not as big a challenge as evaluating and selecting information. Knowledge about how users evaluate information could help design IR functions to facilitate users in evaluating and selecting information.

The analysis also showed that subjects' evaluations of the content of MedlinePlus were dynamic, changing with their experience with the system and the nature of the tasks that they performed. The results suggest that future research on information evaluation not only needs to explore people's use of criteria beyond the basic criteria covered by traditional library training materials, such as credibility, currency, accuracy, and relevance, but also needs to explore how the evaluation is affected by systems and tasks in particular situations. Such research will inform the design of library training materials and information literacy curricula, and the design of functions to support prompt information evaluation as users look for information.

6.5.2.5 Summary of the construction of the content component of the evaluation/emotion dimension

During subjects' interaction with MedlinePlus, they evaluated the content of the system from various aspects: quantity, attributes, quality, utility, and specific sections of the content. In the first 5 minutes of interaction, subjects formed the opinion that MedlinePlus contained a large amount of information. The information was basic, comprehensive, up-to-date, and objective. Some information was available in multiple languages. The information in the system was also of good quality, clear, accurate, reliable, and consistent, and of good utility, useful and usable. These opinions stayed with the subjects through their use of the system over time (T2 and T3).

However, subjects' opinions of the system also developed over time. Their development reflects the effects of tasks on the construction of mental models. At T2, after finishing the first set of tasks, although holding a positive view of MedlinePlus, the complex task group became more critical of the quantity of information, pointing out that the information in MedlinePlus was limited and scant. The group also questioned the comprehensiveness of the content at T2, as did the simple task group at T3, after performing a set of tasks containing both simple and complex tasks. At T2 and T3, in evaluating specific sections of the content, the complex task group focused on specific types of information such as encyclopedia and tutorials, while the simple task group focused on information presentation forms such as summaries and overviews. These differences might be due to the characteristics of the complex tasks: it is difficult to find information to answer the complex tasks and the answers for the tasks need to be formed by integrating information from multiple sources.

6.6 The construction of the information organization component of the structure and evaluation/emotion dimensions

How to organize a large amount of information is a challenge for the creator of any information-rich web spaces. The decisions concerning information organization have direct impact on the usability of a website. This section starts by describing subjects' representations of the information organization component in the structure dimension of mental models at different times (section 6.6.1). Subjects' evaluations of and emotions about information organization in MedlinePlus will be discussed in section 6.6.2. A summary of the changes that the information organization component in both dimensions experienced over time will be provided in section 6.6.3.

6.6.1 The construction of the information organization component of the structure dimension

6.6.1.1 Representations of the information organization component at T1

While information organization is a critical component of a system, subjects' representations of this component in the structure dimension were infrequent at all three data collection points. At T1, subjects contributed 35 concepts to describe the information organization component, which was only 5.6% of the structure dimension. The representations are listed in Table 31. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 31. Subjects' representations of information organization in MedlinePlus at T1

- Categories; Array of topics; Grouped; Organized by topic, concepts
- Headlines organize
- All body systems; anatomy; body parts; body systems
- Alphabetical; Array of letters; Encyclopedic
- Listing
- Health related issues depending on gender/age/other demographics
- Multi-faceted
- Hierarchical
- Cross-listing
- Drugs are listed under both generic and brand names

As shown in the table, despite the lack of emphasis in the subjects' mental models, different information organization schemas in MedlinePlus were well represented at T1. Subjects represented that information in MedlinePlus is organized into categories by subjects, concepts, or medical terms. Furthermore, they were able to articulate specific information organization schemas, such as alphabetical (array of letters, listing, encyclopedic), by body systems (anatomy, body parts), by gender, age, and other demographics, by hierarchical structure, and by generic and brand names of drugs. One subject also noted that multiple facets are used to organize information in MedlinePlus. Meanwhile, several subjects noticed that relevant categories are cross listed.

6.6.1.2 Representations of the information organization component at T2

At T2, subjects' representations of information organization were slightly reduced in number. The simple task group contributed 11 concepts (3.7% of the structure dimension) and the complex task group contributed 13 concepts (5.2% of the structure dimension) to the component. Table 32 shows each group's representations of the information organization component at T2. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 32. Subjects' representations of the information organization in MedlinePlus at T2

Simple task group	Complex task group
- Alphabetical - Alphabetically arranged topics - Categories - Organization by topic - Anatomically arranged topics; Body parts - Types of ailments	- Alphabetical - Lists of diseases - Body systems - Categories; Subcategories - Hierarchal information - Some of the information is organized based on what demographic it applies to most.

It is clear from the table that, after performing a set of tasks using the system, subjects' representations of information organization in MedlinePlus had not improved from T1, when they had used the system for only 5 minutes. On the contrary, some of the schemas mentioned in T1 (such as cross listing) were not mentioned at T2.

At T2, subjects described information organization as by category, topic, alphabetical list, or anatomy. Subjects in the complex task group provided a more comprehensive view of information organization than those in the simple task group. In addition to the general information organization schemas, the complex task group pointed out that information is organized hierarchically, and some of the information is organized based on the demographic to which it applies.

6.6.1.3 Representations of the information organization component at T3

At T3, after using the system to perform a common set of tasks, the subjects' representations of the system's information organization continued to be infrequent. The simple task group listed 16 concepts (5.8% of all the concepts in the structure dimension), and the complex task group listed 7 concepts (2.7%). Table 33 details each group's representations of the information organization component at T3. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 33. Subjects' representations of the information organization in MedlinePlus at T3

Simple task group	Complex task group
- Alphabetical listings	- Category
- Organization by anatomy	- Alphabetical
- Organization by topics; Grouped topics	- Alphabetical listing that breaks down entries according to first two letters of words, not just first letter, is helpful
- Large categories	
- Subheadings under various headings	
- Network	- Lists of diseases
- Table of contents type layout	- Web of information

An inspection of how each group represented information organization revealed that both groups improved their understanding of information organization in MedlinePlus. In addition to typical information organization schemas such as categorization by topic, alphabetical listing, and anatomical organization, some subjects came to realize that information in MedlinePlus is organized in the form of a web or network. As one subject commented in a semi-structured interview:

> If you type in a generalized topic, you go to that page and you have more specifics from that. So it is still hierarchical, but it is not much list form that I thought before. It is more of a web.

Also, a subject in the simple task group represented information organization at the page level, pointing out the table of contents type of layout on health topic pages.

6.6.2 The construction of the information organization component of the evaluation/emotion dimension

When attempting to represent information organization schemas in MedlinePlus, subjects simultaneously evaluated or expressed emotions about the ways information is organized in the system. This section reports the construction of subjects' evaluations and emotions about the information organization component over time.

6.6.2.1 Evaluations of and emotions about information organization at T1

At T1, subjects contributed 15 concepts to express their evaluations of or emotions about information organization in MedlinePlus, which was about 7.4% of the evaluation/emotion dimension of their mental models at T1. Table 34 lists subjects' evaluations of and emotions about information organization in MedlinePlus at T1. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 34. Subjects' evaluations of and emotions about information organization at T1

- Clearly listed
- Good multimedia structure
- Logical
- Simplified categories
- Varied ways of dividing things
- Well organized
- Provides a comprehensive flow map to information
- Not enough subgroups

As shown in the table, in the first 5 minutes of interaction with MedlinePlus, subjects developed a generally positive feeling about its information organization. They agreed that information in MedlinePlus was clearly listed and logically organized. They perceived multiple ways of dividing information, but overall the system provided a comprehensive map to information. Only one subject concerned that there were not enough subgroups.

6.6.2.2 *Evaluations of and emotions about information organization at T2*

At T2, the simple task group contributed 3 concepts (2.2% of the evaluation/emotion dimension) and the complex task group contributed 9 (7.1% of the evaluation/emotion dimension) to express their evaluations of or emotions about information organization in MedlinePlus. Table 35 details the representations. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 35. Subjects' evaluations of and emotions about information organization at T2

Simple task group	Complex task group
- Organized	- Usually well organized - Like the subcategories - Concepts not very tied together - Disorganized - Listings could be broader, then separated into more specific within the topic - Not consolidation of info - Sometimes listings are too specific - Sometimes too much information, too many subgroups - Could make listings more…

As shown in the table, subjects in the complex task group expressed more evaluations of and emotions about information organization in MedlinePlus after the first search session than did subjects in the simple task group. An inspection of the evaluations revealed that the simple task group evaluated information organization at a very general level, commenting only that information in MedlinePlus was well organized. The complex task group evaluated information organization at multiple levels, from subcategories ("like the subcategories") to listings ("listings could be broader") and to general information organization ("well organized").

Furthermore, subjects in the complex task group were more critical of the information organization in MedlinePlus at T2. They pointed out that information in MedlinePlus was not optimally organized and that concepts were not very tied together. They believed that sometimes there was too much information and there were too many subgroups; consolidation of information was needed. Thinking that listings were too specific, a subject suggested that "listings could be broader, and then separated into more specific within the topic".

6.6.2.3 Evaluations of and emotions about information organization at T3

At T3, subjects in the simple task group listed 11 evaluative/emotional concepts (7.4% of the evaluation/emotion dimension) related to information organization and the complex task group listed 6 (5.0% of the evaluation/emotion dimension). Table 36 lists each group's evaluations of and emotions about information organization in MedlinePlus at T3. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 36. Subjects' evaluations of and emotions about information organization at T3

Simple task group	Complex task group
- Distinct use of word relations - Good listing orders - Site is organized logically - Strategic arrangement - Useful categories - Well-organized - Varied arrangement - Disjointed	- Alphabetical listing that breaks down entries according to first two letters of words, not just first letter, is helpful - Both encyclopedia and topics are well structured and organized - Well-organized - More organized - Too many subgroups - Too many subheadings

Compared to T2, subjects in the simple task group evaluated information organization beyond the general level ("well-organized"), and extended the evaluation to categories, word relations, and listing orders. Overall, the simple task group gave positive evaluations to the information organization in MedlinePlus. But subjects in the group also recognized some of the limitations of the information organization in MedlinePlus, pointing out that, in some cases, information was disjointed.

Subjects in the complex task group were very specific in their evaluations of information organization at T3, focusing on either specific types of resources or specific types of organization schema. For example, one subject pointed out that, "Alphabetical listing that breaks down entries according to first two letters of words, not just first letter, is helpful", and another commented that, "Both encyclopedia and topics are well structured and organized". Compared to T2, the complex task group was more positive about the way in which the information in MedlinePlus is organized, but some of them were still critical, pointing out that there were too many subgroups and subheadings.

6.6.3 Discussion of the construction of the information organization component of the structure and evaluation/emotion dimensions

Although the way that information is organized in a system can critically impact the system's usability, the information organization component was consistently the least represented and evaluated component in the structure dimension of subjects' mental models of MedlinePlus over time. There are two possible reasons for the under-representation of information organization. First, the information in MedlinePlus is reasonably well organized so that users do not have to struggle to find their way around the system or to find information for their questions. Consequently, they do not need to pay attention to information organization. This speculation is supported by the fact that at T2, the simple task group had a low representation of information organization and they unanimously evaluated the information in the system as well-organized. Paths for finding answers for simple tasks were well supported by the current information organization of MedlinePlus.

The second possible reason is that it may be end users' nature to down-play information organization when interacting with a system. That is, users do not pay attention to information organization, regardless of how well or how poorly the information in the system is organized. To investigate this possibility, a future study could be designed to explore whether users' representations and evaluations of information organization in their mental models of a system increase when the information is poorly organized.

Subjects' representations of information organization in MedlinePlus did not have much weight in their mental models of MedlinePlus, but they were able to form fairly comprehensive views of the ways in which information in MedlinePlus is organized. They were able to articulate various information organization schemas, even after just 5 minutes of free exploration with the system. Their representations were at two levels. At a general level, subjects identified concepts such as categories, topics, headings, and subheadings. At a more specific level, they were able to articulate specific information organization schemas, such as alphabetical listing, anatomy, body parts, demographic information, and hierarchical structure.

Subjects' representations of information organization at T1 had a very good coverage. The representations at T2 and T3 did not show significant improvement from those at T1. However, there were a few developments in subjects' understandings of information organization over time. For example, for the first time at T3, subjects in both groups realized that information in MedlinePlus was organized in a network manner: the same information can be reached using different routes. Also at T3, a subject in the simple task group mentioned the table of contents format of information organization on the health topic pages.

Task complexity seemed to have an impact on both subjects' representations and their evaluations of the information organization in MedlinePlus. At T2, the complex task group gave information organization slightly more weight in their representations (the complex task group: 13 concepts, 5.2%; the simple task group: 11 concepts, 3.7%) and expressed more evaluations and emotions concerning information organization (the complex task group: 9 concepts, 7.1%; the simple task group: 3 concepts, 2.2%) than the simple task group. The complex task group's evaluations of information organization were also more critical than those of the simple task group. At T3, after the second set of tasks, which included both types of tasks, the simple task group increased their representations (from 11 concepts, 3.7% at T2, to 16 concepts, 5.8% at T3) and evaluations (from 3 concepts at T2, 2.2%, to 11 concepts, 7.4% at T3). The evaluations also became more critical. The difference between the groups at T2 and the difference shown by the simple task group over time (from T2 to T3) might be because the complex tasks demanded more exploration and information integration on the subjects' part, therefore making them pay more attention to the details of information organization in the system and become more critical in evaluating it.

6.7 The construction of the interface component of the structure and evaluation/emotion dimensions

This section starts with reporting subjects' representations of the interface component in the structure dimension of mental models at different times to demonstrate how subjects' representations of the interface of MedlinePlus changed over time (section 6.7.1). Subjects' evaluations of and emotions about the interface of MedlinePlus, as well as their changes over time, will be discussed in section 6.7.2.

6.7.1 The construction of the interface component of the structure dimension

6.7.1.1 Representations of the interface component at T1

At T1, subjects contributed 102 concepts (16.3% of the structure dimension of their mental model of MedlinePlus) to represent the interface component. Within the interface component, 61 of the concepts (59.8%) were about interface elements, 39 (38.2%) were about functions, and only two concepts (2.0%) were about results. Table 37 shows concepts that subjects used to represent each of the three aspects of the interface component.

Numbers in the parentheses are, respectively, the number of concepts used to describe a particular aspect and the percentage of the aspect in the interface component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 37. Subjects' representations of the interface of MedlinePlus at T1

Interface elements (61, 59.8%)	
- Color	- About MedLine
- Links	- About us
- Lines on webpage	- Contact us
- Hospital and doctor listings	- Contact information
- Sections	- Banner
- Homepage	- Tabs
- Images	- Purple
- Related illness	- Java script
- Search bar, Search box	- Ads
- Subject headings	- Disclaimer
- Suggested articles	- Logo
- Links to related topics	- Menus
- Related areas	- Questions and Answers
- Title	- Interface
	- Layout
Functions (39, 38.2%)	
- Search (27)	- Symptom finder (1)
- Local search (5)	- Pill identifier (1)
- Browse: (4)	- Function (note: general description) (1)
Results (2, 2.0%)	
- Search results yield snippets	- Suggested results

As shown in the table, most concepts represented general elements, such as about us, disclaimer, logo, banner, images, colors, "contact us," homepage, links, questions and answers, menus, tabs, subject headings, sections, and titles. Subjects also mentioned a few elements specific to MedlinePlus: purple, lines on webpage, suggested articles, related illness, related areas, and hospital and doctor listings. One subject recognized that some pages in MedlinePlus required java script.

The second row of the table shows concepts that subjects used to represent the functions offered in MedlinePlus. It is apparent that, after a brief interaction with the system, subjects' representations of the functions in MedlinePlus focused on search, mostly general site-wide search. They also recognized that they were able to search for doctors or hospitals by state and search for different medications. Several subjects noted that browsing was an option for looking for information in MedlinePlus. One subject commented that it was easy to "jump" between topics.

Subjects also listed functions not provided by MedlinePlus. For example, subjects listed symptom finder and pill identifier, functions that MedlinePlus does not have. A similar comment was made by one subject in the semi-structured interview at T1:

> […], but they could have symptom checker on the homepage, have instructions on if you want to find this information, go here and if you want to find this information, go there.

The third row of the table shows concepts that subjects used to represent the results in MedlinePlus. At T1, only two subjects mentioned search results: "search results yield snippets" and "suggested results." Subjects had interacted with the system for only 5 minutes when they performed the concept listing protocol, so it was not surprising that few concepts concerning results were provided.

6.7.1.2 Representations of the interface component at T2

At T2, after the first search session, where the simple task group performed simple tasks and the complex task group performed complex tasks, the simple task group contributed 53 concepts (17.8% of the structure dimension) to represent the interface component, and the complex task group contributed 72 concepts (28.6%). Table 38 shows concepts that subjects used to represent the interface component at T2. Terms in bold represent interface elements that did not appear at T1. Numbers in the parentheses are, respectively, the number of concepts used to describe a particular aspect and the percentage of the aspect in the interface component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 38. Subjects' representations of the interface of MedlinePlus at T2

Simple task group	Complex task group
- Interface elements (18, 34.0%) - Links - Related links - Related topics - Subject headings - Search bar - **Cross references** - **Toolbar** - **Subheadings** - **Bold words** - **Clickable images for separating parts of bodily system** - **Landing page (health topic page)** - **Parts for medical care, drug supplements, etc**	- Interface elements (30, 41.7%) - Links - Article titles - Homepage - Tabs - Related topics - Search bar; Search box - Main subject headings - **Letter headings** - **Health issues page** - **Subheadings; Subtopics** - **See also** - **Headers** - **Anatomy links** - **Numbers on left for each category** - **Narrowed search options**
Functions (29, 54.7%) - Search (22) - Browse (6) - Suggest terms (1)	**Functions** (30, 41.7%) - Search (25) - Local search (3) - Browsing (1) - Spelling check (1)
Results (6, 11.3%) - Results - Sometimes hits were in domain sometimes out - Ranking formula - Site hits were major organizations - Pre-formulated search responses - Only provided hits on reliable sites	**Results** (12, 16.7%) - Results; Responses - Relevance - Relatedness - Several links to the same article - Search results listed in categories; Search results broken down into categories

At T2, both groups listed new interface elements (elements that did not appear at T1). Compared to the elements at T1, the newly added ones were more specific to MedlinePlus, such as clickable images for body parts, anatomy links, numbers on the left for each category, and narrowed search options. Unlike elements at T1, which mostly appear on the homepage of MedlinePlus, some new elements were embedded on pages at a deeper level in the website, such as cross-references, subheadings, bold words, letter headings, and "see also" references. Subjects also started recognizing and representing typical web pages in the system, such as a health topic page, and parts of the site dedicated to drug supplements.

There were also some differences between the groups. The simple task group generated fewer concepts related to interface elements than the complex task group. While the two groups shared many interface elements, each also contributed unique ones. For example, the simple task group listed toolbars and a landing page, and the complex task group listed the concepts of tabs and headers.

In representing the functions in MedlinePlus, the two groups listed a similar number of functions. The two groups also did not differ much in the specific functions that they listed. Both groups still focused on search, mainly general site-wide search. Subjects in the complex task group also represented a few local searches for hospitals and doctors. One subject in the simple task group came up with a new function, suggested terms, and one subject in the complex task group came up with another, spelling check. Some functions that subjects represented at T1, such as symptom finder and pill identifier, were phased out from the mental models at T2.

In representing the results, at T2, after the first search session, both groups increased their representations of search results, with the complex task group paying more attention to it than the simple task group. Subjects in both groups talked about results from multiple perspectives. Some mentioned the results in general. Some thought about the organization of results. For example, subjects in the complex task group pointed out "search results broken down into categories". Some thought about sources for the results. For example, subjects pointed out that the results could come from both within and outside of the MedlinePlus domain; many outside results were from major organizations. One subject noticed that, sometimes, several links led to the same articles. Subjects also were concerned about how the results were ranked. One subject in the simple task group expressed curiosity about the "ranking formula". One subject thought the results were ranked by relevance.

6.7.1.3 Representations of the interface component at T3

At T3, after the second search session, the complex task group continued to pay slightly more attention to the interface component (78 concepts, 30.4% of the structure dimension) than the simple task group (54 concepts, 19.6% of the structure dimension). Both groups continued to represent the interface component in relation to three aspects: elements, functions, and results. Table 39 shows concepts that subjects used to represent the interface component at T3. Terms in bold represent new interface elements that did not appear at T1 and T2. Numbers in the parentheses are, respectively, the number of concepts used to describe a particular aspect and the percentage of the aspect in the interface component. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 39. Subjects' representations of the interface of MedlinePlus at T3

Simple task group	Complex task group
Interface elements (23, 42.6%) - Links - Headings - Subheadings - Homepage - Search bar - Cross references - Related articles - Related words - Related topics - **Recommended links** - Section headings - Categories - Outline - Disease main page (health topic page)	**Interface elements** (29, 37.2%) - Contacts - Links - About Medline Plus - Colors - Health issues heading - Homepage - Ailment pages - Health issues pages - Health topic page - Layout - Links to related topics - Sidebar - Tabs - Other names that the topic might be listed under - Related topics - **Drop down menus; Drop-Down Options** - **"did you mean...?"** - **"for you" link** - **"read more"** - **Interactive sections** - **Also known as** - **Other features**
Functions (26, 48.1%) - Search (23) - Browse (3)	**Functions** (39, 50.0%) - Search (31) - Local search (1) - Browse (1) - Sort by (1) - Spelling check; "did you mean...?" function (3) - Function (Note: general) (2)
Results (5, 9.3%) - Results - A kids website seemed to be one of the most popular - Search choices based on relevance	**Results** (10, 12.8%) - Results - Bases relevance on text appearance in an article - Search engine yields external links - Search function turns up results that are categorized - Right amount of links in search results

As shown in the table, subjects in the complex task group listed more interface elements and also more new interface elements. These new elements were mostly embedded deep in the site or situational, such as, drop down menus, “did you mean...”, “read more”, “for you” and “also known as”.

At T3, the complex task group listed more concepts to represent functions in MedlinePlus than the simple task group. However, inspection of the specific functions revealed that, consistent with T1 and T2, subjects in both groups still focused on search functions, mostly site-wide search. Only one subject in the complex task group mentioned local search and anther subject in the complex task group identified the “sort by” function, which was not mentioned previously.

At T3, both groups contributed about the same number of concepts to the results aspect as they did at T2. Also similar to T2, subjects represented the results from multiple perspectives. Some thought about results in general by listing the term “results”. Some expressed concern about the ranking of the results by listing phrases such as “search choices based on relevance” and “bases relevance on text appearance in article”. One subject in the complex task group talked about the presentation of search results by pointing out that “search results are categorized”. Subjects also were concerned about where the results came from. One subject in the complex task group pointed out that the “search engine yields external links”. One subject in the simple task group noted that “a kids’ website seemed to be one of the most popular”. The same subject made a similar comment in the semi-structured interviews:

> It seems [they] keep bringing a kids' website for a return of a search, and I like that because it simplifies things that might be hard to understand.

6.7.1.4 Discussion of the construction of the interface component of the structure dimension

Table 40 shows an overview of the relative magnitude of subjects' representations of the three aspects of the interface component – elements, functions, and results – at T1, T2, and T3. The numbers in the columns are the number of concepts that subjects used to describe each specific aspect of the interface component and the percentage of that particular aspect in the interface component of the structure dimension of subjects' mental models of MedlinePlus. Though it is included in the discussion, T0 is not included in the table since it was based on subjects' written answers to a question in the demographic questionnaire rather than concept listing protocols.

Table 40. Subjects' representations of the interface component at T1, T2, and T3

	T1	T2		T3	
		Simple	Complex	Simple	Complex
Elements	61 (59.8%)	18 (34.0%)	30 (41.7%)	23 (42.6%)	29 (37.2%)
Functions	39 (38.2%)	29 (54.7%)	30 (41.7%)	26 (48.1%)	39 (50.0%)
Results	2 (2.0%)	6 (11.3%)	12 (16.7%)	5 (9.3%)	10 (12.8%)

In the initial models (T0), subjects represented the interface of information-rich web spaces in relation to two aspects: elements and functions. After they interacted with MedlinePlus, subjects represented the interface of MedlinePlus in relation to three aspects: interface elements, functions, and results. Interface elements are design elements of the interface, such as tabs, menus, and links. Functions refer to possible behaviors of the systems enabled by interface elements, such as search and navigation. Results representations are about various aspects concerning results, such as presentation and access.

From the table, it is clear that subjects paid a great deal of attention to the interface elements. In the initial model (T0), subjects represented menus, sidebars, tabs, a homepage, links, subheadings, and bold fonts. These elements are general and widely used to help users navigate through websites. At T1, subjects continued to represent general interface elements or elements at the surface of the website, such as “About MedlinePlus”, “Contact Us”, “Disclaimer”, “tabs”, “logo”, and “search bar”. They also noticed a few elements that were embedded at second- or third-level web pages, such as “related illness” and “hospital and doctor listings”.

At T2, both groups reduced their representations of interface elements. An examination of the specific interface elements revealed that the elements that subjects listed at T2 became more specific. Most of the elements went beyond the homepage and appeared on deeper level webpages in MedlinePlus, such as cross references, “see also” links, subheadings, and anatomy links. This finding suggests that, after performing a set of tasks, subjects were getting to know the system better. Subjects also started representing typical pages within the system, such as health topic pages and the pages for drugs and supplements. In MedlinePlus, every disease or condition has its own topic page, and these health topic pages are based on the same template. The representations of typical pages

(pages sharing the same templates, such as the health topic pages) in the system signifies that subjects started having a sense of how information is presented in the site and where to find the information. This understanding could impact subjects' future use of the site.

At T3, the two groups paid comparatively equal attention to the interface elements. Both groups represented the interface elements beyond the surface of the website. Subjects in the complex task group were able to provide more new elements, such as "read more" and "for you".

In the initial model (T0), subjects thought of search, suggestions of pages of interest, and frequently searched links and topics. At T1, after subjects' used MedlinePlus for 5 minutes, subjects focused on search and browse. Meanwhile, they mentioned another two functions: "symptom finder" and "pill identifier". The two functions were not available in MedlinePlus, which indicates that functions that users represented in their mental models do not necessarily belong to the system – they could come from other similar systems. This observation suggests that the construction of users' mental models of a system is affected by their prior knowledge base. Users' mental models of a system mimic the structure-relation of the system, but not the details of the system.

At T2, both groups increased their representations of the functions in MedlinePlus. The increase might be because subjects had used the system to perform a set of tasks. In the process of solving problems, they used various functions provided by the system. At T2, subjects brought up a few new functions, such as spelling check; however, they still focused on the search functions. At T3, the two groups continued to contribute nearly half of the concepts associated with interface component to the function aspect. The complex task group was able to come up with a new function, sort by.

The result aspect did not appear in subjects' initial model (T0) of information-rich web spaces. At T1, subjects listed 2 concepts (2.0% of the interface component) to represent search results in MedlinePlus. The limited attention that the result aspect received at this time could be because subjects were not assigned specific tasks to perform. In the 5-minute exploration, they might have browsed around the site rather than performing specific searches and examining search results.

At T2, after performing a set of tasks, subjects in both groups increased their representations of the results in MedlinePlus. The increase is natural because in the process of performing tasks, subjects were more likely to inspect search results returned by the search engine. When paying more attention to results, subjects also represented results from different aspects. They thought about the results in general, about ranking and organization of the results, and about the sources of the results. At T3, both groups maintained a similar level of attention to the result aspect and their representations of the results were similar to what they were at T2. This indicates that users form an understanding of the results of an IR system after using the system for completing just a few tasks and the understanding might not change much with more experience with the system.

6.7.1.5 Summary of the construction of the interface component of the structure dimension

Based on their initial models of the interface of information-rich web spaces, subjects constructed their mental models of the interface of MedlinePlus. They represented the interface component in relation to three aspects: elements, functions, and results, with the majority of the attention being paid to the first two aspects.

In representing the interface elements, at T1, subjects tended to list general elements that appeared on the surface of the website, such as “Contact Us” and “Disclaimer”. At T2 and T3, the representations of the interface elements were reduced; but more elements were from deep level webpages in the system. Subjects also began to represent typical pages in MedlinePlus, indicating that they developed a sense of how information in MedlinePlus is presented. Such development helps subjects recognize the information that they need when performing tasks.

In representing the functions at T1, subjects focused on search functions. Some of them brought up functions not available in MedlinePlus. At T2 and T3, both groups continued to focus on the search functions in the system. Meanwhile, they were able to bring up new functions, such as spell checking and sorting.

Subjects paid little attention to the display of results in MedlinePlus at T1. At T2 and T3, they gave more emphasis to this aspect and represented the displays of results from different perspectives, including how the results were ranked, how they were organized, and what sources the results came from.

6.7.2 The construction of the interface component of the evaluation/emotion dimension

When talking about different aspects of the interface of MedlinePlus, specifically the interface elements, functions, and results, subjects simultaneously expressed their evaluations of and emotions about the interface. This section reports how subjects evaluated the interface over time.

6.7.2.1 Evaluations of and emotions about the interface component at T1

At T1, 52 concepts (25.6%) of the evaluation/emotion dimension of subjects' mental models of MedlinePlus were about interface. Table 41 shows concepts that subjects used to evaluate the interface of MedlinePlus in terms of three aspects: look and feel, navigation, and search. Numbers in the parentheses are, respectively, the number of concepts used to describe each aspect of the evaluations of and emotions about the interface of MedlinePlus and the percentage of the aspect in the interface component of the evaluation/emotion dimension. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 41. Subjects' evaluations of and emotions about the interface of MedlinePlus at T1

Look & Feel (35, 67.3%)	
- Clear layout	- Plain
- Clarity	- Old fashion, A little outdated
- Clean lines on website	- Not extremely visually appealing
- Simple	- Lack of pictures, Need some design improvement such as addition of images, Few pictures
- Welcoming	- Needs new website banner
- Pretty, Colorful	- Word-intensive, Text-heavy
- Visually pleasing	- Sometimes too much information on a single page, Many words on each page
- Catchy title	- Some pages overwhelming with choices of topics
- No advertising!	
- Interactive	
- No clutter	
- Easy on eyes	
Navigation (14, 26.9%)	
- Streamlined	- Decent flexibility
- Easy to navigate	- Like the links to other pages
- Ability to jump topics with ease	- The variety of links on each page is helpful.
- Easy to follow from link to link	- Helpful links
- Easy interface	
Search (3, 5.8%)	
- Lacking search functions	- No advanced search
- Limited searchability	

As shown in the table, subjects evaluated the interface from three aspects: look and feel, navigation, and search. Look and feel refers to users' perceptions about the visual appearance of the interface. At T1, subjects dedicated the majority (35 concepts, 67.3%) of the evaluations and emotions concerning the interface of MedlinePlus to its look and feel. They expressed mixed feelings about this aspect. Some subjects thought that the interface was simple, clean, easy on the eyes, interactive, and visually pleasing; while some subjects said that the interface was plain, a little outdated, cluttered, and lacked pictures.

During their brief interactions with MedlinePlus, subjects also paid attention to the navigation in the system: 14 concepts (26.9%) of the evaluations of the interface were about navigation. Overall, subjects expressed positive feelings about the navigation in

MedlinePlus. They generally agreed that the interface was streamlined, easy to navigate, and had decent flexibility.

Subjects also evaluated the search function of MedlinePlus. It is worth noting that, in the initial encounter with the MedlinePlus system, they were critical of the search function. Subjects commented that the system had limited searchability and there was no advanced search.

6.7.2.2 Evaluations of and emotions about the interface component at T2

At T2, the simple task group dedicated 42 concepts (30.2%) of the evaluation/emotion dimension to evaluate the interface of MedlinePlus, and the complex task group used 30 concepts (23.8%). Table 42 lists concepts that subjects in the simple and complex task groups used to evaluate the interface of MedlinePlus. Numbers in the parentheses are, respectively, the number of concepts used to describe each aspect of the evaluations of and emotions about the interface of MedlinePlus and the percentage of the aspect in the interface component of the evaluation/emotion dimension. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 42. Subjects' evaluations of and emotions about the interface of MedlinePlus at T2

Simple task group	Complex task group
Look & Feel (14, 33.3%) - Clear - Clearly lists available options on the website - Decent page size - Simple - Liked subheadings for articles - Numerous links - Needs more pictures - Wish the pages were laid out more attractively (visuals, pictures) - Old - Text-heavy - Puts me to sleep - Not futuristic - Not perfect - Colors get repetitive	**Look & Feel** (6, 20.0%) - Like subheadings - Like subtopics - Layout pretty simple - Lack of diagrams - Lack of pictures - Not visually appealing - Needs more pictures/diagrams
Navigation (9, 21.4%) - Easy to navigate, Ease of navigation - Easy transition from one item to the next - Home is easy to get to - Navigateable - Streamlined - Easy to follow links - Drug summaries easy to navigate and especially good	**Navigation** (5, 16.7%) - Easy to navigate - Layout logical - Helpful links from main site - Hard to navigate to find specific info - Easy to find components of a system, but harder to find functions of a system
Search (13, 31.0%) - Easy to search - Effective search engine - Fairly accurate searches - Search bar very useful - Search works well - Nice search engine - Liked guided searches over browsing before - Actually narrower searches--ones that shows specifically what you are looking for instead of being up unrelated topics - Split for easy searching - Specific search difficult - Needs broader searches	**Search** (7, 23.3%) - Easy to search; Search easy to use - Effective search engine - Search effective in finding outside info - No higher level search functions - Basic search functions
Browsing (3, 7.1%) - Browsing is time consuming - Browsing less useful - Liked guided searches over browsing before	**Browsing** (1, 3.3%) - Easy browsing
Results (3, 7.1%) - Only provided hits on reliable sites	**Results** (11, 36.7%) - Lots of responses to search

- Site hits were major organizations - Curious about page ranking formula	- Very relevant articles - Top-Related - Organized search results - Specific results - Relatedness - Limited results - Too many results - Search results broken down into categories on left not helpful - Sometimes conflicting results - The search function tends to bring up several links to the same article (on different sites).

As shown in the table, at T2, in addition to look and feel, navigation, and search, subjects in both groups also evaluated two additional aspects of the interface: browsing and the results returned by the search function. At T2, both groups reduced their representations of the look and feel of the interface and navigation, while giving more weight to search, browsing, and results.

Subjects largely inherited the opinions that they developed at T1 about the look and feel of the interface of MedlinePlus. They still showed mixed feelings about the interface, with negative feelings outnumbering positive ones. One subset of subjects found the interface to be clear, simple, and decent, while another subset suggested including more diagrams, pictures, enhancement of the visual cues, and a more futuristic design.

Both the simple and complex task groups slightly reduced their representations of the navigation aspect of the interface. Examination of the specific evaluations revealed that the two groups held different opinions of the navigation in MedlinePlus. Consistent with T1, subjects in the simple task group agreed that the interface was streamlined and easy to navigate. However, subjects in the complex task group became more critical. One subject pointed out that “it was easy to find components of a system, but harder to find functions of a system”. Another commented that it was “hard to navigate to find specific info”.

Subjects by and large showed positive feelings about the search function in MedlinePlus, unlike the views that they expressed at T1. At T2, they thought that the search was easy to use, effective, and accurate. However, subjects also reported difficulties with the search function, especially with its limited ability to define the specificity of a search. This concern is illustrated well by two quotes from subjects in the simple task group: “specific search difficult” and “needs broader searches”. One subject’s comment about searching in MedlinePlus in the semi-structured interviews at T2 echoed the concern:

> I am not sure how to specifically use it. Sometimes, it is a bit too specific; sometimes, it is a bit too general. What's in between seems to be hard to find.

At T2, subjects in both groups made a few evaluative comments about browsing in MedlinePlus. Subjects in the simple task group expressed negative feelings about using browsing to look for information in the system. A couple of them reported that browsing was “time consuming” and “less useful”. However, one subject in the complex task group found browsing easy.

Subjects also evaluated results at T2, with the complex task group giving more attention to this aspect than the simple task group (11 concepts, 36.7% vs. 3 concepts, 7.1%). Inspection of the specific evaluations suggests that the simple task group did not have strong feelings about the results. Subjects in the group observed that the results were from major and reliable organizations. One subject in the simple task group also expressed curiosity about the ranking mechanisms. Subjects in the complex task group expressed positive feelings about results, but also made some critical comments. One subject in the complex task group reported conflicting results and another reported repetitive hits. Subjects in this group also commented that there were too many results and organizing results into categories was not helpful.

6.7.2.3 Evaluations of and emotions about the interface component at T3

At T3, the simple task group generated 34 concepts (22.8% of the evaluation/emotion dimension) regarding the evaluation of the interface, and the complex task group generated 25 concepts (20.7%). Table 43 lists concepts that subjects in the simple and complex task groups used to evaluate the interface of MedlinePlus. Numbers in the parentheses are, respectively, the number of concepts used to describe each aspect of the evaluations of and emotions about the interface of MedlinePlus and the percentage of the aspect in the interface component of the evaluation/emotion dimension. When similar expressions were used to describe one concept, only one expression is listed in the table.

Table 43. Subjects' evaluations of and emotions about the interface of MedlinePlus at T3

Simple task group	Complex task group
Look & Feel (14, 41.2%) - Simple; Simple format - Uncomplicated user interface - Colorful - Technologically friendly interface - Good layouts - Pleasurable - Interactive - Boring - Plain - Limited interactivity - Uncolorful	**Look & Feel** (9, 36.0%) - Pretty colors - Simple interface - Simplified - Needs more diagrams - Less links - Too many links per page - Too much info per page - Pages often too crowded with info - Too many links on each topic page
Navigation (10, 29.4%) - Streamlined - Easy to navigate - Logical - Interconnected - Easy maneuvering - Ease of transition - Easy to follow train-of-thought - Section headings useful - Intuitive	**Navigation** (4, 16.0%) - Easy to use interface - Health issues heading most effective - Everything flows and connects with each other - Integrated links
Search (7, 20.6%) - Useful search - Good keyword searches - Search is most important - Navigatable search bar - Like the search function - Search engine not as effective - Searches brought back really broad results that sometimes had no relation it seemed to my search question	**Search** (5, 20.0%) - Search categories helpful - Search engine most useful tool - Easy to search - Will be nice if you can search the topic or encyclopedia through the search engine - Search engine is plain
	Browsing (1, 4.0%) - Browsing sometimes frustrating
Results (2, 5.9%) - Search choices come up in a logical order - Broad results	**Results** (4, 16.0%) - Right amount of links in search results - Limited results - Does not rank relevance - Random articles
Others (1, 2.9%) - Widget less	**Others** (2, 8.0%) - Other features interesting, but not as relevant for consistent users

As shown in the table, at T3, in addition to the five aspects on which subjects evaluated the interface of MedlinePlus at T2, subjects in both groups started evaluating other aspects of interface in a general way, which were categorized as "others" in the table. One subject in the simple task group commented that the system did not provide many widgets. One subject in the complex task group commented that "other features [were] interesting, but not as relevant for consistent users".

At T3, in evaluating the look and feel of the interface of MedlinePlus, subjects still had mixed feelings. On one hand, they thought that the interface was simple, interactive, and pleasurable. On the other hand, they thought that it was plain and boring. In contrast to T2, when subjects expressed a strong desire for more pictures and diagrams, at T3, subjects called for reducing the amount of information on the web pages.

At T3, both groups generated about the same number of concepts evaluating the navigation aspect of the interface as they did at T2. Similar to T2, subjects in the simple task group agreed that the interface was intuitive, interconnected, and easy to navigate. It was easy for them to "follow train-of-thought." Different from T2, the complex task group expressed only positive evaluations at T3. For example, one subject commented that "everything flows and connects with each other." Another subject in the complex task group echoed this in the semi-structured interview at T3, stating that:

> It is easy to navigate. If you know what you are looking for, it is easy to find.

Both groups slightly reduced their evaluations of search at T3. Subjects in both groups recognized the usefulness of the search function provided by MedlinePlus. However, both groups also questioned the effectiveness of the search engine. One subject in the simple task group commented that “searches brought back really broad results that sometimes had no relation it seemed to my search question.” One subject in the complex group commented that it “will be nice if you can search the topic or encyclopedia through the search engine.” It is worth pointing out that the search engine in MedlinePlus does search topics and the encyclopedia concurrently, and returns relevant articles from both sections. This observation suggests that mental models sometimes contain misconceptions, which to a certain degree makes mental models a good tool for usability testing. In this case, the design could be improved by making it explicit for end users what sources the search engine of MedlinePlus searches against.

At T3, subjects paid little attention to browsing. Only one subject in the complex task group commented on browsing: “browsing sometimes frustrating.”

At T3, both groups also reduced their evaluations of results. Both groups expressed mixed feelings about the results. Subjects in the simple task group thought that the results were organized in a logical order, but sometimes the results could be too broad. One subject in the complex task group felt that there were the “right amount of links in search results,” but more commented that the results were limited, random, and not ranked by relevance.

6.7.2.4 Discussion of the construction of the interface component of the evaluation/emotion dimension

Table 44 shows the distribution of subjects' evaluations of and emotions about the interface of MedlinePlus at the three data collection points for both groups. The numbers in the columns are the number of concepts that subjects used to describe each specific aspect of the interface component and the percentage of that particular aspect in the interface component of the evaluation/emotion dimension of subjects' mental models of MedlinePlus. Though it is included in the discussion, T0 is not included in the table since it was based on subjects' written answers to a question in the demographic questionnaire rather than concept listing protocols.

Table 44. Subjects' evaluations of and emotions about the interface component at T1, T2, and T3

	T1	T2		T3	
		Simple	Complex	Simple	Complex
Look & feel	35 (67.3%)	14 (33.3%)	6 (20.0%)	14 (41.2%)	9 (36.0%)
Navigation	14 (26.9%)	9 (21.4%)	5 (16.7%)	10 (29.4%)	4 (16.0%)
Search	3 (5.8%)	13 (31.0%)	7 (23.3%)	7 (20.6%)	5 (20.0%)
Browsing		3 (7.1%)	1 (3.3%)		1 (4.0%)
Results		3 (7.1%)	11 (36.7%)	2 (5.9%)	4 (16.0%)
Others				1 (2.9%)	2 (8.0%)

In their initial model (T0), subjects evaluated the interface of information-rich web spaces in relation to two aspects: look and feel and navigation. As their experience with MedlinePlus increased, subjects were evaluating more and more aspects of its interface. At T1, after using the system for 5 minutes, subjects expanded their initial evaluations of the interface to include the search function. At T2, after subjects performed a set of tasks using the system, their evaluations of and emotions about the interface incorporated two additional aspects: browsing and search results. At T3, after performing a second set of tasks, subjects started evaluating some other general features (represented by the terms "widgets" and "other features") of the interface of MedlinePlus.

In the initial models, subjects evaluated the look and feel of the interface of information-rich web spaces as cluttered and distracting. At T1, after a brief interaction with the platform for the study, MedlinePlus, subjects dedicated the majority of their attention to evaluating the look and feel of its interface (35 concepts, 67.3% of the interface component in the evaluation/emotion dimension). Mixed feelings were expressed about the interface of MedlinePlus. Some subjects thought the interface simple, clean, and welcoming, but others thought the interface was plain, outdated, and cluttered. At T2, both groups reduced their evaluations of the look and feel of the interface. But they largely inherited the opinions that they developed at T1 about the look and feel of the interface of MedlinePlus, demonstrating mixed feelings about the interface. Several subjects expressed a desire for more pictures and diagrams.

At T3, both groups gave a little more emphasis to this aspect of the interface, but their opinions about the look and feel of the interface remained the same as at T2. Subjects called for reducing the amount of information on the web pages in the site. The development of subjects' evaluations of the look and feel of the MedlinePlus interface suggests that, in interacting with information-rich web spaces, users form their feelings/opinions about the look and feel of the interface at a very early stage of the interaction. Such feelings were comparatively stable over the two search sessions. In the early stage of the interaction, users tend to pay more attention to the presence of pictures, graphics, and diagrams; as their experience with using the system increases, users start to be concerned about the density of information on web pages.

In representing the navigation aspect of the interface, at T1, subjects had a good feeling about the navigation in MedlinePlus, feeling that its interface was streamlined and easy to navigate. At T2 and T3, subjects in the simple task group continued their positive feelings toward the navigation. Different from the simple task group, the complex task group's evaluations and emotions about navigation in MedlinePlus changed in response to the tasks that they performed. At T2, after performing a set of complex tasks, the complex task group started feeling that the navigation was sometimes hard. The negative feelings might be due to the difficulties associated with finding answers for the complex tasks. At T3, after performing the second set of tasks, the complex task group's feelings about navigation became more positive, which could be because of the easiness of finding answers for the simple tasks in the second task set.

In representing search, subjects in both groups increased their representations at T2 and reduced them at T3. At T1, when subjects just encountered MedlinePlus, they focused on pointing out the limitations of the search function, such as a lack of advanced search. However, at T2 and T3, they had a more balanced view of the search capabilities of MedlinePlus. They expressed both pros (e.g., easy to use, effective, and accurate) and cons (e.g., difficult, not as effective, and broad) of the search function. At both T2 and T3, subjects showed conflicting feelings about the effectiveness of the search function in MedlinePlus in dealing with the specificity of searches. Some thought that specific search was difficult in the system, while some thought that general search was difficult. It is worth pointing out that the difficulty in defining the specificity of a search is not a problem unique to the search function in MedlinePlus. Allowing users to define the level of specificity in their searches is a research and design issue that challenges information search behavior researchers, user interface designers, and system designers alike.

At T2, after performing a set of tasks, subjects started representing browsing, but these representations faded away at T3. The hierarchical and hyperlink-based organization of medical information in MedlinePlus is intended to support easy browsing. Ironically, subjects generally felt browsing in MedlinePlus was time consuming, frustrating, and less useful than searching. This result suggests that browsing might not be a preferable means for end users to access information in IR systems. Even in MedlinePlus, a system hand-crafted by a group of librarians and intentionally designed for accessing information by browsing, subjects showed overall negative feelings about browsing and expressed a preference for search functions (as one subject in the simple task group commented that he/she liked guided search better than browsing).

Also at T2, subjects started representing results returned by searches in MedlinePlus, and slightly reduced the representations at T3. At both times, the complex task group gave more emphasis to this aspect of the interface than the simple task group. The complex task group was also more critical of the results, commenting that the results were limited, random, and not relevant. The differences between the groups in the magnitude and tone of the evaluations of and emotions about the results could be due to the nature of the tasks: the difficulties associated with finding answers for the complex tasks could force subjects in the complex task group to spent more effort fumbling through the results to find answers, therefore reflecting more on the results. A comment made by a subject in the complex task group in the semi-structured interview reinforces this speculation:

> I got a lot of results for all of my searches, almost too many. Some are off topic.

6.7.2.5 Summary of the construction of the interface component of the evaluation/emotion dimension

Subjects evaluated the interface of MedlinePlus in relation to six aspects: look and feel, navigation, search, results, browsing, and other elements or functions on the interface. At T1, after a brief interaction with MedlinePlus, subjects were able to establish a balanced opinion about the look and feel of the interface of MedlinePlus, pointing out that the interface was simple and clear, but boring, needing more graphics and less text. The opinions remained consistent at T2 and T3, after subjects had more experience with the system.

Subjects' evaluations and emotions about the navigation and results were affected by the efforts and difficulties associated with finding answers for their assigned tasks. For navigation, the simple task group consistently showed positive feelings across time (e.g., streamlined and easy to navigate), while the complex task group thought the navigation difficult at T2, after performing a set of complex tasks, and regained positive feelings about it after performing a set of simple tasks. For the results aspect, the complex task group gave more emphasis to it and was more critical (e.g., limited, random, and lacks relevance) than the simple task group.

When subjects first encountered MedlinePlus, they were critical about the search function, pointing out that the system lacked searchability. After using the system to solve a set of tasks, subjects developed a balanced opinion about search, recognizing both pros and cons of this function. A common concern of subjects in both groups was the difficulty associated with defining the specificity of the search. This difficulty is not unique to MedlinePlus. Research is needed to explore mechanisms for users to define the specificity of their searches in the context of a particular IR system. Regardless of the limitations of the search function, it was the primary means by which subjects accessed information in MedlinePlus. MedlinePlus was intentionally designed to support users in finding information through browsing, but subjects thought browsing was frustrating and less helpful than the search overall.

6.8 Two additional components of the evaluation/emotion dimension of mental models: heuristics and self-reflection on the experience

The previous four sections (6.4-6.7) described and discussed the changes experienced by the four components, system, content, information organization, and interface, in the two dimensions (structure and evaluation and emotion) of subjects' mental models of MedlinePlus. Recall that at T2, after performing a set of tasks using the system, subjects began to develop heuristics for using the system and to reflect on their experiences of using the system. This section describes and discusses the changes over time in these two additional components in the evaluation/emotion dimension.

6.8.1 Heuristics

Heuristics are subjects' perceptions about how to use the system. They are rules of thumb about what is good to do and what is not, what is easy and what is difficult to do within a system. After the first 5-minute interaction with MedlinePlus (T1), subjects did not report any heuristics for using the system. One possible reason is that a 5-minute exposure to the system was too short for subjects to get to know how the system works. The other possible reason is that subjects were not assigned specific tasks to perform, and so were not focused on the procedural aspects of system use.

At T2 and T3, after using the system to solve some assigned tasks, subjects formed heuristics for using MedlinePlus. The heuristics might have an impact on the ways that subjects interact with the system in the future. This section reports heuristics developed by subjects at T2, after they performed the first set of assigned tasks (simple tasks for the simple task group and complex tasks for the complex task group), and T3, after they performed the second set of assigned tasks (a combination of simple and complex tasks for both groups).

6.8.1.1 Heuristics at T2

Table 45 lists the heuristics reported by subjects at T2. The simple task group generated 7 heuristics, 5.0% of all the concepts generated for the evaluation/emotion dimension at T2. The complex task group generated 5 heuristics, 4.0% of all the concepts generated for the evaluation/emotion dimension at T2. It is apparent that, at T2, heuristics were only a small portion of subjects' mental models.

Table 45. Heuristics for using MedlinePlus at T2

Simple task group	Complex task group
Looking for information	**Looking for information**
General approaches	*Easy for …*
- Medline is good for guided searches	- Easy to find information on specific topics
- Medline is not good for browsing	
Easy for …	*Difficult for …*
- Search engine was good if what you were looking for was specific	- Very specific info not that accessible
- Easy for general search	- Hard to find specific answers
	- Hard to navigate to find specific info
Difficult for …	**Processing information**
- Specific search difficult	*Hard to compare*
	- Harder to relate topics to one another
Use of information types	
- Hard to find answers outside of FAQ sites	
Accessing information	
- Everything you need basically if you are willing to read multiple articles sometimes	

As shown in the table, at T2, after performing the first set of tasks, subjects developed heuristics concerning a series of decisions in seeking information in an IR system, including looking for information (what approaches to take, what it is easy to search for, and what it is difficult to search for), use of information types, accessing information, and processing information.

Subjects in the simple task group formed heuristics about the effectiveness of two general approaches to look for information, search and browse, in MedlinePlus. One subject felt that MedlinePlus was “not good for browsing”; another concluded that MedlinePlus was “good for guided search”. Subjects in the complex task group did not form such heuristics.

Search was the most prominent means to look for information in MedlinePlus, as illustrated in subjects’ representations of the search function in the interface component and their expected behavior for solving the hypothetical tasks. At T2, subjects developed heuristics about the effectiveness of the search engine in MedlinePlus in finding information with different levels of specificity. Conflicting views were expressed by both groups. In the simple task group, one subject commented that it was “easy for general search”, and another commented that the “search engine was good if what you were looking for was specific”. In the complex task group, one subject commented that it was “easy to find information on specific topics”, while another felt the opposite, that “very specific information [was] not that accessible.”

As shown in the table, a subject in the simple task group developed heuristics for the use of specific types of information in MedlinePlus, commenting that it was “hard to find answers outside of FAQ sites”. Another subject in the simple task group formed a heuristic for accessing information in the system, commenting that he/she can get everything if he/she is “willing to read multiple articles sometimes”. Subjects in the complex task group did not form heuristics for these two aspects; instead, they formed heuristics for information processing. He/she commented that it was “harder to relate topics to one another”. The fact that the two task groups formed heuristics concerning different aspects of system use might be due to the cognitive demand imposed by the tasks: the simple tasks do not require subjects to relate or compare different medical conditions or diseases (information processing), and the complex tasks were designed to encourage relating and comparing.

Based on the analysis, it is reasonable to speculate that subjects developed the heuristics for using MedlinePlus based on their experience of solving tasks using the system. The development of heuristics was affected by the nature of the tasks.

6.8.1.2 Heuristics at T3

At T3, after performing the second set of tasks, subjects in both groups generated more heuristics for using MedlinePlus. The simple task group listed 18 heuristics (12.1% of the evaluation/emotion dimension) and the complex task group listed 13 (10.7% of the evaluation/emotion dimension). Table 46 details the heuristics reported by both groups at T3.

Table 46. Heuristics for using MedlinePlus at T3

Simple task group	Complex task group
Looking for information *Easy for …* - Broad, more well known concepts easier - Great for simple things - If you need something easy, its easy to find - Simplest search is best - Easy for general info *Difficult for …* - Difficulty finding very specific subjects - Tough to find specific answers - Not good with VERY specific questions - Not good for finding information on medical myths - Not good for trivia *What works in the system* - Typing in a question does not work well **Accessing information** *General* - Looking at multiple pages - Multiple places to look - Specific details on linked websites - Processing information *Hard to compare* - Hard to compare two diseases - It is harder to compare diseases to one another - No pages comparing illnesses	**Looking for information** *Easy for …* - Good for simple searches - Good for single term searches - Good for straightforward searches *Difficult for…* - Specifics difficult to find - Not good for searches involving multiple steps - Not good for searches involving multiple terms - Search multiple words - Not good for terms involving complex ideas - For very specific or uncommon questions, it would probably be at least as useful to start with a large search engine like Google and screen your results from there. *What works in the system* - Must search for statements, not questions **Accessing information** - Click on multiple links to find desired results **Use of information types** - The basic articles that the site provides on medical conditions are good for finding answers to common and expected questions. - Encyclopedia articles good for general information - Searching by both encyclopedia and topics is useful

Heuristics that subjects developed at T3 still revolved around the series of decisions in using IR systems, including looking for information, use of particular information types, accessing information, and processing information. At this time, the heuristics were more comprehensive and more consistent than they were at T2.

Subjects in both groups developed similar heuristics for looking for information in

MedlinePlus. They generally agreed that it was easy to search for simple, broad, and general information in MedlinePlus, while difficult to find specific information and complex information that involves multiple concepts or ideas. At T3, subjects also developed heuristics about how to formulate searches in MedlinePlus. One subject in the simple task group said that “typing in a question does not work well”, and one subject in the complex task group commented that one “must search for statements, not questions”.

There were more heuristics related to accessing information at T3 than at T2. The subjects in both groups pointed out that it is necessary to look at multiple pages to find desired results. Furthermore, one subject in the simple task group developed the idea that “specific details [are] on linked websites”. Apparently, both heuristics would have an impact on subjects’ behavior when accessing information in MedlinePlus.

In contrast to T2, the complex task group formed several heuristics about using different types of information in MedlinePlus, while the simple task group did not report such heuristics. Subjects commented on the use of the encyclopedia, basic articles provided by MedlinePlus, and health topics. For example, one subject commented that “encyclopedia articles [are] good for general information”, and another said “the basic articles that the site provides on medical conditions are good for finding answers to common and expected questions”.

Concerning the processing of information, the subjects in the simple task group found that it was difficult to compare diseases to one another. At T3, the complex task group did not comment on this aspect (they had generated related heuristics at T2). The report of this heuristic by subjects in the simple task group at T3 may be because they performed two complex tasks in the second search session, which required them to compare and relate information from different places.

The analysis of the heuristics that subjects developed at T3 reinforced the fact that subjects developed heuristics from their experience of using the system to perform specific tasks. The nature of the tasks helped to shape which heuristics the subjects developed.

6.8.2 Self-reflection on their experience with MedlinePlus

In the concept listings, subjects also expressed their evaluations of and emotions about their experience of using MedlinePlus to perform the tasks. The self-reflections did not appear at T1; they began to appear at T2. This section reports subjects' self-reflections on their experience with the system.

6.8.2.1 Self-reflections on their experience with MedlinePlus at T2

Table 47 lists subjects' self-reflections on their experience with MedlinePlus at T2. Subjects in the simple task group listed 4 self-reflections (2.9% of the evaluation/emotion dimension). Subjects in the complex task group generated no self-reflections at T2.

Table 47. Subjects' self-reflections on their experience of using MedlinePlus at T2

Simple task group	Complex task group
- A little frustration - Generally not much effort - Some prior knowledge might be necessary for few tasks - More specific questions took most time	

As shown in the table, at T2, one subject in the simple task group expressed his/her frustrations with using the system, but another subject claimed that performing the tasks did not take much effort. Subjects also reflected on their experience with particular tasks. One commented that "some prior knowledge might be necessary for [a] few tasks", and another said that the "more specific questions took [the] most time."

6.8.2.2 Self-reflections on their experience with MedlinePlus at T3

At T3, the simple task group generated 13 self-reflections (8.7% of the evaluation/emotion dimension) and the complex task group generated 1 (0.8% of the evaluation/emotion dimension).Table 48 lists subjects' self-reflections on their experience with MedlinePlus at T3.

Table 48. Subjects' self-reflections on their experience of using MedlinePlus at T3

Simple task group	Complex task group
- Frustrating, Frustration - Not as satisfied this time - Not sure how to improve user search for complex issues - Want to use it more to make a better opinion - A kid's website seemed to be one of the most popular - Getting an error when I clicked a link - Homepage not used - Similar links every time - Right in front of you	- Becomes easier with more use

At T3, after performing the second set of search tasks (including both simple and complex tasks), subjects in the simple task group continued to express frustration and dissatisfaction with MedlinePlus. They also pondered how to improve search for complex issues due to the difficulties they experienced in finding answers for complex tasks.

Meanwhile, subjects reflected on their experience with using the system by recalling a set of memorable instances that happened during the interaction. For example, one subject noted “homepage not used”, one commented that “a kid’s website seems to be one of the most popular”, and another commented on “getting an error when I clicked a link”.

The complex task group contributed only one self-reflection on experience. In contrast to the simple task group, whose evaluations were largely negative, the subjects in the complex task group felt that it “becomes easier with more use”.

6.8.3 Summary

Subjects developed heuristics concerning a series of decisions and actions involved in seeking information using IR systems: looking for information (by browsing or by searching), accessing information, selecting and using particular information types, and processing information. The heuristics arose from subjects’ empirical experience with using the system to solve particular tasks and were affected by the nature of the tasks that subjects had performed. As subjects’ experience with the system increased, they developed more heuristics and the heuristics became more consistent across subjects.

Subjects also reflected on their overall experience with the system by directly expressing feelings (frustration, dissatisfaction, and ease of use) or recalling instances that appeared in their interaction with the system (e.g., “getting an error when I clicked a link”). Only a few self-reflections were generated overall and almost all of those were generated by the simple task group. Thus, no general statements can be made related to these findings.

Both heuristics and self-reflections on experience could affect subjects' expectations of certain functions in the system and plausibly their future behavior in using the system. Future studies could be performed to explore whether these two components of the evaluation/emotion dimension have an impact on users' actual behavior in using an IR system.

Chapter 7: Conclusions and implications

In the previous two chapters, the context for examining mental models construction (the characteristics of subjects, tasks, and subjects' perceptions of MedlinePlus) and the construction of and changes in subjects' mental models of MedlinePlus at different time points were reported and discussed. This chapter sets out to discuss the conclusions and implications of the study.

The first two sections of this chapter (sections 7.1 and 7.2) discuss, respectively, the findings concerning the two research themes of the study: (a) the construction and changes of users' mental models of an information-rich web space, and (b) the impact of assigned tasks on mental models construction. The third section (section 7.3) reviews the limitations of the study. The results of this study have important implications for two groups: HCI and IR researchers and HCI and IR system designers. Thus, this chapter concludes by discussing the implications of the research results for design (section 7.4) and for future research (section 7.5).

7.1 Conclusions: construction of and changes in mental models

This section reports conclusions concerning the first theme of this study, the construction and changes of subjects' mental models of information-rich web spaces during a search session, from two aspects: (1) the structure and elements of subjects' mental models of information-rich web spaces and (2) the cognitive, emotional, and behavioral development of mental models during subjects' interactions with MedlinePlus. At the end of the section, the characteristics of subjects' mental models of MedlinePlus and mental model construction will be reviewed. Connections of the results of this particular study with the existing research on mental models will be made and the contributions of this study to our understanding of the construct of mental models will be discussed.

7.1.1 Users' mental models of information-rich web spaces

7.1.1.1 The three dimensions of users' mental models

Subjects in the study were a group of homogenous undergraduate students. They were not heavy users of medical information. General web search engines, WebMD, and Wikipedia were the main online resources they had previously used to look for medical information. Before the study, they had never used the MedlinePlus system. This group of subjects had an initial model of information-rich web spaces. The initial model was general and basic; however, it established the primary three-dimension and four-component framework of subjects' mental models of MedlinePlus, a typical information-rich web space.

The three dimensions were structure, evaluation and emotion, and (expected) behavior.[1] The structure dimension encompassed subjects' representations of physical or abstract elements of the system. At the same time that they represented the elements of the system, subjects formed opinions or feelings about the represented elements. These opinions or feelings constituted the evaluation/emotion dimension of their mental models. The first two dimensions emerged from subjects' concept listings. When given a task, subjects formed strategies to solve the task using the system. The expected behavior dimension represented users' perceptions of procedures for performing a task using the system. The (expected) behavior dimension was informed by mental model theory, which suggests that people's mental models of a system encompass procedural knowledge of how to use the system to solve problems, and was constructed from an analysis of the strategies that subjects reported they would use to solve a hypothetical task.

The emergence of the two dimensions, structure and evaluation and emotions dimensions, from the concept listing data suggests that there are two parallel cognitive processes involved in constructing mental models of an information-rich web space: representing elements of the system and forming evaluations and emotions about certain elements. Users might tend to represent more elements than they evaluate. These two processes were also identified in studies of people assessing the credibility of websites. Fogg (2003) pointed out that two things happen when people assess credibility online: (1) the user notices something, and (2) the user makes a judgment about it. His findings

[1] In this study, subjects were not required to actually perform the hypothetical task; instead, they were asked only to describe their strategies to solve the task. Therefore, the dimension is named expected behavior, instead of behavior.

support the current finding that people integrate representations and evaluations and emotions in their mental models of an information-rich web space.

IR systems are designed for people to look for information to solve tasks. Solving a particular task using an IR system involves decisions about strategies to access information in the system (e.g., search or browse) and decisions about a series of actions to execute the selected strategy. The decisions about the strategies and actions are inevitably based on users' understanding of the structure of the system and their evaluations of and emotions about the system. For example, only when subjects represent the encyclopedia in the structure dimension of their mental models could they possibly use the source to find related information. Therefore, although it was informed by mental model theory, the (expected) behavior dimension is well supported by the empirical data. Subjects' (expected) behavior of using a system is affected by the first two dimensions and is an integral part of people's mental models of an IR system.

It is worth pointing out that the comparative weight of the first two dimensions was stable over time. About 70% of the concepts that subjects contributed at each time point were about the structure dimension (T1: 75.5%; T2: the simple task group: 68.2%; the complex task group: 66.7%; T3: the simple task group: 64.9%; the complex task group: 68.0%) and the rest of the concepts were about the evaluation/emotion dimension. However, the composition and the content of these two dimensions of mental models (at T1, T2, and T3) did change over time, as has been discussed in detail in Chapter 6. Also changed was the third dimension (expected behavior dimension) of the mental models, in which subjects significantly changed their planned strategies to solve a common task over time (at T1, T2, and T3, as measured by subjects' descriptions of their strategies to solve a hypothetical task rather than their actual behavior of performing the task). The relationships between the three dimensions, particularly the relationship between the previous two and the behavior dimension, need further investigation.

7.1.1.2 The four components of users' mental models

Subjects' representations of the elements in the system (the structure dimension) clustered into four components: system, content, information organization, and interface. The four components were represented from different aspects, as summarized in Table 49. Some aspects did not appear at all three time points (T1, T2, and T3), when subjects' mental models of MedlinePlus were measured.

Table 49. Subjects' representations of the components of the structure dimension

Components	Aspects of subjects' representations of the components	
System	• System structure • Audience • The usage of the site	• Agencies involved • System behavior • Similar sites
Content	• General information • Subject • Specific content	• Type • Format • Presentation
Information organization	• Information organization schema	
Interface	• Interface elements • Functions	• Results

Similarly, subjects' evaluations and emotions about MedlinePlus were clustered around the same four components, system, content, information organization, and interface. When subjects' had used MedlinePlus to solve a set of assigned tasks, they began to form heuristics of what works and what does not, what is easy and what is difficult to do in the system. In addition, they began to reflect on their experiences with the system. The aspects of users' evaluations and emotions that appeared in their interactions with MedlinePlus are summarized in Table 50.

Table 50. Subjects' representations of the components of the evaluation/emotion dimension

Components	Aspects/perspectives of subjects' representations of the components	
System	• Attributes • Usefulness	• Usability • Public awareness
Content	• Quantity • Quality • Utility • Specific content in MedlinePlus	• Attributes • Comprehensiveness • Objectivity • Currency • Depth of info. • Language • Presentation
Information organization	• Information organization	
Interface	• Look and feel • Navigation • Search	• Browsing • Results
Heuristics	• Look for information • Use of information types	• Access information • Process information
Self-reflections	• Self reflection of their experience with the system • Memorable instances that subjects met in using the system	

At each time (T1, T2, and T3), subjects listed about 21-24 concepts, on average, to represent their thoughts about the system, which indicates that subjects' mental models of MedlinePlus were simple and parsimonious. However, from the two tables, it is clear that subjects represented and evaluated many different aspects of the system. At a high level, they represented and evaluated the overall system, the content of the system, the information organization in the system, and the interface of the system. At a second level, they represented multiple aspects of each component and evaluated each component in relation to these different aspects.

The comparatively small number of concepts that subjects employed to represent their perceptions of MedlinePlus and the comprehensiveness of their collective mental models of the system indicate that there might be significant differences between individual subjects in their mental representations. Many factors affect the likelihood of an element being noticed, such as involvement, topic, task, experience of the user, and individual differences (Fogg, 2003). Further analysis of the individual subjects' mental models will be conducted in the future. Efforts will be made to classify users based on their mental models of the system.

The three-dimension and four-component framework was established in the subjects' initial models of general information-rich websites and inherited by their mental models formed after using MedlinePlus. Therefore, it is plausible that this overall framework of people's mental representation of an information-rich web space is independent of the search tasks assigned.

7.1.2 Construction and changes of mental models over time: cognitive, emotional, and behavioral development

Three developmental processes involved in subjects' construction of mental models over time were identified: cognitive development, evaluative and emotional development, and behavioral development. Each of these processes corresponds to one of the three dimensions included in subjects' mental models of information-rich web spaces.

7.1.2.1 Cognitive development

The cognitive development of subjects' mental models of MedlinePlus is mainly reflected by the changes of their representations of the system, that is, the structure dimension of their mental models.

In developing the structure dimension of their mental models of MedlinePlus, subjects constantly assimilated new information into their existing mental models. The assimilation process was reflected at two levels, the concept level and aspects of the components level. At the concept level, subjects (over time) incorporated new concepts that describe new features or characteristics of the system. For example, at T3, a subject listed "table of content type layout" as a way to organize information in MedlinePlus; this feature did not appear at T1 or T2. At the level of aspects of the components, subjects (over time) developed and incorporated new aspects to represent a component in subjects' mental models. For example, at T1, after subjects had used MedlinePlus for 5 minutes, they developed a new aspect, agencies that created or contributed to the system, to represent MedlinePlus as an integrated system. This aspect did not appear in subjects' initial model of information-rich web spaces.

A parallel process in developing the representations of the structure dimension is phasing information out of the current mental models. The phasing out process was also reflected at two levels: individual concepts, and aspects of the components. At the concept level, subjects phased out meaningful concepts from their mental models. For example, a "symptom finder" function showed up at T1, but disappeared at T2 and T3. At the level of aspects of the components, subjects removed some aspects that they used to represent a component in the previous model. For example, at T2, the simple task group left out the audience of the system, one aspect of the system component included in their mental model of MedlinePlus at T1. At T3, the complex task group left out another aspect of the system component, system behavior, from their representations.

The two processes, assimilating new objects or aspects of a component into the mental models and phasing out some existing objects or aspects of the models, reflected the dynamics involved in the construction of mental models. Meanwhile, the two processes, in parallel, kept the number of concepts that subjects' contributed to represent their mental models of MedlinePlus stable over time. If the number of concepts that subjects expressed is an indicator of the actual magnitude of their mental models, the magnitude of subjects' mental models of MedlinePlus was stable over time. Such stability ensures the runnability of mental models over time. Keeping the model runnable is critical for a knowledge structure that supports inference, reasoning, and learning.

The cognitive development of subjects' mental models of MedlinePlus is not only reflected in the occurrence and disappearance of certain concepts or aspects, but is also reflected in the change of the emphases that subjects put on the different dimensions and the components in each dimension. For example, at the dimension level, compared to T1, both groups reduced the number of concepts contributed to the structure dimension to represent elements in MedlinePlus at T2 and T3 (the reduction of the emphasis on the structure dimension was statistically significant at T2 for the complex task group). At the component level, at T2, subjects in the complex task group contributed significantly fewer concepts to represent the content component than at T1.

Most importantly, cognitive development is reflected in the development of subjects' understanding of MedlinePlus. Recall that, in the initial model, subjects' representations of information-rich web spaces were general and basic. When they encountered MedlinePlus and then gained more experience with it, their representations of certain aspects of MedlinePlus became more and more specific, illustrating their improved understanding of the system. For example, in the initial model, subjects' representations of interface elements focused on general elements, such as menus, tabs, and links. At T1, subjects represented interface elements that appeared on the homepage of MedlinePlus, such as "about us," contact, and disclaimer. At T2, subjects incorporated elements that were more specific to MedlinePlus, such as cross references and clickable images. At T1 and T2, subjects listed various forms of information organization in MedlinePlus, such as alphabetical listing, listing by topics, and listing by anatomy. At T3, subjects in both groups recognized that information in MedlinePlus was organized in a network manner and they could reach the same information through different routes.

7.1.2.2 Evaluative and emotional development

The evaluative and emotional development of subjects' mental models of MedlinePlus is mainly reflected by the changes of their evaluations of and emotions about MedlinePlus, that is, the evaluation/emotion dimension of their mental models. Similar to the cognitive development, subjects' evaluative and emotional development also involved assimilation and phasing out processes.

The assimilation process was reflected at three levels, from low to high: the concept level, aspects of the components level, and the component level. At the concept level, subjects (over time) incorporated new concepts that evaluated elements in the system. For example, at T2, subjects in both groups evaluated the search engine in MedlinePlus as an "effective search engine", a concept that did not appear at T1. At the level of aspects of the components, subjects (over time) developed and incorporated new aspects to evaluate a component in subjects' mental models. For example, at T1, subjects evaluated the interface of MedlinePlus from three aspects: look and feel, navigation, and search. At T2, they evaluated two additional aspects: browsing and the display of results. At T3, they began to evaluate other general features of the interface. At the component level, subjects developed and incorporated new components into the evaluation/emotion dimension of their mental models. This development is reflected by the observation that, at T2, subjects began to form heuristics concerning various actions (e.g, search and access information) involved in seeking information using MedlinePlus.

The phasing out process in evaluative and emotional development was reflected at two levels: individual concepts, and aspects of the components. At the concept level, subjects phased out meaningful concepts from the evaluation/emotion dimension of their mental models. For example, a subject questioned the credibility of information in MedlinePlus at T1, but this concept was absent at T2 and T3. At the level of aspects of the components, subjects removed some aspects from which they evaluated a component in the previous model. For example, at T3, the simple task group stopped evaluating the browsing aspect of the interface of MedlinePlus, which they had evaluated at T2.

The evaluative and emotional development of subjects' mental models of MedlinePlus is also reflected in the change of the emphasis that subjects put on the different dimensions and the components in the different dimensions. For example, at the dimension level, compared to T1, subjects in the simple task group contributed significantly more concepts to the evaluation/emotion dimension. At the component level, compared to T1, the complex task group significantly increased the number of concepts evaluating the content component at T2.

Another development of subjects' evaluations of and emotions about MedlinePlus or their experience with the system is reflected in the change of values of certain feelings. For example, one subject in the complex task group commented at T3 that it "becomes easier with more use". Another subject in the complex task group echoed the comment by stating in the semi-structured interview that:

> I felt more comfortable with the system, the more [I] used it, no matter what the questions are.

One subject in the simple task group commented in the semi-structured interview that:

> When I first started, I just click on things, I didn't know what bring it up, if I am going to get a list of things, are there going to have any links to them? But now, I am very certain if I bring an alphabetical list, I know some of them would be more, look at this part as opposed to clicking on the direct link.

Such emotional development is echoed by the results from the user experience questionnaires. Recall that subjects were asked to fill out a user experience questionnaire after the first search session (T2) and after the second search session (T3). As has been reported in Chapter 5, the analysis showed that the simple task group became significantly more satisfied with the content of MedlinePlus from T2 to T3 and the complex task group indicated stronger intentions to use the system in the future. Although not statistically significant, subjects in both groups reported increasing enjoyment of the site, satisfaction with the overall experience with the site, and pleasure when using the system.

7.1.2.3 Behavioral development

The subjects' behavioral development is reflected in the development of their expected behaviors when using MedlinePlus to solve a task, rather than their actual behaviors of performing the task using the system. The development of subjects' expected behavior in the process of constructing mental models was demonstrated by two observations: a) subjects' descriptions of the steps that they planned to take to solve a hypothetical task, and b) the development of heuristics for using the system.

As reported in the overview of the construction of the (expected) behavior dimension of mental models in section 6.3.3, subjects in both the simple and complex task groups became more likely to use the more sophisticated strategy B over time. In the simple task group, at T1, no one planned to use strategy B; at T2, 57.9% of the subjects (11 subjects) planned to use it; and the number increased to 68.4% (13 subjects) at T3. The complex task group showed a similar development pattern, from 5.3% (1 subject) of the subjects planning to use strategy B at T1, to 21.1% (4 subjects) at T2, and 26.3% (5 subjects) at T3.

As reported in section 6.8, at T2, after the first search session, subjects in both groups began to develop heuristics for using the system. Heuristics reflected subjects' perceptions about what to do and what not to do, what is easy and what is difficult in the MedlinePlus system. At T3, after the second search session, the number of heuristics increased for both groups. Here are several exemplar heuristics that subjects developed in interacting with MedlinePlus:

- "Typing in a question does not work well"
- "Search for statements, not questions"
- "It is harder to compare diseases to one another"

The heuristics are likely to have an impact on subjects' future behavior in MedlinePlus. At a minimum, the increased number of heuristics suggests that subjects might make more planned use of the system over time.

7.1.3 Characteristics of mental models and of model construction

Johnson-Laird (1981) pointed out that a mental model plays a direct representational or analogical role. In HCI and IR research, there are conflicting views on whether mental models are metaphorical. One group of researchers argues that metaphors or analogies function as tools to help the user to understand an unfamiliar domain (Carroll & Olson, 1983; Gentner & Gentner, 1983; Young, 1983). Another group argues that mental models are not inherently metaphorical in nature (e.g., Savage-Knepshield, 2001).

This study found that mental models overall are representational, delineating the elements and structure of the current system; however, they also encompass metaphors (or analogies). The function of metaphors could be different at different stages of interaction with the system. When subjects first encounter a system, they employ analogies (similar sites) to help them understand the subject area of the web space; in this study, subjects used WebMD and PubMed to help them understand the subject area of MedlinePlus. When they gain more experience with the system, they use analogies to help them understand the structure of the web space; in this study, subjects used Yahoo and Google to help them understand the structure of MedlinePlus, and cited WebMD and Google as alternative sources to which they would go for the tasks at hand.

Consistent with the findings from other studies (e.g., Mayer & Bayman, 1981; Norman, 1983), this study found that subjects' mental models of MedlinePlus contain misconceptions. For example, some subjects confused two information types in MedlinePlus, health topics and an encyclopedia, regarding them as one source of information. Some subjects pointed out that the search engine in MedlinePlus does not search the health topics and encyclopedia, but in fact, the search engine brings up articles from both sources.

As can be seen from the two tables (Table 49 and Table 50), subjects represented many aspects of MedlinePlus. However, their mental models of MedlinePlus were incomplete. For example, some subjects admitted in the interviews that they did not actually think about the search mechanisms available in MedlinePlus. Some subjects pointed out that some links led to pre-formulated Medline searches and some links led to clinical trials; however, none of them pointed out the mechanism, NLM's Medical Subject Headings (MeSH), used to link many of these services to specific health topics (Marill, Miller, & Kitendaugh, 2006). This observation confirmed the notion suggested by Mayer and Bayman (1981) and Rosson (1983) that even an extensive amount of experience does not necessarily lead the user to a complete or consistent mental model. There are some things about a system that most users never learn. Training is often needed to help people to identify certain parts – especially invisible parts – of a system.

Being incomplete might be an inherent nature of mental models. Researchers (e.g., Norman, 1983; Ehrlich, 1996) agree that being simple and incomplete enables a mental model to be useful in inference and reasoning, because dynamically running the models would not be cognitively demanding. In this study, it was observed that at each time, subjects employed a similar small number of concepts (on average, 21-24) to represent their mental models of MedlinePlus. During their interactions with the system, subjects assimilated new elements, objects, evaluations, or emotions into their models, but also phased out some previously represented elements, objects, evaluations, or emotions from the models.

Mental models are dynamic. In this study, the construction of mental models reflected the interaction between people's current knowledge, the system, tasks, system feedback, and time. Before they encountered MedlinePlus, subjects held an initial model of information-rich web spaces. This initial model provided a foundation for them to construct their mental models of MedlinePlus, an instance of information-rich web spaces. When they initially saw the MedlinePlus system, they represented physical or abstract elements in the system. The representations of some elements, such as the content of the system, were often affected by the tasks that they were performing. The tasks' effects on mental models were also reflected by the fact that subjects' descriptions of the steps that they planned to use to solve a hypothetical task were directly influenced by the actions that they took to perform the prior assigned tasks. A detailed discussion of tasks' impact on mental model construction will be provided in the next section, section 7.2.

System feedback is considered a very important source for constructing mental models (Muramatsu & Pratt, 2001). Savage-Knepshield (2001) pointed out that providing users with visible feedback in response to their actions is key for facilitating users' construction of appropriate mental models of a system. This study supports this assertion. In the study, subjects pointed out that it was not clear how the system selected resources. Providing visible clues for users to decide how the system selects sources would make the model construction process smoother. Also, as mentioned earlier, subjects thought the search engine did not search the health topics and the encyclopedia together. If clearer system feedback, such as an indication of the sources in the search results, was provided, this type of misconception could be minimized.

The cognitive and emotional development embedded in mental models construction is not only a process coordinated by internal (people's initial model) and external structures (system, system feedback, and tasks), but also a process distributed through time, in the sense that earlier mental models impact later ones. For example, subjects' initial models established the structure for their subsequent models of the MedlinePlus system. As time went on and their experience with the system increased, subjects' mental models of MedlinePlus became more critical and substantive; subjects were more and more likely to use more sophisticated system-specific strategy to access information in MedlinePlus (indicated by subjects' descriptions of the strategies that they were going to employ to solve a hypothetical task using MedlinePlus). The notion that mental models are constructed from distributed cognitive processes (internal, external, and distributed over time) supports Hollan, Hutchins, and Kirsh's (2000) proposal of distributed cognition.

7.2 Conclusions: the impact of task complexity on mental models construction

This section reports conclusions concerning the second research theme of the study, the impact of tasks on mental models construction. Two types of tasks were defined in the study, simple and complex tasks. Simple tasks are well defined questions. Answers to them are located on one page and easy to recognize. Complex tasks are open-ended questions.

Answers to these questions are located on multiple pages. High-level cognitive effort, such as knowledge synthesizing, is required to solve complex tasks. The complexity of each individual task in the study was judged by two medical information professionals and later confirmed by subjects' assessments of task difficulty and mental efforts required by the task.

In this study, the assigned tasks had distinct impacts on all three dimensions of subjects' mental models of MedlinePlus. In the structure dimension, at T2, after completing corresponding tasks, the complex task group contributed significantly fewer concepts to the structure dimension than at T1, while the simple task group remained consistent with T1. This difference might be related to the cognitive load associated with the two types of tasks, with the load associated with the complex tasks consuming the subjects' ability to represent elements in the system.

At T2, the complex task group contributed significantly fewer concepts to represent the content in MedlinePlus (the content component) at the specific level. An inspection of the listed concepts suggested that subjects' representations of the content at the specific level were closely related to the tasks assigned. The 12 simple tasks provided the simple task group with more scenarios and medical terms than the 3 complex tasks provided to the complex task group.

In the structure dimension, the two task groups also differed in their qualitative representations and understanding of some aspects of MedlinePlus. For example, when representing the structure aspect of the system component, subjects in the complex task group developed a more in-depth understanding, pointing out that, in addition to pulling from outside information, MedlinePlus has its own information. When representing the system behavior dimension of the system component, the complex task group focused more on specific instances, such as “some links did not work” and “the search function tends to bring up several links to the same article (on different sites),” while the simple task group was more focused on general behavior of the system, such as pop-up windows. When representing types of information included in MedlinePlus, in addition to the basic information types, such as dictionary, encyclopedia, and news, the simple task group also listed fact sheets, FAQs, reports, and what-to-do articles. This difference might be because these sources provide the types of information useful for answering the simple tasks. The complex task group, in addition to the basic information types, listed scholarly and academic articles, clinical trials, and tutorials, which could be due to the fact that these sources provide more in-depth information required to answer the complex tasks.

In the evaluation/emotion dimension, after completing corresponding tasks, the two groups differed in the number of concepts contributed to evaluating MedlinePlus as an integrated system, with the simple task group listing significantly more concepts than the complex task group.

The two groups also differed in their attitudes about information organization and results in MedlinePlus. When evaluating information organization, the simple task group spent less energy evaluating the system's information organization, evaluating it at a general level with unanimous agreement that the information was well organized. However, subjects in the complex task group evaluated information organization at multiple levels, from subcategories to listings to general information organization. They were also more critical of the information organization in MedlinePlus, pointing out limitations of the information organization and providing recommendations for improving it. When evaluating search results, subjects in the simple task group were very neutral about the results, they expressed curiosity about the ranking mechanism, and they recognized that the results were from major and reliable organizations. However, subjects in the complex task group were more critical. They pointed out conflicting and repetitive results. They thought that there were too many results and they questioned the usefulness of organizing results into categories.

In addition, in the evaluation/emotion dimension, the two groups developed different heuristics concerning processing and use of information in MedlinePlus. Subjects in the complex task group recognized that it was hard to relate topics to one another in the system, while none in the simple task group pointed this out. This difference is because the complex tasks encouraged relating, comparing, and synthesizing information from different places, while the simple tasks were mostly fact finding and subjects could find each complete answer at a single place.

The tasks' impact on the subjects' strategies (expected behaviors) for solving future tasks was reflected by their adoption of a more sophisticated and system-specific strategy to solve a hypothetical task. At T2, after completing corresponding tasks, more subjects in the simple task group planned to adopt the system-specific strategy to access information in MedlinePlus, while more subjects in the complex task group planned to adopt the general search strategy. The differences between the groups in the planned adoption of search strategies might be due to the fact that paths to the answers of the simple tasks helped reveal the overall information structure of MedlinePlus, therefore encouraging the adoption of the more sophisticated system-specific strategy. A comment from a subject in the simple task group supported this speculation:

> The tasks are an effective way of learning how to navigate the site. That helped a lot. If I hadn't had that purpose, I would have wandered a lot more.

7.3 Limitations of the study

As with any other research, this study has some limitations. First, the platform web space in the study was MedlinePlus, a general consumer health information website created by the NLM. The site is database driven, but the content is manually maintained and organized by a group of dedicated librarians specializing in medical information. Therefore, the results based on this site may not be generalizable to other information-rich websites (particularly in non-medical domains), and probably not to other kinds of web spaces. Meanwhile, MedlinePlus is a hyper-link based web space facilitated by a simple search function (no advanced search functions were provided in the system). The design of its interface encourages end users to follow links in the system to find information. Therefore, MedlinePlus is different from classic search-oriented IR systems, such as OPACs and experimental search systems, used in the past mental models studies (e.g., Borgman, 1986; Savage-Knepshield, 2001). The results from this study might also not be generalizable to other IR systems.

Second, the subjects in this study were a small and homogenous group of undergraduate students at a major research university. They had more computer experience and were more skilled with using the web to access information than the general population. They also demonstrated a higher spatial ability than subjects in many other studies (Ekstrom, et al., 1976). Meanwhile, most of them were less motivated to seek medical information because of their young age and comparatively good health status. Thus, they were different from typical users of MedlinePlus. The results from this study might not be generalizable to other people, particularly those with different motivations for health information seeking.

Third, the behavior dimension of subjects' mental models was not constructed from their real behavior of solving a hypothetical task; rather, it was constructed from subjects' descriptions of the strategies that they would plan to use to solve the hypothetical task (thus the dimension is named the expected behavior dimension in the study). This limitation restricts the potential of this study to investigate the relationship between cognition, emotion, and actual behavior in the context of IR.

It is worth noting that this study explored mental models construction during a short search period: subjects spent 5 minutes exploring the system on their own and spent about 32 minutes using the system to perform two sets of tasks. Investigating mental models construction in such a short period of time is justified by the fact that, in the web environment, if a person cannot engage with a system in the first 10-20 minutes, there is a good chance that he/she will give up the site and try somewhere else. Therefore, it is important for us to understand how subjects' mental models of a generic web-based IR system evolve during this short period of time. However, a trade-off for this research design is that the study could not detect how subjects' mental models of an information-rich web space developed over a longer period of time, during which a person could develop from a novice user to an expert user of the system. Probing mental models development over a longer period of time will require a longitudinal study design.

7.4 Implications for design

The concept of mental models was adopted by HCI and IR researchers two decades ago for its potential to inform the design of technologies (Brehmer, 1987). However, the impact of mental models research on system design and user training has been limited (Roger et al., 1994). Thus, one of the main motivations for this study was to explore mental models' potential as a tool to improve practices in HCI and IR. This section discusses the implications of this study for system design, system evaluation, user modeling, and user training.

7.4.1 Mental models as a tool for system design, system evaluation, and user modeling

In the current HCI literature, the implications of mental models research for system design are often proposed in the form of general design guidelines, or proposed to enhance single functions or elements of a system. For example, in exploring users' mental models of an experimental IR system, Savage-Knepshield (2001) suggested that providing users with visible feedback in response to their actions is the key to help users develop accurate mental models of a system. By investigating an IR system's demands on mental models, Katzeff (1990) pointed out that the system should improve the presentation order of the search results to improve the usability of the system.

However, mental models have rarely been used as a tool or a framework as, for example, task analysis was used to generate system requirements and direct a system design from an overarching perspective. This study's findings suggest that mental models are not only able to guide the design of specific functions, elements, or sections of a system, but are also able to serve as a framework to support designers' decisions on many aspects of the system.

As has been discussed, subjects' mental models of information-rich web spaces have a three-dimension and four-component framework. The three dimensions, structure, evaluation/emotion, and (expected) behavior, correspond to three aspects involved in people's information search process: cognition, emotion, and actions. Therefore, the framework is able to provide a holistic view of end users in the context of searching for information. The framework was established in subjects' initial model of information-rich web spaces and inherited in the subjects' subsequent mental models of a particular information-rich web space, MedlinePlus. The composition and content of the mental models were affected by tasks that subjects performed, but to a large extent, the structure of the framework remained stable over time. Within the dimensions and components, subjects represented the system from many different aspects. Therefore, the mental models are a rich representation of end users' perceptions of the system.

The integration of perceptions, cognitions, and emotions involved people's information search process, the comparatively stability of the framework (to a large extent independent of the tasks assigned), and the richness of the representations encompassed in the framework make mental models a potentially powerful cognitive tool/framework for assisting designers in the design and evaluation of information-rich web spaces. Table 51 illustrates the mental model framework that could be used to support system design and evaluation of information-rich web spaces.

Table 51. Mental model framework to be used as a tool for system design and evaluation

Components	Structure dimension	Evaluation/emotion dimension
System	• System structure • Audience • The usage of the site • Agencies involved • System behavior • Similar sites	• Attributes • Usefulness • Usability • Public awareness
Content	• General information • Subject • Specific content • Type • Format • Presentation	• Quantity • Quality • Utility • Specific content in MedlinePlus • Attributes - Comprehensiveness - Objectivity - Currency - Depth of info. - Language - Presentation
Information organization	• Information organization schema	• Information organization
Interface	• Interface elements • Functions • Results	• Look and feel • Navigation • Search • Browsing • Results

The use of mental models as a framework for systematic guidance of system design fits well with the user-oriented approach to information system design. In designing a new information-rich web space, at the stage of eliciting users' requirements, the structure dimension of mental models outlined in Table 51 could be used as a tool to elicit potential users' expectations of and requirements for the system. For example, concerning the whole system, the designers (UI designers and user experience specialists) could elicit users' expectations of the structure of the system, their perceived audience for the system, the purposes for which they would use the system, what other similar sites they think of, given the designer's descriptions of the new site, and whether they expect to see the site connect to other organizations or other sites. Concerning the content, the designer or the information architect of the site could ask the potential users what subject areas they expect to see on the website, what types of information they think should be provided, and what format they prefer for information presentation. Concerning the information organization in the system, the designer or the information architect could ask the potential users about their expectations for how information in the site should be organized. Concerning the interface, the designer could ask users what interface elements they expect to see, what functions they believe the system should have, and how the results should be presented.

Based on interviews with a group of software designers, Hammond et al. (1983) found that providing designers with a user model is a better way to improve interface design than context-free design guidelines or task-action analysis methods, such as GOMS. Requirements elicitation based on the various facets shown in Table 51 can be backed up by how users are actually going to perceive and represent the system. Users' answers to these questions could quickly help designers form an appropriate model of the potential end users' expectations for the system.

If the designer has a prototype system for users to review, a parallel set of questions derived from the structure dimension of Table 51 could be used to elicit users' comments about different aspects of the design, which could then be used to improve the usability of the system. For example, in the study, when representing the content of MedlinePlus, a subject suggested that "additional content other than medical data might make it more interesting, such as an article about health insurance". Another example is that, when representing the functions of MedlinePlus, a subject suggested a symptom finder, a function that is not provided by MedlinePlus.

Questions also could be formulated based on the evaluation/emotion dimension of subjects' mental models (the third column in Table 51). For example, concerning the content, the designer could ask users what they think about the quality, utility, comprehensiveness, or objectivity of the content, or how they think about a particular section of the content. Concerning the interface, the designer could ask how the users think about the look and feel of the interface and how they feel about the navigation, search, browsing, and results access provided by the interface. These evaluations could help the designer to re-think and re-design certain elements to more accurately shape future users' perceptions of the content of the system. For example, one subject in the study suggested that categorizing results into different categories does not really help.

It is clear that the framework derived from the two dimensions of mental models is useful at the stage of eliciting user requirements and the stage of evaluating system prototypes. Because the two dimensions describe how end users are going to perceive, represent, and evaluate the system, the questions generated based on them could provide comparatively structured data about users' thoughts and feelings about various aspects of the system.

It has been discussed that the construction of mental models is dynamic, reflecting the interaction of the user and the context in which he/she is situated. When users only had a brief interaction with the system, they were able to represent the system and evaluate it from different aspects, but these representations and evaluations were basic, general, and limited. When users had more interaction with the system, their representations of the system became more fine-grained, and their evaluations became more substantive and critical. The dynamics of users' representations and evaluations and their developmental nature suggest that an iterative approach needs to be considered when using the framework to support system design and evaluation.

In addition to providing a framework for designers to design and evaluate information-rich systems at the points when the system is designed or evaluated, mental models also have implications for developing user modeling mechanisms that are able to dynamically adapt to end users during their interaction with the system. Traditional information sources for user modeling include users' domain knowledge, interests, goals, beliefs, and preferences (Brusilovsky & Millan, 2007). Many current user modeling systems are behavior oriented (Kobsa, 2001). As illustrated in the framework, when users interact with an information system, they not only represent the objects in the system, but also evaluate and express emotions about the objects in the system. This suggests that user models in information retrieval systems or hypermedia-based systems need to take into consideration the affective states of end users toward the system during the interaction. Effective adaptations could be made to a user by detecting the correspondence of affective states and the objects and elements in the system or facts in the user-system interaction.

As has been reviewed in the previous section, the nature of tasks has an impact on people's representations and evaluations of many aspects of an information retrieval system, such as type of content, information organization, results, and heuristics of using the system. People's representations and evaluations of the system could provide informative suggestions about what users need at a certain point in their interaction with a system. Therefore, tasks could be a very good information source for user modeling in information systems, particularly information retrieval systems.

Compared to research on applying mental models to improve the usability of a system, the research on mental models as a source for user modeling is scarce. More research is needed to bridge these two areas. Two directions for future analysis of the data from this study could provide more insights to user modeling in IR systems: (1) the analysis of individual users' mental models of MedlinePlus, and (2) the analysis of users' behaviors while interacting with MedlinePlus, particularly the relationships between users' mental models and their behaviors.

7.4.2 Mental models as a tool for user training

It has been expected by researchers and designers that, if a system is well designed and conforms to users' expectations, users should be able to use the system without any formal training. This is true for applications that people use to achieve a simple goal, such as ATMs. But for most applications, especially systems encompassing a large amount of information and complex navigation and search functions, training is necessary for superior performance.

In the current study, subjects' mental models became more detailed as their experience with MedlinePlus increased. However, they only represented the visible parts of the website (such as interface elements and functions), or guesses about the system based on system feedback (for example, that the site connects to outside information, and that javascript is needed to view some pages), or their common knowledge about web technologies (databases). They did not have thoughts about how information was put together. They did not have representations about metadata, and did not know that MeSH connects other government medical information resources, such as PubMed and ClinicalTrial, to the health topics in MedlinePlus. Thus, self-guided learning of a system can help users reach only a certain level of understanding of the system; training is necessary to help users achieve a higher level of understanding of such systems.

Most existing research on mental models and training focuses on how to develop training materials to help users develop better mental models of a system, such as providing users with appropriate conceptual models (Borgman, 1983; Savage-Knepshield, 2001). Only a few studies explored how research on mental models can inform the design of training materials. For example, Henderson and Tallman (2006) used stimulated recall methods to examine teacher-librarians' mental models about pedagogy, practice, and self-reflections when teaching computer information literacy skills. They then used the results to help teachers learn more about their teaching and what their students were thinking so that they could individualize their teaching strategies and troubleshoot problems with student misunderstandings. Based on a two-stage mental model construction theory (stage 1, the learner builds components models for each part in the system; stage 2, the learner builds a causal model of the entire system), Mayer, Mathias, and Wetzell (2002) proposed that, in designing computer-based multimedia instructional materials for a physical system, presenting the learners with component models followed by a presentation of the causal relationships between the components could reduce the cognitive load of the learners.

By the same token, this study revealed subjects' mental models of MedlinePlus – how they represented the system and how they evaluated it. In this process, it also disclosed the limitations of subjects' understandings of the system. Instructors could design teaching strategies that address the weaknesses in student understandings. Compared to imposing a conceptual framework on users, this bottom-up approach is based on their current knowledge status and, therefore, might be more constructive.

Metaphors help users understand and learn to use a new system (Foss, Rosson, & Smith, 1982). This study found that subjects considered similar sites during their interactions with the system. As was introduced in the first section of this chapter, at the early stage of interaction, subjects tended to think of sites with similar subject matter. When they had some experience, they tended to think of sites with similar structure to the system of interest. This observation suggests that, when encountering a new IR system, users might try to understand the subject matter of the information in the system, then the structure of the system. Metaphors could be useful to help users make sense of the system. At the early stage, instructors could describe sites that share similar subject matter with the system and later describe sites with similar structure.

7.5 Implications for research

This section discusses the implications of this study for research in terms of: (1) the effectiveness of the concept listing method for eliciting mental models, and (2) directions for future studies.

7.5.1 An effective method for eliciting mental models: concept listing

In this study, three methods were used to elicit subjects' mental models of MedlinePlus: a concept listing protocol, semi-structured interviews, and a drawing protocol. In the results reported here, subjects' mental models were constructed based primarily on the data produced by the concept listing protocols. Traditionally, semi-structured interviews are the main method for eliciting mental models in IR research. Thus, in this study, transcripts of the interviews were analyzed to provide a comparison point with the data from the concept listing protocols, as well as to support the interpretation of some listed concepts. Results from the drawings will be reported elsewhere.

As has been reviewed in Chapter 2, people's verbal protocols, such as interviews and think-aloud protocols, are usually the main source for eliciting people's mental models.

Verbal protocols have some obvious limitations. For example, semi-structured interviews tend to impose a pre-existing structure on subjects. In a think-aloud protocol, subjects tend to rationalize their behavior, producing a biased account of subjects' cognitive structures and activities. One of the common limitations for both methods is that transcribing and analyzing verbal protocols is time consuming. How to effectively measure people's mental models of an IR system remains a challenge for IR researchers. This study suggests that concept listing could be an efficient and effective alternative to verbal protocols for eliciting people's mental representations of an IR system.

The effectiveness of the concept listing protocol is reflected in the nature of the method, the way that the method was implemented in this study, and the results produced in this study. During the concept listing protocol, subjects had full control of what concepts to contribute, in what order, and at what pace. There were no interruptions during the task. Thus, there was no pre-imposed structure on how subjects presented their thoughts. Mental models frameworks were allowed to emerge naturally from the data. The richness of subjects' representations of the MedlinePlus system reflected in the concept listing data indicates that a concept listing protocol has sufficient power to elicit complex mental activities. In addition, the effectiveness of the method was cross validated by a preliminary analysis of the interview transcripts, which produced a similar framework to the one that emerged from the concept listings.

The efficiency of the concept listing protocol is reflected in the implementation of the method. In the study, each concept listing session took subjects just five minutes to accomplish. During the session, it was observed that it was easy for subjects to learn to complete the protocol. Before the first concept listing protocol, subjects were given a short demonstration of how to perform the concept listing using an in-house developed computer program. No subjects reported difficulties with understanding how to perform the task or with using the concept listing program.

When using a concept listing protocol to elicit mental models, researchers are recommended to employ a strong prime. Pejtersen (1991) pointed out that a strong prime might be helpful in preventing subjects from generating random, common sense or stereotypical information that is not related to the object or system under study. In this study, before each concept listing task, subjects were given clear instructions concerning what is expected in the protocol. Subjects were explicitly told that the concepts could be, but were not limited to, the system's component parts, objects in the system, its working mechanisms, its functions, and related processes. To prevent subjects from thinking only about these aspects of the system, at the end of the instructions, they were reminded that they could list any concepts that they believed were important in representing their thoughts about the system. The results showed that the prime was successful. Only 4.1% of the listed concepts were too general to infer the subject's intended meaning.

It is worth noting that, when listing concepts, subjects did not necessarily list them in the form of a single word or a phrase. They also provided phrases and sentences that included multiple concepts. This practice should be allowed in a concept listing protocol. In future studies, instructions could be provided to subjects to indicate that they are allowed to input phrases or sentences if they feel they need to do so, as one subject commented in the exit interview:

> The first time, I just put the word down, then I […] started typing annotated sentences, instead of single words. I found that to be more indicative than what I was actually thinking or feeling about the website.

Concept listing is an effective and efficient method for eliciting mental models. However, in order to develop a better understanding of users' mental models of a system, concept listing protocols should be supplemented with semi-structured interviews. In concept listing, subjects rarely mentioned their understanding of the system's working mechanisms, such as how search engines work, and how the results were ranked. It is possible that users do not think about (or do not care about) these aspects of a system when they use the system. However, sometimes it is necessary for researchers or designers to learn about users' understanding of these underlying mechanisms to inform particular design decisions. For training purposes, augmentation of the concept listing protocol with semi-structured interviews may be even more useful.

7.5.2 Future research

This study suggests several directions for future studies. First, additional analysis could be performed to identify patterns of individual users' mental models and categorize users based on the characteristics of their mental models. Understanding different patterns of users' mental models would help to investigate whether different mental model groups show different behavior and performance. Furthermore, tests could be performed to investigate whether mental models mediate the relationships between individual differences, such as spatial ability, or environmental factors, such as tasks, and people's information searching behavior and performance. If the mediating effect of mental models is established, the characteristics of the mental model could be powerful indicators of why people show different behaviors and performance with information searching in a particular IR system.

As has been pointed out, in this study, expected behaviors (subjects' descriptions of steps that they planned to employ to solve a hypothetical task using MedlinePlus) were used as indicators of subjects' procedural knowledge of performing tasks using MedlinePlus. In future studies, it would be worthwhile to observe subjects' actual information searching behavior. Such observations could foster the examination of the relationships between the three dimensions, particularly how the structure and evaluation/emotion dimensions of mental models affect the subjects' behavior of solving particular tasks using a particular system.

It has been mentioned that subjects in the study were a group of homogenous undergraduate students with comparatively good health status, so the results of this study may not be generalizable to other populations. It would be interesting in future studies to explore how other more typical users of MedlinePlus, who often do not have extensive experience with computers and who badly need medical information, use the site and how they construct their mental models of the site.

Finally, future user studies or usability testing could be designed to test whether the framework (Table 51) based on the structure and evaluation/emotion dimensions of subjects' mental models could effectively help designers build a model of end users of the system and elicit the users' specifications of their requirements or their evaluations of an information-rich web space. Future studies also could be conducted to see whether the framework could be effectively applied to inform the design and evaluation of other types of web-based IR systems.

Appendix A: Demographic questionnaire

1. Your age: ______

2. Your sex:
   ___ Female
   ___ Male

3. Your current class status
   - ❒ Freshman
   - ❒ Sophomore
   - ❒ Junior
   - ❒ Senior

4. Your major ____
   Minor, if any _______________________________________
   Occupation, if any ___________________________________

5. About how many years have you been an internet user? ____________ years

6. How often do you look for medical information online (choose one to fill; if you)
   - ❒ _______times/day
   - ❒ _______times/week
   - ❒ _______times/month
   - ❒ _______times/year
   - ❒ Never

7. If you look for medical information, what sources do you use? Please check the sources that you use and list the frequency of your usage. Provide additional sources and the associated frequency of usage in the blank if they are not on the list.

❒ **General web search engines (e.g., Google, Yahoo, AOL, or ASK)**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Wikipedia**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **PubMed**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **MedlinePlus**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **WebMD**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **UNC's health websites**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Yahoo! Health website (health.yahoo.com)**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Family and friends**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Doctors**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Books (including books from libraries)**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Other: ______________________________________**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Other: ______________________________________**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Other: ______________________________________**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Other: ______________________________________**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Other: ____**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

❒ **Other: ____**

❒ Daily ❒ Weekly ❒ Monthly ❒ Yearly

---

8. This set of questions asks for your impressions and opinions about information rich websites.

Information rich websites are websites that contain a large amount of information (often thousands or millions of web pages) in a particular domain or across domains. The information on the sites could be in various formats, such as text, image, graphs, or videos. Information-rich websites are often designed for a diverse set of users. Users can access the information in the sites by searching or by browsing (following hyperlinks provided on the website). Examples of such information-rich websites include wikipedia, Library of Congress' American Memory project, WebMD, and the IBM company website.

a. Please describe characteristics of information-rich websites based on your own experience and understanding.
[You can type your answer on your computer, if you want]

b. Please describe your general approach to use this type of website.
[You can type your answer on your computer, if you want]

Appendix B: Mental models measurements

Part I: Concept Listing

In this task, you are asked to illustrate your mental image of the current system by listing meaningful concepts. The concepts could be, but are not limited to, the system's component parts, objects in the system, its working mechanisms, functions, and processes. Remember that you can list any concepts that you believe are important in representing your thoughts about the system.

Part II: Semi-structured interviews

a. What information is provided by MedlinePlus?
b. How do you think information in MedlinePlus is organized?
c. How do you think the system works?

Imagine that your friend asked you about the symptoms of anxiety and the treatment for it. You decide to investigate the question by using MedlinePlus.

What steps would you take in order to find information for your friend's request? Write down each of the steps that you would follow as if you were actually using the system to find related information.

<Note: Speak out: Recording>

Part III: Drawing

Please draw a diagram or a picture of your perceptions about MedlinePlus

[Note: enough blank space for drawing]

Please describe/explain the drawing.

Example for Concept Listing

This example intends to show you the way the Concept Listing works.

In this example, you are asked to list concepts concerning the field of Psychology to illustrate your understanding of the field. You will need to list the concepts one by one in the order that they come to your mind:

The answer could be:

Psychoanalysis
Experimental studies
Intelligence
Mental imagery
Instinct
Anxiety
Consciousness
Mind
Cognition
Thinking

Appendix C: Tasks

Session 1: The simple task group performs simple tasks and the complex task group performs complex tasks.

Simple task group: 12 simple tasks

1. One of your friends told you that she been exposed to hepatitis B because her husband is positive with hepatitis B. She was immunized some time ago with a hepatitis B vaccine. But now she is wondering how long a hepatitis B vaccine is usually good for.
2. Your friend is an athlete. Now he wants to increase muscle mass. He has been training without creatine, but would like to start a regimen. He is seeking your advice on this. You decide to find out first what the side effects of taking creatine are.
3. A heart attack is a medical emergency and prompt treatment increases the chance for survival. According to the American Heart Association, heart attacks cause 1 out of every 5 deaths. According to the National Institutes of Health (NIH) more than 1.2 million heart attacks occur each year in the United States and about 460,000 of these are fatal. Approximately 300,000 people die annually from heart attacks before they can receive medical treatment. To be prepared for possible emergencies, you decide to find out what to do when a person around you has a heart attack.
4. Amyotrophic Lateral Sclerosis (ALS), sometimes called Lou Gehrig's disease, is a rapidly progressive, invariably fatal neurological disease that attacks the nerve cells responsible for controlling voluntary muscles. ALS causes weakness with a wide range of disabilities. Symptoms may include twitching, cramping, or stiffness of muscles. One of your friends told you that recently, he is feeling a lot of muscular movements like tics that affects several muscular groups. He is concerned that it might relate to ALS. You decide to help him find out what tests are used to diagnose Amyotrophic Lateral Sclerosis.
5. Protein is a “building block” nutrient. Your body uses protein to build tissue, such as white and red blood cells, other cells in the immune system, skin, hair, and muscle. Given the importance of protein in your body, you want to find out how much protein an average person needs each day.
6. Imagine that one of your close family members recently had kidney failure. You are wondering whether it is an indication that you might be at a high risk of having kidney disease.
7. Years ago, if a woman was HIV+, family planning was the last thing on her mind; HIV and having a child just didn't mix. The fear of transmitting HIV to her unborn baby was too great for most women. But with the advent of HIV medications, you wonder whether it might be possible for an HIV positive woman to carry a baby. You decide to find out what are the possible ways to prevent an HIV positive mom from passing HIV to her unborn child?
8. Low blood pressure can be a boon when it results from a healthy lifestyle. But in some instances, low blood pressure can be a sign of serious, even life-threatening disorders. The reason for low blood pressure is not always clear, but following a healthy diet is always helpful in improving the

condition. Considering the potential risk of low blood pressure, you decide to find out what to eat when you have low blood pressure.

9. Your friend's doctor put him on Amoxicillin for an infection and he found that he is having diarrhea. You want to find out for him whether Amoxicillin can cause the side effect of diarrhea?

10. Imagine that one of your cousins, who is under 18 years old, is considering taking Vitamin A. You heard that taking too much Vitamin A could cause bone loss, blurred vision, hair loss, and damage to internal organs, particularly the liver. Considering the potential risks of overdosing, you decide to find out what is the recommended daily amount of Vitamin A for children under 18.

11. The American College of Sports Medicine (ACSM) defines aerobic exercise as "any activity that uses large muscle groups, can be maintained continuously, and is rhythmic in nature." ACSM also pointed out that it is a type of exercise that overloads the heart and lungs and causes them to work harder than while at rest. Imagine that you are interested in taking aerobic classes offered on campus, and you decide to use MedlinePlus to see what aerobic exercise does for your health?

12. HIV (Human Immunodeficiency Virus) is the virus responsible for the condition known as AIDS (Acquired Immunodeficiency Syndrome). You have a vague idea that HIV has to do with one's immune system. You decide to use MedlinePlus to find out generally how the HIV suppresses the immune system.

Complex task group: 3 complex tasks

1. Imagine that a friend of yours is studying the roles that insulin plays in the liver and the kidney. He particularly wants to know what is the primary function of the liver and the kidney. What are the roles of insulin in the liver and kidney respectively, and why would insulin be needed there? Is insulin related to liver and kidney diseases? You decide to use MedlinePlus to find information to help him answer these questions.

2. Over the last decade there has been increasing interest in the clinical association between hypertension and diabetes. You want to know what is the relation between diabetes, Type I diabetes and Type II diabetes respectively, and hypertension? And how do they affect each other. You decide to use MedlinePlus to find as much information as you can to make sense of these questions.

3. Since its prohibition in 1937, marijuana's use as a medicine became restricted. However, in recent years, some states (e.g., California) legalized the smoking of marijuana by certain patients. Thus medical marijuana has become a subject of contentious debate. You want to understand the arguments for and against the use of marijuana for medical purposes. Therefore, you decide to do some research on this subject using MedlinePlus.

Session 2: Two groups perform the same set of tasks

Both groups: 4 simple tasks and 2 complex tasks

Simple tasks

1. Healthcare these days, Computed Tomography scans, or CT scans as they are more commonly referred to, are being prescribed by doctors quite frequently. Imagine that one of your family members is scheduled for a CT scan next week, and she is anxious about what she needs to do. So you decide to find out for her how she can prepare for a CT scan.
2. Blood pressure is important in making sure that blood can move against gravity to the brain and also to maintain normal functioning of the kidney and other organs. One of your friends gets his blood pressure reading as 111/69. You want to find out approximately whether it falls in the normal range.
3. The Centers for Disease Control and Prevention has released recommendations for the use of the hepatitis A vaccine by travelers. People traveling to any country other than Australia, Canada, Japan, New Zealand and countries in Western Europe and Scandinavia should receive either the hepatitis A vaccine or immune globulin before departure. Imagine that you are travelling to a place not listed above and you decide to take the hepatitis A vaccine before your travel. So you decide to find out, for best protection, how long before traveling should you start taking the Hepatitis A vaccine series?
4. Back pain affects 80% of Americans at some time in their lives. Given the troubles that back pain can possibly cause, you plan to take actions to improve your back health. So you decide to find information about back health using MedlinePlus, in particular, what types of exercises help improve back health.

Complex tasks

1. Imagine that your friend recently was diagnosed with Asthma and was put on two inhalers. But he thinks it is chronic bronchitis. So he wants to know what the similarities and differences between asthma and chronic bronchitis are. Also, he wants to know various means to treat or soothe asthma, such as medicine, diet, alternative medicines, exercise, etc. You want to help him by finding as much information as possible in MedlinePlus.
2. There is an interesting observation that many ophthalmologists always wear glasses and do not use lasik surgery, soft contact lenses, or hard contact lenses. You are interested to know if there are any reasons for this phenomenon. You decide to use MedlinePlus to find related information.

Appendix D: After-each-task questionnaire

How difficult was this task?

- ❒ Very easy
- ❒ Easy
- ❒ Neutral
- ❒ Difficult
- ❒ Very difficult

How much mental effort did you use in doing this task?

- ❒ Very small amount
- ❒ Small amount
- ❒ Some
- ❒ Large amount
- ❒ Very large amount

How satisfied are you with your performance on this task?

- ❒ Very disappointed
- ❒ Disappointed
- ❒ Neutral
- ❒ Satisfied
- ❒ Very satisfied

Appendix E: User experience questionnaire

1. This MedlinePlus website is similar to other websites that I have used.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

2. Learning to operate the MedlinePlus website was easy for me.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

3. Using this website enables me to find information more quickly.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

4. I found it easy to get the website to do what I wanted it to do.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

5. Using this website improves my performance in finding information.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

6. My interaction with MedlinePlus was clear and understandable.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

7. I found the website to be flexible to interact with.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

8. MedlinePlus makes it easier to find information.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

9. The information on MedlinePlus is understandable for me to use.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

10. The website worked as I expected it would.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

11. MedlinePlus is having a simple layout for its content.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

12. It would be easy for me to become skillful at using this website.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

13. I think the website is engaging.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

14. Using this website enhances my effectiveness in finding information.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

15. I found MedlinePlus easy to use.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

16. The information on MedlinePlus is reliable for me to use.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

17. I understood the internal workings of the MedlinePlus website.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

18. I will recommend MedlinePlus to my friends if they are looking for medical information.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

19. I find this website useful for finding information.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

20. Using this website increases my productivity in finding information.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

21. How much about the overall MedlinePlus do you feel you learned through your searches?

❒ Very little ❒ A little ❒ Some ❒ A lot ❒ A great amount

22. The information on MedlinePlus is useful for me to use.

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

23. MedlinePlus is of a clear design

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

24. I enjoyed using the website

❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

25. I will use MedlinePlus for medical information again.

- ❒ Strongly disagree ❒ Disagree ❒ Neutral ❒ Agree ❒ Strongly agree

26. My overall experience of using MedlinePlus is

Satisfied	extremely	quite	slightly	neither	slightly	quite	extremely	Dissatisfied
Pleased	extremely	quite	slightly	neither	slightly	quite	extremely	Displeased
Contended	extremely	quite	slightly	neither	slightly	quite	extremely	Frustrated
Delighted	extremely	quite	slightly	neither	slightly	quite	extremely	Disappointed

Appendix F: Exit interview

1. Please reflect on the session that you just experienced. Do you have any comments on search tasks, MedlinePlus system, or your search process?
2. Did you anticipate that the measurements would repeat themselves during the session at three different time spots? If so, did you try to come up with concepts or prepare for the interview questions when you performed the search tasks

Appendix G: Coding schema for concept listing data

1. System

Definition: subjects' representation of attributes or characteristics of the overall MedlinePlus system.

- System: structure

  Definition: MedlinePlus is a database driven website. It selects, compiles, organizes, and maintains links to information at many government health agencies and professional health organization websites. MedlinePlus has its own content, but is limited to summary and overview information. MedlinePlus also license content, such as encyclopedia, dictionary, and news feeds. Concepts describe any of these aspects of the structure of MedlinePlus will be coded under this category.

  Examples: 1) Database
  2) Outside resources

- System: audience

  Definition: subjects' thoughts of audience of MedlinePlus.

  Examples: 1) Layman
  2) Patient

- System: agencies

  Definition: concepts concerning agencies involved in creating MedlinePlus or contributing information to the system.

  Examples: 1) Medical associations
  2) National Institute of Health

- System: system's behavior

  Definition: concepts describe the way the system behaves, such as pop-up window when clicking on a link

  Examples: 1) Search results yield snippets
  2) Redirected to another page

- System: usage of the system

  Definition: concepts express subjects' thoughts of what MedlinePlus could be used for.
  Examples: 1) Used to see if any major health issues

2) Self-diagnosing

- System: similar sites

    Definition: concepts describe other websites that are similar to MedlinePlus

    Examples: 1) WebMD
    2) Pubmed

- System: evaluation

    Definition: Subjects' evaluation of the MedlinePlus as a system, not any particular part of the system.

    Examples: 1) Helpful
    2) Quick access

2. Content

Definition: subjects' representation of subjects, attributes, or characteristics of the content/information in MedlinePlus.

- Content: general

    Definition: concepts that describe information in general.

    Examples: 1) Information
    2) Advice

- Content: subject

    Definition: concepts that describe the subject of information in MedlinePlus

    Examples: 1) Symptoms
    2) Treatments

- Content: specific

    Definition: concepts that describe specific content, such as specific diseases, conditions, or drugs in MedlinePlus.

    Examples: 1) Diabetes
    2) Insulin

- Content: type

    Definition: concepts describe the different types of information, such as dictionary, encyclopedia, news, journals, tutorials, and so on.

Examples: 1) Encyclopedia
2) News

- Content: format
  Definition: concepts describe the format of information in MedlinePlus, such as text, video, and images

  Examples: 1) Video
  2) PDF

- Content: presentation
  Definition: concepts describe the way information is presented, such as summary and overviews.

  Examples: 1) Overview
  2) Summary

- Content: evaluation
  Definition: subjects' evaluation of various aspects of content/information in MedlinePlus

  Examples: 1) Informative
  2) Reliable sources

3. Information organization

Definition: Concepts describes how information in MedlinePlus is organized.

- IO: schema
  Definition: the way in which information in MedlinePlus is organized, such as hierarchical and alphabetical

  Examples: 1) Alphabetical
  2) Organization by anatomy

- IO: evaluation
  Definition: Subjects' evaluation of information organization in MedlinePlus

  Examples: 1) Good listing orders
  2) Well-organized

4. Interface

Definition: concepts that reflect subjects' understanding of interface elements, functions embedded in the elements, navigation tools, and results access.

- Interface: elements

  Definition: Design elements of interfaces, such as tabs, menus, and links.

  Examples: 1) Links
  2) Search bar

- Interface: functions

  Definition: functions that made available through interface elements, such as search and navigation.

  Examples: 1) Search
  2) Spelling check

- Interface: results

  Definition: concepts describe the results, presentation of results, and results access.

  Examples: 1) Does not rank relevance
  2) Search choices come up in a logical order

- Interface: evaluation

  Definition: Subjects' evaluation of the interface elements, functions, and results.

  Examples: 1) Too many links per page
  2) Pretty colors

5. Heuristics

Definition: Heuristics are subjects' perceptions about how to use the system. They are rules of thumb about what is good to do and what is not, what is easy and what is difficult to do within the system.

Examples: 1) Typing in a question does not work well
2) Search for statements, not questions

6. Experience of using the system

Definition: Experience of using the system is evaluations of and emotions about their experience of using MedlinePlus to perform the tasks

Examples: 1) Not as satisfied this time
2) A kid's website seemed to be one of the most popular

References

Ahmed, I, & Blustein, J. (2006). Influence of Spatial Ability in Navigation: Using look-ahead Breadcrumbs on the web. *International Journal of Web Based Communities, 2*(2), 183-196.

Allen, B. L. (1996). *Information Tasks: Toward a User-centered Approach to Information Systems*. San Diego: Academic Press.

Allen, R. B. (1997). Mental models and user models. In M. Helander, T.K. Landauer, & P., Prablu (Eds.), *Handbook of Human-Computer Interaction* (2nd ed., pp. 49-63). New York, NY: Elsevier.

Alonso, D. (1998). *The Effects of Individual Differences in Spatial Visualization Ability on Dual-Task Performance*. Ph.D. Dissertation. University of Maryland.

Bayman, P., & Mayer, R.E. (1983). A diagnosis of beginning programmers' misconceptions of BASIC programming statements. *Communications of the ACM, 26*(9), 677-679.

Bayman, P., & Mayer, R. E. (1984). Instructional manipulation of users' mental models for electronic calculators. *International Journal of Man-Machine Studies 20*, 189-199.

Belkin, N.J., Brooks, H.M., & Oddy, R.N. (1982). Ask for information retrieval. *Journal of Documentation, 38*(2), 61-71.

Bell, D.J., & Ruthven, I. (2004). Searchers' assessments of task complexity for web searching. In S. McDonald & J. Tait, J. (Eds.), *Advances in Information Retrieval, 26th European Conference on Information Retrieval. Lecture Notes in Computer Science 2997* (pp. 57-71). Sunderland, UK, Springer.

Besnard, D., Greathead, D., & Baxter, G. (2004). When mental models go wrong: co-occurrences in dynamic, critical systems. *International Journal of Human-Computer Studies, 60* (1), 117-128.

Bibby, P.A. (1992). Mental models, instruction and internalization. In Y. Rogers, A. Rutherford & P. A. Bibby (Eds.), *Models in the Mind: Theory, Perspective and Application* (pp. 154-172). London: Academic Press.

Bilal, D. (2000). Children's use of the Yahooligans! Web search engine: I cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science, 51*(7), 646-665.

Bilal, D. (2001). Children's use of the Yahooligans! Web search engine. II. Cognitive and physical behaviors on research tasks. *Journal of the American Society for Information Science and Technology, 52*(2), 118-136.

Bilal, D. (2002). Children's use of the Yahooligans! Web search engine. III. Cognitive and physical behaviors on fully self-generated search tasks. *Journal of the American Society for Information Science and Technology, 53*(13), 1170-1183.

Borgman, C.L. (1984). *The User's Mental Model of an Information Retrieval System: Effects on Performance*. Unpublished Doctoral Dissertation, Stanford University.

Borgman, C. L. (1986). The users mental model of an information-retrieval system - an experiment on a prototype online catalog. *International Journal of Man-Machine Studies, 24*(1), 47-64.

Borgman, C.L. (1989). All users of information retrieval systems are not created equal: an exploration into individual differences. *Information Processing and Management, 25*(3), 237-251.

Borgman, C.L. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science, 47*(7), 493-503.

Borlund, P. (2000). *Evaluation of Interactive Information Retrieval Systems*. Åbo, Finland: Åbo Akademi University Press.

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

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review 85*, 1-21.

Brandt, D. S., & Uden, L. (2003). Insight into mental models of novice Internet searchers. *Communications of ACM, 46*(7), 133-136.

Brehmer, B. (1987). System design and the psychology of complex systems. In: J. Rasmussen & P. Zuhde (Eds.), *Empirical Foundations of Information and Systems Science* (pp. 21-32). New York: Plenum Press.

Brown (1986). From cognitive to social ergonomics and beyond. In D.A. Norman & S. W. Draper (Eds.), *User Centered System Design: New Perspectives on Human-Computer Interaction* (pp. 457-486). Hillsale, NJ: LEA.

Bruce, H. (1999). Perceptions of the Internet: What people think when they search the Internet for information. *Internet Research, 9*(3), 187-199.

Brusilovsky, P., & Millan, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science, 4321,* 3-53. Springer-Verlag.

Byrne, R. M. J. (1992). The model theory of deduction. In Y. Rogers, A. Rutherford & P. A. Bibby (Eds.), *Models in the Mind: Theory, Perspective and Application.* London: Academic Press.

Byström, K., & Hansen, P. (2005). Conceptual framework for tasks in information studies.

Journal of the American Society for Information Science & Technology, 56(10), 1050-1061.

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management, 31*(2), 191-213.

Campagnoni, F., & Ehrlich, K. (1989). Information retrieval using a hypertext-based help system. *ACM Transactions on Information Systems, 7*, 271-291.

Cacioppo, J. T., von Hippel, W., & Ernst, J. M. (1997). Mapping cognitive structures and processes through verbal content: The thought-listing technique. *Journal of Consulting and Clinical Psychology*, *65*, 928-940.

Campbell, D.J. (1988). Task complexity: A review and analysis. *Academy of Management Review, 13*(1), 40-52.

Capra, R., Marchionini, G., Oh, J. S., Stutzman, F., & Zhang, Y. (2007). Effects of structure and interaction style on distinct search tasks. *Proceedings of ACM/IEEE-CS Joint Conference on Digital Libraries* (JCDL07), 442-451.

Carbonell, J. (1979). *Subjective Understanding: Computer Models of Belief Systems*. Technical Report 150. New Haven, Conn.: Yale University, Department of Computer Science.

Card, S., & Moran, T. (1986). User technology: from pointing to pondering. *Proceedings of the ACM Conference on the History of Personal Workstations*, 183-198.

Card, S.K., Moran, R., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-analytic Studies*. NY: Cambridge University Press.

Carroll, J. M, & Thomas, J.C. (1982). Metaphor and the cognitive representation of computing systems. *IEEE Transactions on Systems, Man, and Cybernetics, SMC 12*(2), 107-116.

Carroll, J. M., & Olson, J. R. (1987). Mental models in human-computer interaction: Research issues about what the user of software knows. Committee on Human Factors, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, DC: National Academy Press

Chang, S. N. (2007). Externalizing students' mental models through concept maps. *Educational Research, 41*(3): 107-112.

Chen, C. (2000). Individual differences in a spatial-semantic virtual environment. *Journal of the American Society for Information Science, 51*(6), 529-542.

Chen, C., & Czerwinski, M. (1997). Spatial ability and visual navigation: An empirical study. *New Reviews of Hypermedia and Multimedia, 3*, 67-89.

Chen, C., & Rada, R. (1996). Interacting with hypertext: A meta-analysis of experimental studies. *Human-Computer Interaction, 11*, 125-156.

Chen, H., & Dhar, V. (1990). User misconceptions of information retrieval systems. *International Journal of Man-Machine Studies, 32*, 673-692.

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

Clement, J. (1983). A conceptual model discussed by Galileo and used intuitively by physics students. In D. Gentner & A.L. Stevens (Eds.), *Mental Models* (pp. 325-340). Hillsdale, NJ: LEA.

Conant, R. C., & Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science, 1*(2): 89-97.

Cool, C., Park, S., Belkin, N. J., Koenemann, J., & Ng, K. B. (1996). Information seeking behavior in new searching environments. *Proceedings of Conceptions of Library Science*, 403-416.

Cooper, A. (1995). *About Face: The Essentials of User Interface Design*. Foster City, CA: Programmer's Press.

Craik, K. (1943). *The Nature of Explanation*. Cambridge: Cambridge University Press.

Cramer, P. (1968). *Word Association*. NY: Academic Press.

Crudge, S. E., & Johnson, F. C. (2004). Using the information seeker to elicit construct models for search engine evaluation. *Journal of the American Society for Information Science and Technology, 55*(9): 794-806.

Czerwinski, M. & Larson, K. (1997). The new Web browsers: They're cool but are they useful? *Proceedings of HCI'97*. Berlin: Springer Verlag.

Dahlbäck, N., Höök, K., & Sjölinder, M. (1996). Spatial cognition in the mind and in the world- the case of hypermedia navigation. *Proceedings of the 18th annual meeting of the cognitive science society*.

deKleer, J. & Brown, J. S. (1983). Assumptions and ambiguities in mechanistic mental models. In D. Gentner & A.L. Stevens (Eds.), *Mental Models* (pp. 155-190). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dillon, A. (2000). Spatial-semantics: How users derive shape from information space. *Journal of the American Society for Information Science, 51*(6): 521-528.

Dillon, A., Richardson, J., & McKnight, C. (1990). Navigation in hypertext: a critical review of the concept. In D. Diaper, D. Gilmore, G. Cockton & B. Shackel (Eds.), *Human-computer Interaction - INTERACT'90* (pp. 587-592). North Holland: Amsterdam, 587-592.

Dimitroff, A. (1990). *Mental Models and Error Behavior in an Interactive Bibliographic Retrieval System*. Unpublished doctoral dissertation, The University of Michigan.

Dimitroff, A. (1992). Mental models theory and search outcome in a bibliographic retrieval system. *Library and Information Science Research, 14*(2), 141-156.

diSessa, A. (1986). Models of computation. In D. A. Norman & S. W. Draper (Eds.), *User-Centered System Design: New Perspectives in Human-Computer Interaction*. Hillsdale, NJ: LEA.

diSessa, A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 15-33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Doyle, J. K., & Ford, D. N. (1998). Mental models concepts for system dynamics research. *System Dynamics Review, 14*(1), 3-29.

DuBoulay, B., O'Shea, T., & Monk, J. (1981). The black box inside the glass box : Presenting computing concepts to novices. *International Journal of Man-Machine Studies, 14*, 237-249.

Eckstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Kit of Factor-referenced Cognitive Tests*. Princeton, N.J.: Educational Testing Service.

Efthimiadis, E. N., & Hendry, D. G. (2005). Search engines and how students think they work. *Proceedings of SIGIR'05* (pp. 595-596). New York: ACM.

Ehrlich, K. (1996). Applied mental models in human-computer interaction. In J. Oakhill & A. Garnham (Eds.), *Mental Models in Cognitive Science* (pp. 223-245). Hove, East Sussex, U.K.: Psychology Press.

Farris, J. S., Jones, K. S., & Elgin, P. D. (2002). Users' schemata of hypermedia: What is so spatial about a website? *Interacting with Computers, 14*(5), 487-502.

Fein, R. M., Olson, G. M, & Olson, J. S. (1993). A mental model can help with learning to operate a complex device. *INTERACT '93 and CHI '93 conference companion on Human factors in computing systems*, 157-158.

Fischer, G. (1991). The importance of models in making complex systems comprehensible. In M. J. Tauber and D. Ackermann (Eds.), *Mental Model and Human-Computer Interaction 2* (pp. 3-36). Amsterdam, North-Holland.

Fischer, G. (2001). User modeling in human-computer interaction. *User Modeling and User-Adapted Interaction, 11*, (1-2), 65-86.

Fogg, B. J. (2003). Prominence-Interpretation theory" Explaining how people assess credibility online. *Proceedings of the CHI-2003*, 722-723.

Forbus, K.D. (1983). Qualitative reasoning about space and motion. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 53-72). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fox, S. (2005). *Pew Internet & American Life Project: Health Information Online*. http://www.pewinternet.org/pdfs/PIP_Healthtopics_May05.pdf (Accessed Sep. 5, 2008).

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 99-130). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D., & Stevens, A. L. (1983). *Mental Models*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Glaser, B.G., & Strauss, A.L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* New York: Aldine.

Glenberg, A. M., and Langston, W. E. (1992) Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language, 31*, 129-151

Graham, J, Zheng, L. Y, & Gonzalez, C. (2006). A cognitive approach to game usability and design: Mental model development in novice real-time strategy gamers. *CyberPsychology & Behavior, 9*(3): 361-366.

Gray, S. H. (1990). Using protocol analyses and drawings to study mental models construction during hypertext navigation. *International Journal of Human-Computer Interaction, 2*, 359-377.

Grunig, J.E., Ramsey, S., & Schneider, L. A. (1985). An axiomatic theory of cognition and writing. *Journal of Technical Writing and Communication, 15*(2), 95-131.

Halasz, F. G., & Moran, T. P. (1983). Mental models and problem-solving in using a calculator. In *Proceedings of CHI'83 Human Factors in Computing Systems*. New York: ACM.

Hancock-Beaulieu, M., Fieldhouse, M., & Do, T. (1995). An evaluation of interactive query expansion in an online library catalog with graphical user interface. *Journal of Documentation, 51*, 225-243.

Hammond, N., Jorgensen, A., MacLean, A., Barnard, P. & Long, J. (1983). Design practice and interface usability: Evidence from interviews with designers. *CHI'83 Proceedings*, 40-44.

Hanisch, K. A., Kramer, A. F., & Hulin, C. L. (1991). Cognitive representations, control, and understanding of complex systems: A field study focusing on components of users' mental models and expert/novice differences, *Ergonomics, 34*(8), 1129-1145.

Henderson, L., & Tallman, J. (2006). *Stimulated Recall and Mental Models: Tools for Teaching and Learning Computer Information Literacy*. Lanham, Maryland: The Scarecrow Press, Inc.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human*

Interaction, 7(2), 174-196.

Huang, M. S. (1992). *Pausing Behavior of End-users in Online Searching*. Unpublished Doctoral Dissertation. The University of Maryland, College Park.

Hui, B., Partridge, G., & Boutilier, C. (2009). A probabilistic mental model for estimating disruption. *Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI)*, 287-296.

Ingwersen, P. (1992). *Information Retrieval Interaction*. London: Taylor Graham.

Ingwersen, P., & Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context.* Netherlands: Springer.

Johnson-Laird, P. N. (1981). Mental models of meaning. In A. Joshi, B. Webber & I. Sag (Eds.), *Elements of Discourse Understanding*. Cambridge, UK: Cambridge University Press.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge, Massachusetts: Harvard University Press.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99*, 418-439.

Katzeff, C. (1990). System demands on mental models for a fulltext database. *International Journal of Man-Machine Studies, 32*, 483-509.

Kellar, M., Watters, C. & Shepherd, M. (2007). A field study characterizing Web-based information seeking tasks. *Journal of the American Society for Information Science and Technology, 58(*7), 999-1018.

Kelly, D. & Belkin, N. J. (2002). A user modeling system for personalized interaction and tailored retrieval in interactive IR. In *Proceedings of Annual Conference of the American Society for Information Science and Technology (ASIST '02)*, 316-325.

Kerr, S. T. (1990). Wayfinding in an electronic database: the relative importance of navigational cues VS. mental models. *Information Processing & Management, 26* (4), 511-523.

Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science 8*(3): 355-273.

Kim, K. S. (2000). Users, tasks, and the Web: Their impact on the information-seeking behavior. *Proceedings of the 21st National Online Meeting*, 189-198.

Kim, K. S., & Allen, B. (2001). Cognitive and task influences on web searching behavior. *Journal of American Society for Information Science and Technology, 53*(2). 109-119.

Kobsa, A. (2001). Generic user modeling systems. *User Modeling and User-Adapted Interaction, 11*, 49-63.

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press.

Krippendorff, K. (2004). *Content Analysis: An Introduction to its Methodology*. Thousand Oaks, California: Sage Publication, Inc.

Landy, M. S., Maloney, T., & Pavel, M. (1996). *Exploratory Vision: The Active Eye*. Springer Verlag..

Langan-Fox, J., Platania-Phung, C., & Waycott, J. (2006). Effects of advance organizers, mental models and abilities on task and recall performance using a mobile phone network. *Applied Cognitive Psychology, 20*: 1143-1165.

Lohman, D. F. (1979). *Spatial Ability: A Review and Reanalysis of the Correlation Literature*. Stanford, CA: Aptitude Research Project, School of Education, Stanford University Technical Report No.8.

Lindberg, D. A. B. (2000). Internet access to the national library of medicine. *Effective Clinical Practice, 3*(5), 256-260.

Marill, J. L., Miller, N., & Kitendaugh, P. (2006). The MedlinePlus public user interface: studies of design challenges and opportunities. *Journal of Medical Library Association, 94*(1), 30-40.

Makri, S., Blandford, A., Gow, J., Rimmer, J., Warwick, C., & Buchanan, G. (2007). A library or just another information resource? A case study of users' mental models of traditional and digital libraries. *Journal of the American Society for Information Science and Technology, 58*(3), 433-445.

Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge, England: Cambridge University Press.

Marchionini, G. (1989a). Making the transition from print to electronic encyclopedias - Adaptation of mental models. *International Journal of Man-Machine Studies, 30*(6), 591-618.

Marchionini, G. (1989b). Information seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science, 40*(1), 54-66.

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM, 49*(4), 41-46.

Mayer, R. E. (2002). *Multimedia Learning*. NY: Cambridge University Press.

Mayer, R. E. & Bayman, P. (1981). Psychology of calculator languages: A framework for describing differences in users' knowledge. *Communications of the ACM, 24*, 511-520.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper learning of multimedia messages? *Journal of Educational*

Psychology, 93(2), 390–397.

Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages though pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied, 8*(3), 147-154.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The Psychology of Computer Vision* (pp. 211-277). New York: McGraw-Hill.

Modjeska, D., & Chignell, M. (2003). Individual differences in exploration using desktop VR. *Journal of the American Society for Information Science and Technology, 54*(3), 216-228.

Moray, N. (1987). Intelligent aids, mental models and the theory of machines. *International Journal of Man-Machine Studies, 27*, 619-629.

Muramatsu, J., & Pratt, W. (2001). Transparent queries: investigating users' mental models of search engines. *Proceedings of the Twenty-fourth International ACM Conference on Research and Development in Information Retrieval SIGIR 2001* (pp. 217-224). New York: ACM.

Neumann, L. J., & Ignacio, E. N. (1998). Trial and error as a learning strategy in system use. *ASIS '98: Proceedings of the 61st ASIS Annual Meeting, 35,* 243-252.

Nielsen, J. (1989). The matters that really matter for hypertext usability. *Proceedings of ACM Hypertext ’89*, 239-248.

Nielsen, J. (1990). A meta-model for interacting with computers. *Interacting with Computers 2*: 147-160.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 7-14). Hillsdale, NJ: Erlbaum.

Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User-Centered System Design: New Perspectives in Human-Computer Interaction*.. Hillsdale, NJ: Lawrence Erlbaum Associates.

Norman, D. A. (1988). The Psychology of Everyday Things. New York: Basic Books. Also published in paperback as D. A. Norman. (1990). The Design of Everyday Things. New York: Doubleday.

Norman, D. A., & Bobrow, D. G. (1979). Descriptions: an intermediate stage in memory retrieval. *Cognitive Psychology, 11*, 107-123.

O'Regan, J. K., & Noe, A. (2001) A sensorimotor account of vision and visual consciousness. *Behavior and Brain Sciences, 24*(5)

Otter, M., & Johnson, H. (2000). Lost in hyperspace: metrics and mental models. *Interacting With Computers, 13*, 1-40.

Pak, R., Rogers, W. A., & Fisk, A. D. (2006). Spatial ability subfactors and their influences on a computer-based information search task. *Human Factors, 48*(1), 154-165.

Papastergiou, M. (2005). Students' mental models of the Internet and their didactical exploitation in informatics education. *Education and Information Technology, 10*(4), 341-360.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. NY: Oxford University Press.

Pejtersen, A. M. (1991). *Interfaces Based on Associative Semantics for Browsing in Information Retrieval*. Roskilde: Risø Laboratory.

Pejtersen, A.M., & Fidel, R. (1998). A framework for work centered evaluation and design: A case study of information retrieval on the Web. Working Paper for *MIRA workshop*.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88*, 16-45.

Pylyshyn, Z. (1999). What's in your mind? In: E. Lepore & Z. Pylyshyn (Eds.), *What is Cognitive Science?* Black-well.

Pylyshyn, Z. (2002) Mental imagery: In search of a theory. *Behavioral and Brain Sciences 25*, 157-182

Qiu, L. (1993). Analytical searching vs. browsing in hypertext information retrieval systems. *Canadian Journal of Information and Library Science 18*(4), 1-13.

Rasmussen, J. (1979). *On the Structure of Knowledge – A Morphology of Mental Models in Man-Machine System Context* (Tech. Rep. No. Riso-M-2192). Roskilde, Denmark: Riso National Laboratory. (Cited from Rouse and Morris, 1986)

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction*. Amsterdam: North-Holland.

Rasmussen, J., Pejtersen, A.M., & Schmidt, K. (1990). *Taxonomy for Cognitive Work Analysis* (Rise-M-2871). Roskilde, Demark: Rise National Laboratory.

Riley, M. S. (1986). User Understanding. In D. A. Norman & S. W. Draper (Eds.), *User-Centered System Design: New Perspectives in Human-Computer Interaction* (pp157-169). Hillsdale, New Jersey: Prentice Hall.

Roger, Y. (1992). Introduction. In Y. Rogers, A. Rutherford & P. A. Bibby (Eds.), *Models in the Mind: Theory, Perspective and Application.* London: Academic Press.

Roger, Y., Sharp, H., Benyon, D. Holland, S., & Carey, T. (1994). *Human-Computer Interaction*. Menlo Park, CA: Addison-Wesley.

Rosson, M.B. (1983). Patterns of experience in text editing. In *Proceedings of the CHI-83 Conference on Human Factors in Computing* (pp. 171-175). New York: ACM.

Rouse, W. B., & Morris, N. M. (1986). On Looking Into the Black-Box - Prospects and Limits in the Search for Mental Models. *Psychological Bulletin, 100*(3), 349-363.

Rumelhart, D.E. (1980). Schemata: the building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale: NJ: Erlbaum.

Rumelhart, D. E. and Norman, D. A. (1978). Accretion, tuning and restructuring: three modes of learning. In J.W. Cotton & R.L. Klatzky (Eds.), *Semantic Factors in Cognition*. Hillsdale. N.J.: LEA.

Rumelhart, D.E. and Ortony, A. (1977) The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro & W.E. Montague (Eds.), *Schooling and the Acquisition of Knowledge*. Hillsdale, N.J.: LEA.

Rutherford, A., & Wilson, J.R. (1992). Searching for mental models in human-machine systems. In Y. Rogers, A. Rutherford & P. A. Bibby (Eds.), *Models in the Mind: Theory, Perspective, & Application* (pp.195-223). SD, CA: Academic Press Inc.

Sanderson, P. M. (1990). Knowledge acquisition and fault diagnosis: Experiments with PLAULT. *IEEE Transactions on Systems, Man and Cybernetics 20*, 225-242.

Sasse, M.A. (1991). How to t(r)ap users' mental models. In M. J. Tauber and D. Ackermann (Eds.), *Mental Models and Human-Computer Interaction 2*. (pp. 59-79). North-Holland, Amsterdam: Elsevier Science.

Savage-Knepshield, P. A. (2001). *Mental Models: Issues in Construction, Congruency, and Cognition*. Unpublished doctoral dissertation, Rutgers, The State University of New Jersey.

Sein, M. K., Olfman, L., Bostrom, R. P., & Davis, S. A. (1993). Visualization ability as a predictor of user learning success. *International Journal of Man-Machine Studies, 39*, 599-620.

Schank, R. C., & Abelson, R.P. (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. New York: Academic.

Schank, R.C., & Childers, P. (1984). *The Cognitive Computer on Language, Learning, and Artificial Intelligence*. CA: Addison-Wesley Publishing Company, Inc.

Schank, R.C. (1999). *Dynamic Memory Revisited*. NY: Cambridge University Press.

Sein, M. K., Olfman, L., Bostrom, R. P., & Davis, S. A. (1993). Visualization ability as a predictor of user learning success. *International Journal of Man-Machine Studies, 39*, 599-620.

Schvaneveldt, R.W. (1990). *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, NJ: Ablex.

Sharit, J, Hernández, M.A., Czaja, S.J., & Pirolli, P. (2008). Investigating the roles of knowledge and cognitive abilities in older adult information seeking on the web. *ACM Transactions on Computer-Human Interaction, 15*(1), 1-25.

Shneiderman, B. (1997). Designing information-abundant web sites: issues and recommendations. *International Journal of Human-Computer Studies, 47*, 5-29.

Shrager, J. and Klahr, D. (1983). Learning in an instructionless environment: observation and analysis. In *Proceedings of the 1983 CHI Conference on Human Factors in Computing*. NY: ACM.

Slone, D. J. (2002). The influence of mental models and goals on search patterns during Web interaction. *Journal of the American Society for Information Science and Technology, 53*(13), 1152-1169.

Solomon, P. (2002). Discovering information in context. *Annual Review of Information Science and Technology (ARIST), 36*, 229-264.

Staggers, N., & Norcio, A. F. (1993). Mental Models - Concepts for Human-Computer Interaction Research. *International Journal of Man-Machine Studies, 38*(4), 587-605.

Stanney, K. M., & Salvendy, G. (1995). Information Visualization: Assisting low spatial individuals with information access tasks through the use of visual mediators. *Ergonomics, 38*(6), 1184-1198.

Streitz, N. A. (1988). Mental models and metaphors: implications for the design of adaptive user-system interfaces. In M. Heinz & L. Alan (Eds.). *Learning Issues for Intelligent Tutoring Systems*. NY: Springer-Verlag.

Swan, R. C., & Allan, J. (1998). Aspect windows, 3-D visualizations, and indirect comparisons of information retrieval systems. *SIGIR 1998*, 173-181.

Stanovich, K. E. (1999). *Who is Rational? Studies of Individual Differences in Reason*ing. New Jersey: Lawrence Erlbaum Asspciates, Publisher.

Thatcher, A., & Greyling, M. (1998). Mental Models of the Internet. *International Journal of Industrial Ergonomics, 22*(4-5), 299-305.

Thelen, E. and Smith, L.B. (1994) *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press.

Tognazzini, B. (1992). *Tog on Interface*. Reading, MA: Addisson-Wesley.

Vakkari, P. (1999). Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval. *Information Processing and Management 35*(6): 819-837.

Vakkari, P. (2003). Task-based information searching. *Annual Review of Information Science & Technology, 37*, 413-464.

van der Veer, G. C., & Puerta Melguizo, M. C. (2002). *Mental models.* In: J.A. Jacko & A.

Sears (Eds.) *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Uitgever: Lawrence Erlbaum & Associates.

van der Velden, J. M., & Arnold, A.G. (1992). Mental models and the evaluation of user interfaces: A case study of a library system. In M. J. Tauber & D. Ackermann (Eds.), *Mental Models and Human-Computer Interaction 2*. (pp. 179-189). North-Holland, Amsterdam: Elsevier Science.

Vicente, K. J., Hayes, B. C., & Williges, R. C. (1987). Assaying and isolating individual differences in searching a hierarchical file system. *Human Factors, 29*, 349-359.

Vicente, K. J. & Williges, R. C. (1988). Accommodating individual differences in searching a hierarchical file system. *International Journal of Man-Machine Studies, 29*, 674-688.

Waern, Y. (1985). Learning Computerized Tasks as Related to Prior Task Knowledge. *International Journal of Man-Machine Studies, 22*(4), 441-455.

Wang, P., Bales, S., Reiger, J., & Zhang, Y. (2004). Survey of learners' knowledge structures: Rationales, methods and instruments. *Proceedings of the 67th Annual Meeting of the ASIS&T* (pp. 218-228). Medford, NJ: Information Today.

Wason, P.C., & Johnson-Laird, P.N. (1972). *The psychology of Deduction: Structure and Content*. Cambridge, MA: Harvard University Press, London: Batsford.

Westbrook, L. (2006). Mental models: a theoretical overview and preliminary study. *Journal of Information Science, 32*(6), 563-579.

Westerman, S. J. (1997). Individual differences in the use of command line and menu computer interfaces. *International Journal of Human-Computer Interaction, 9*, 183-198.

Westerman, S. J. (1998). A comparison of the cognitive demands of navigating two-versus three-dimensional spatial database layouts. *Ergonomics, 41*(2), 207-212.

Westerman, S. J., Collins, J., & Cribbin, T. (2005). Browsing a document collection represented in two- and three-dimensional virtual information space. *International Journal of Human-Computer Studies, 62*, 713-736.

White, R. W., Ruthven, I., & Jose. J. (2005). A study of factors affecting the utility of implicit relevance feedback. *SIGIR'05*, 35-42.

Wildemuth, B. M., & Hughes, A. (2005). Perspectives on the tasks in which information behaviors are embedded. In K. E. Fisher, S. Erdelez & L. McKechnie (Eds.), *Theories of Information Behavior* (pp. 275-279). Medford, NJ: Information Today.

Wilensky, R. (1978). *Understanding Goal-based Stories.* Technical Report 140. New Haven, Conn.: Yale University, Department of Computer Science.

Williams, M. D, Hollan, J. D., & Stevens, A. L. (1983). Human reasoning about a simple physical system. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 131-153). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wilson, J. R., & Rutherford, A. (1989). Mental models - theory and application in human factors. *Human Factors, 31*(6), 617-634.

Wright, P., & Bason, G. (1982). Detour routes to usability: a comparison of alternative approaches to multipurpose software design. *International Journal of Man-Machine Studies, 18*, 391-400.

Young, R. M. (1983). Surrogates and mappings: Two kinds of conceptual models for interactive devices. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 35-52). Hillsdale, NJ: Lawrence Erlbaum Associates.

Zhang, X. (1997). *A Study of the Effects of User Characteristics on Mental Models of Information Retrieval Systems*. Unpublished doctoral dissertation, University of Toronto.

Zhang, Y. (2008a). The influence of mental models on undergraduate students' searching behavior on the Web. *Information Processing & Management 44* (3):1330-1345

Zhang, Y. (2008b). Undergraduate students' mental models of the Web as an information retrieval system. *Journal of the American Society for Information Science and Technology, 59*(13), 2087-2098.

Zhang, Y., & Wildemuth, B. (2009). Thematic content analysis. In B. Wildemuth (Ed.), *Applications of Social Research Methods to Questions in Information and Library Science* (pp. 222-231). Westport, CT: Libraries Unlimited.