# Health Information Searching Behavior in MedlinePlus and the Impact of Tasks

Yan Zhang
School of Information
University of Texas at Austin
Austin, TX 78701-1213
1 512 471 9448
yanz@ischool.utexas.edu

Peiling Wang
School of Information Sciences
University of Tennessee
Knoxville, TN 37996-0341
1 865 974 3700
pelingw@utk.edu

Amy Heaton & Heidi Winkler
School of Information
University of Texas at Austin
Austin, TX 78701-1213
1 512 471 9448

## ABSTRACT
Consumer health information is increasingly available online, but this vast amount of information is not necessarily accessible to general consumers. To design effective health information websites, it is critically important to gain an in-depth understanding of how consumers search for health information in these systems. This study is an attempt to explore consumer health information searching behavior in web-based health information spaces by observing their search behaviors in MedlinePlus. Nineteen undergraduate students accomplished three search tasks in MedlinePlus. The participants used both searching and browsing as interaction strategies. This paper reports on the findings of their searching behaviors, particularly query construction, query reformulation, and access to results; and their browsing behaviors, particularly access to different resources, health topics, and related topics. Furthermore, we examined if the number of concepts involved in search tasks had any impact on search behaviors.

## Categories and Subject Descriptors
H.1.2 [**Information Systems**]: User/Machine Systems – *human factors, human information processing.*

## General Terms: Design, Human Factors.

## Keywords: Consumer health informatics, health information searching behavior, search tasks, health information systems, MedlinePlus

## 1. INTRODUCTION
In the past decade, the Web has become one of the most important channels for general consumers to research medical problems and seek health-related advice. A recent Pew study revealed that, as of 2009, more than 80% of the Internet users in the U.S. look online for health information, up from 25% in 2000 [13]. Furthermore, health information seeking was uniformly ranked, across all age groups, the third most popular online activity, after email and search engine use [36]. The information found online was found to have a significant impact on consumers' decisions on treatment choices, how to manage a chronic condition, and how to maintain their own health or the health of someone in their care [13].


*IHI'12*, January 28-30, 2012, Miami, Florida, USA.


In response to this surge of demand for health information and the great impact this information has on health-related decisions, numerous consumer health information websites and information systems were developed or renovated to provide consumers access to medical and health information in the past several years [16]. Being available, however, does not necessarily guarantee being easy to access or use. Consumers continued to report difficulties in finding relevant information using the existing sources [33]; and they still felt confused and frustrated about their search experience [1].

To change this situation, efforts were made to improve consumers' experience with searching for health information. These efforts mainly focused on designing better query processing and query recommendation mechanisms [18, 33], developing better ways to organize and display search results [18], and constructing effective information architecture to integrate various medical information sources and to help users navigate through a health information space [19]. However, in this process, relatively little attention has been paid to issues such as, how consumers actually search and move around in a health information web space and what strategies they use to find relevant information in such an unfamiliar space [9, 11]. Knowledge of these subjects is necessary for developing a user-friendly and easy-to-access consumer health information system.

This paper reports on part of a larger project on users' mental models of a health information website, MedlinePlus [35]. Developed by the National Library of Medicine (NLM), the site provides authoritative health information to the public. As a typical health information web space, it contains various types of resources, such as medical news, health topics, dictionary, and encyclopedia organized in categories for navigation and a search engine. This paper focuses on analyzing users' interaction behaviors in MedlinePlus, and the impact of tasks on strategies. Specifically, we address the following research questions:

1. How do users search MedlinePlus?
2. How do users browse MedlinePlus?
3. How do search tasks influence interaction strategies?

Through the close examination of the interaction processes and strategies, we aim at an in-depth look into the problems searchers encountered in using MedlinePlus and the ways by which they handled the problems. The results can shed light on the design of similar consumer health information web spaces in particular and Web information search systems in general.

## 2. RELATED LITERATURE
Consumer health informatics is becoming an increasingly important research area in health informatics; it studies

technologies and communications for bridging the gap between patients (and the general public) and health information [10]. As more and more users are searching health information online by themselves, researchers are paying more attention to consumers' online health information searching behavior [e.g., 26]. Researchers have observed search topics, search processes, queries and reformulation of queries, access to search results, and evaluation of health information.

An early study [25] analyzing queries submitted to Excite and Ask Jeeves found that consumers were looking for information about general medical and health, human relationships, weight, reproductive health/puberty, and pregnancy. Zhang, et al. [34] analyzed queries submitted to HealthLink website and found that the most searched key words included *stroke*, *depression*, *cholesterol*, *hip*, and *stomach*. A recent survey by the Pew Research [13] reported that Americans searched online for information about specific diseases, medical problems, treatments, prescriptions, and over-the-counter drugs as well as information about doctors, hospitals, other medical facilities, and health insurance. Consumers also searched online for wellness information, such as exercise, fitness, diets, and information concerning how to lose weight and how to stay healthy on a trip overseas.

When searching for health information, consumers tended to begin with a general web search engine, instead of a medical website. In an experimental study, Eysenbach and Kohler [11] reported that none of their participants used medical portals or the sites of medical societies or libraries as a starting point for their searches. Sillence, et al. [24] found similar results: none of their participants, patients with menopause, used medical portals as a starting point to look for menopause information. Although a few of them attempted to go the National Health Service (NHS) website, they failed to type in the correct address. In a study of undergraduate and graduate students' searching for health information, Efthimiadis [9] found that about 80% of the students started their searches by going to a general search engine or a directory-based search; the rest started their searches from the websites created by the university, not-for-profit professional organizations, or government.

Query formulation is another aspect of health information searching behavior that has drawn a substantial number of studies. Consumer health queries were often characterized as simple and short. Zeng, et al. [31] found that about 90% of the queries that users submitted to a hospital website and MedlinePlus contained 1-3 terms. Hong et al. [14] found that the average length of queries submitted to HealthLink, a consumer health information resource, was 2.1 words. The analysis of HealthLink search logs in 2005 (377,000 queries) by Wang and associates found the average query length to be 1.78 terms and each session contained 2.34 queries [27]. Spink, et al. [25] reported an average of 2.2 terms per health query in general-purpose web search engines. A more recent analysis of health-related queries from general web search engines reported an average number of 3.3 terms per query [28]. Another recent observational study pointed out that the length of queries was dependent on search tasks. In the study, the query length ranged from 1.6-5.9 terms for different tasks [9].

Formulating health-related queries involves different levels of difficulties for general consumers. At a conceptual level, consumers have their own understandings and hypotheses about a particular condition. This understanding shapes their search strategies. But most likely, this understanding deviates from an expert's knowledge [14]. Such discrepancy could lead to less optimal search results. At the linguistic level, sometimes, users could not find specific or effective keywords to describe their intentions [5, 33, 34]. Their vocabulary often does not match with the medical terminologies that are found in medial taxonomies or used in indexes of medical information collections. This problem affects the success of searching for health information. At the query level, misspellings, partial words, run-together words, and abbreviations are common, which also cause search failures [4, 21].

Query reformulation is an important aspect of information searching behavior as it reflects tactical or conceptual changes that users experience during a search session [2, 12, 29]. However, only a few studies briefly touched on this aspect. Toms and Latter [26] pointed out that consumers' formulation of queries was a trial-and-error process. Sillence et al. [24] briefly described their observation of query reformulation in their study of fifteen menopause patients searching for disease-related information. They reported that the search strategies used by participants varied by their specific areas of interest. But they often began with simple searches using single keywords, then modified their search by making use of Boolean operators and regularly altered search terms to alter search results.

Accessing search results is an important step in the process of information searching [20. 23]. Eysenbach and Kohler [11] observed that the participants in their study usually chose one of the first results, with only a few checking out the results on the second or following pages. Differently, Toms and Latter [26] reported that their participants accessed, on average, 5.4 pages of results, with only 16% examining just the first page of results. Consumers employed various heuristics in determining which links to click on a result list. Some tended to rely on scanning the titles, summaries, and URLs, while paying little attention to information such as dates, size of the site, and type of file [26]. Some tended to look for certain keywords and source identifiers to select a link [24].

Consumers' evaluation and appraisal of health information from search results is another aspect of consumer health information searching that draws comparatively wide attention. Using a focus group technique, Eysenbach and Kohler [11] found that consumers used varies source features to judge the credibility of a source, such as authority of the source, layout and appearance, advertising, readability, and picture of the site owner. Based on the concurrent think-aloud protocols and after-search group discussions, Sillence et al. [24] found that poor design features (e.g., layout, navigation aids, and visual design) often led to the quick rejection of a health website. Content factors, such as informative content, relevant illustrations, a wide variety of topics covered, unbiased information, and clear, simple language, were more important than the design features in determining whether a site could be trusted and if they would explore the site in more depth. By analyzing post-search interviews, Toms and Latter [26] corroborated these findings and reported that participants' using or rejecting a site depended on relevance, quality, and presentation of the information as well as accessibility of the site.

It must be pointed out that studies used different data collection methods reported discrepant behaviors. The same participants in Eysenbach and Kohler's study [11] mentioned above failed to use the important quality indicators they had suggested during the previous focus group interviews. For example, they did not check the owner of the sites or read disclosure statements, and they neglected to find information about how the sources had been compiled when they were observed in real searching settings. In examining twenty participants' comments on a consumer health information system after they completed several search tasks, Williams, et al. [30] found that there was a lack of awareness of content quality.

It is apparent that an in-depth understanding of the consumer health information search process can reveal not only consumers' behavior in using health information systems but also the problems they encounter during the search. Such research eventually can help improve the design of health information systems. Although the effort to examine the consumer health information search process is increasing, there is still a lack of understanding on many aspects of this phenomenon, such as how users reformulate queries, how they make use of different types of resources, how tasks impact the deployment of search strategies, how individual differences affect searching, and how searchers adapt to a particular information system. This study presents an effort to explore some of these issues. As pointed out by Eysenbach and Kohler [11], in order to design innovative technologies for high quality consumer health information, more studies are needed to observe consumers behavior of using existing systems.

# 3. RESEARCH METHOD

## 3.1 Participants

Twenty undergraduate students, 11 females and 9 males, from a major research university participated in the study. The participants were recruited by sending out an email to the campus-wide listserv. The participants were from the mathematics, journalism, communication studies, psychology, business, history, Spanish, political science, and sociology departments. By design, none of the participants had used MedlinePlus before the study.

## 3.2 System Used: MedlinePlus

Launched in 1998, MedlinePlus (www.medlineplus.gov) is a website created by the NLM to provide free and trusted consumer health information to the public. Information resources include publications of medical research sponsored by the National Institutes of Health's (NIH) and publications of professional medical societies and voluntary health agencies without commercial or business motives [17]. Additional resources include licensed medical dictionaries, encyclopedia, and free information such as prescription drugs, directories of health professionals and hospitals, and news feeds from third-party content providers.

During the time of data collection in late 2008, MedlinePlus was a browsing oriented system plus a simple search engine. The left column of the homepage listed major information sources, such as health topics, encyclopedia, drugs & supplements, dictionary, etc. Within each source, information was organized in alphabetical order or by subject. A search box was placed under the title of the homepage. The simple search engine supported word and phrase searches but not the advanced features such as field search or setting filters.

## 3.3 Search Tasks

Participants performed three assigned search tasks in the study. To reflect real information needs, the search questions were adopted from Yahoo! Answers, a social question and answer website where general consumers ask questions and receive answers from peer users. The original questions were slightly modified to suit the scope of MedlinePlus. For each search task, a scenario was added to provide the context as follows:

1. 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.
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. 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.

These three search tasks represent different levels of multiplicity. Task 1 involves finding information on one drug's specific facet: marijuana-medical use. Task 2 involves two health conditions, hypertension and diabetes. Task 3 involves one hormone and two organs, insulin, liver, and kidney. The complexity increases as multiplicity increases from Task 1 to Task 3.

## 3.4 Settings and Procedure

The data collection took place in a private lab. All participants performed the three tasks individually. The computer for data collection was equipped with the Windows Vista operating system. The Internet Explorer (IE) was set as the default Internet browser, with the starting page set to the MedlinePlus homepage. All the participants followed the same following procedure.

Upon arrival, the participant was welcomed and received a brief introduction to the study. The participant then was asked to review and sign an informed consent form. After the consent, the participant finished the ETS VZ-2 paper-folding test [7], which measures individuals' spatial ability. Previous research suggested that spatial ability might affect searching behavior in a hypertext environment [e.g., 6]. After the VZ-2, the participant completed a demographic questionnaire on his/her major field of study, computer experience, and experience with medical information searching. Then, the participant was directed to the testing computer to spend five minutes to explore MedlinePlus in the way he/she preferred.

After the 5-minute free-style exploration, the participant was given the three tasks at once. The order of the three tasks was randomized among participants to balance potential order effect. The participant was instructed to take as much time as he/she wanted to finish the tasks. Upon the completion of each task, the participant rated the difficulty of the task, the mental effort required to accomplish the task, and the satisfaction with his or her search performance, all on a 5-Likert scale. The search session was video-captured using Camtasia software. After the completion of all search tasks, the participant was interviewed about his/her impressions on the tasks, the MedlinePlus website, and the search process. At the end of the session, the participant was thanked and debriefed about the study.

## 3.5 Data Analysis

Three types of data were collected from the participants: demographics and spatial ability, video-recordings of search sessions, and interviews. Both videos recordings and interviews were transcribed for analysis. Each webpage was an analytical unit: when a participant moved to a new webpage, one record was created. The actions executed on the webpage, such as clicking on links and typing in queries, were transcribed and incorporated into the record. The queries were also transcribed. In most sessions, the participants reformulated queries. The behavior of query reformulation was analyzed in terms of the actions that participants performed to change the queries, such as adding a concept, deleting a concept, repeating a concept, and changing to a new concept. In addition, the conceptual changes that resulted from these actions were examined. The analysis adopted both top-down and bottom-up approaches. That is, the coding schema for query reformulation developed by Rieh and Xie [22] (e.g., specification and generalization) was used as the initial framework for coding; at the same time, we expanded the schema to incorporate new types of conceptual changes that emerged from the data. Each query reformulation instance is a unit of analysis for both actions and conceptual changes. Two coders coded all the query reformulation instances with the coding agreement at 96.4%. The discrepancies were solved by discussion.

The participants' interaction strategies are analyzed in terms of searching or browsing. Searching refers to submitting a query to the search box in MedlinePlus website while browsing refers to following the sequential or semantic paths/links, such as alphabetical lists and subject hierarchies. The participants' interaction strategies also include the sequential transitions from searching to searching, searching to browsing, browsing to browsing, or browsing to searching.

The transcripts of the interviews were analyzed using the open coding method. Due to the space limit, the interview results are only briefly reported in this article.

# 4. RESULTS

## 4.1 Characteristics of Participants

One male participant had a negative spatial ability score. The after-search interview revealed that an earlier medical condition affected his spatial ability. Thus his data was excluded from the analysis. The rest 19 participants' spatial ability scores range from 6.8 to 17.6 (*Mean* = 12.42; *SD* = 4.04). Their ages range from 18 to 21 years (*Mean* = 20.37; *SD* = 1.21) and their internet experiences range from 6 to 13 years (*Mean* = 10; *SD* = 2.03).

When it comes to looking for medical or health-related information, the participants referred most frequently to doctors, followed by family and friends. Nevertheless, all participants, except one, had searched for medical information online on a yearly or monthly basis. The sources they used include general web search engines, mostly Google, Wikipedia, WebMD, and health information provided on the university's website. Two participants also had used PubMed.

## 4.2 Session Length and Perceptions of Tasks

The sessions lasted between 11.63 and 29.01 minutes (*Mean* = 19.86; *SD* = 7.20). Table 1 shows the breakdown on each task, as well as their ratings on difficulty of each task, mental efforts to complete the task, and satisfaction with his/her task performance.

**Table 1. Time spent on tasks and perceptions of the tasks**

| Mean (SD) | T (mins) | Difficulty[a] | Men. eff.[b] | Satisf.[c] |
|---|---|---|---|---|
| Task 1 | 5.31 (3.22) | 2.00 (.58) | 2.68 (.67) | 3.89 (.56) |
| Task 2 | 5.59 (4.23) | 2.94 (.97) | 2.84 (.76) | 3.52 (.70) |
| Task 3 | 8.63 (2.36) | 3.63 (.60) | 3.26 (.56) | 2.94 (.77) |

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

The participants spent significantly more time on Task 3 than on either Task 1 ($t(18) = 3.25$; $p<.005$) or Task 2 ($t(18) = 4.56$; $p<.00$). The difference in time spent for Task 1 and Task 2 is not statistically significant.

Paired *t*-tests suggest that the participants experienced differences in task difficulties. Task 3 was significantly more difficult than either Task 1 ($t(18) = 10.40$; $p<.00$) or Task 2 ($t(18) = 2.97$; $p<.008$); Task 2 was significantly more difficult than Task 1 ($t(18) = 3.66$; $p<.002$). Likewise, Task 3 demanded more mental efforts than Task 1 ($t(18) = 3.28$; $p<.004$) or Task 2 ($t(18) = 2.19$; $p<.05$). But the difference in mental efforts for Task 1 and Task 2 is not statistically significant. The participants were least satisfied with their performances on Task 3 and the differences are significant between Task 3 and either Task 1 ($t(18) = 5.30$; $p<.00$) or Task 2 ($t(18) = 2.63$; $p<.02$). The difference between their satisfaction with their performances on Task 1 and Task 2 is not statistically significant.

## 4.3 Search Behavior

The analysis of searching behavior examined four aspects: query features, search terms, query reformulation, and access to results.

### *4.3.1 Query features: number of queries and terms*

A total of 192 queries were executed in the search sessions. Table 2 shows the breakdown of queries and query terms by tasks.

**Table 2: Query features: number of queries and terms**

| | Participants | No. of Q (per P) | No. of T (per Q) |
|---|---|---|---|
| Task 1 | 17 | 32 (1.88) | 66 (2.06) |
| Task 2 | 18 | 58 (3.22) | 176 (3.03) |
| Task 3 | 19 | 102 (5.37) | 238 (2.33) |
| *Total* | | *192* | *480 (2.50)* |

Task 3 had the most queries, which almost twice as many as Task 2 and three times of Task 1. Paired *t*-tests show that the participants issued significantly more queries for Task 2 than for Task 1($t(18) = 2.77$; $p<.02$); for Task 3 than for Task 1 ($t(18) = 4.41$; $p<.00$); for Task 3 than for Task 2 ($t(18) = 2.83$; $p<.02$). Overall, queries were short with the average of 2.5 terms. The queries for Task 1 were the shortest (mean: 2.06) and queries for Task 2 the longest (mean: 3.03). But the differences in query length between tasks are not statistically significant.

### *4.3.2 Search terms*

The query terms can be categorized into three types (Table 3): keywords that convey semantic meanings (meaningful terms), stop words (no semantic meanings), and search operators (AND, OR). These terms, including typos, are shown as submitted with their occurrences in parentheses.

Meaningful terms represent participants' understanding of each task. The majority of these terms were from the task descriptions. Nevertheless, in some occasions, participants used synonyms instead of the terms in the task description. For example, In Task 1, *benefit or benefits* and *pro* were used to search for arguments supporting the use of marijuana for medical purposes. In Task 2, *high blood pressure* substituted hypertension in their searches. In both Task 2 and Task 3, when searching for relations between conditions or organs, participants also used terms *affecting* and *effects*, in addition to terms in the task descriptions (relation and related).

**Table 3. Search terms**

| Type | Task 1 | Task 2 | Task 3 |
|---|---|---|---|
| Key-words | marijuana (30)<br>medical (12)<br>use (3)<br>benefits (2)<br>Marijuana (2)<br>legalization (2)<br>against (1)<br>arguments (1)<br>benefit (1)<br>legalized (1)<br>medicinal (1)<br>purposes (1)<br>pro (1) | diabetes (44)<br>hypertension (30)<br>blood (9)<br>pressure (9)<br>high (8)<br>coronary (5)<br>heart (5)<br>disease (4)<br>type 1 (4)<br>type 2 (4)<br>type II (3)<br>2 (2)<br>smoking (2)<br>Type II (2)<br>type I (2)<br>type ii (2)<br>diabaetes (1)<br>tobacco (1)<br>failure (1)<br>affecting (1)<br>related (1) | insulin (67)<br>liver (61)<br>kidney (31)<br>disease (11)<br>function (6)<br>effects (3)<br>primary (3)<br>role (3)<br>functions (2)<br>kindey (2)<br>diabetes (1)<br>facts (1)<br>insuline (1)<br>insuling (1)<br>organ (1)<br>pancreas (1)<br>problems (1)<br>productin (1)<br>production (1)<br>relation (1)<br>related (1) |
| Stop words | of (5)<br>for (2)<br>the (1) | | of (8); in (4);<br>on (2); to (2);<br>and (5) |
| Operators | | + (3); AND (1)<br>and (15) | + (9); “ ” (1);<br>AND (2) |

It is also found that some participants searched concepts associated to the concepts in the task descriptions. For example, in Task 2, the participants searched on new concepts related to hypertension: *coronary heart disease* and *heart failure* and two searched *smoking* and *tobacco*. In Task 3, some participants searched *pancreas*, *diabetes*, and *organ*. These new concepts and terms seem to come from their personal knowledge and experiences that enabled them to go beyond the given task descriptions. Along the same line, one participant expressed the type of information (*facts*) expected in search results.

The use of stop words, such as *of*, *for*, *in*, and *the* in queries suggests a tendency to use natural language in search. The participants also used Boolean operators, particularly *and*, *AND* and +, to connect query words and one used a quotation mark *“ “ to* search for a phrase. It is worth noting that, although being provided with the task descriptions, some participants still misspelled the words, such as *diabaetes*, *insuling*, *insuline*, *kindey*, and *productin*.

### *4.3.3 Query reformulation*

During a search session, the first query represents a user’s initial understanding of the problem. As his/her understanding of the task develops or mental model of the search system changes, the user reformulates the query to iterate the search. Thus, analyzing query reformulations can reveal the changes of users’ perceptions of the task or mental models of the system during the searching process. The analysis of query reformulations, in this study, focused on (1) the executed actions that modified the queries and (2) the subsequent conceptual changes resulted from these actions. Each pair of queries, $Q_i$ and $Q_{i+1}$, was compared to identify changes.

#### *4.3.3.1 Query reformulation: Actions*

There are three types of actions to modify queries: (1) related to concepts, such as add, delete, repeat a concept, or change to new concept(s) not appearing in the previous query; (2) related to forms of terms such as from singular to plural, (3) related to conceptual relationships, such as Boolean operators. Table 4 reports the query reformulation actions and their frequencies.

Actions in relation to concept changed the semantic building blocks of a query: concepts. Query reformulations in this study demonstrated five different patterns of actions, adding concept(s), deleting concept(s), replacing concept(s), repeating concept(s), and changing to new concept(s) not included in the previous query. Across the three tasks, about 30% of the query reformulations was replacing concept(s) in the previous query with new concept(s). About the same percentage (28.9%) of the query reformulations was adding one or two concepts to the previous query. The third popular action pattern was changing to new concept(s) not appeared in the previous cycle (about 20.7%). The least used pattern was repeating the previous query. Among the three tasks, Task 3 involved the most number of query reformulations (Task 1: 13; Task 2: 39; Task 3: 83) and the query reformulations for Task 3 involved more patterns of action.

Two types of actions in relation to forms were observed: a change to the form of a term (e.g., “*type ii diabetes*” was modified to “*type 2 diabetes*”), and a correction to misspellings in the previous query (e.g “*liver + insuling*” was modified to “*liver + insulin*”).

The change made on Boolean operators is conceptual at the level of relationships among concepts, whether or not done correctly. For example, the query “*liver disease AND insulin*” was modified to “*insulin liver disease*” in which the searcher dropped Boolean *AND* and also changed the order of the words. Only two cases were observed.

#### *4.3.3.2 Query reformulation: Semantic analysis*

Actions in relation to concept often lead to conceptual or semantic changes in queries. We observed the changes fit well to the four categories of conceptual changes in query formulation adopted from Rieh and Xie [22]: (1) Specification (the inquiry becomes more specific), (2) Generalization (the inquiry becomes more general), (3) Parallel movement (the reformulated query has a partial overlap with the previous query; the two queries deal with somewhat different aspects of one concept), and (4) Replacement with synonym (replace the current terms with words with similar meaning). Synonyms are different terms of the same concept, which is considered as query iteration rather than conceptual change in this study (Table 5).

In addition, this study found that, sometimes, the participants came up with a new concept that did not overlap with the concepts in the previous query. This conceptual change is called “switching topic.” Another phenomenon of interest is that the participants re-executed the previous query after having clicked through the results in order to get back to some results. This is called “re-execution.” In some cases, the subsequent query has more than one type of change. For example, a query was modified from "hypertension" to "high blood pressure diabetes." This reformulation involved two changes: (1) hypertension was replaced by the synonym high blood pressure (Replacement with synonym), and (2) the query becomes more specific when diabetes was added (Specification). Both were counted in analysis and the total number of changes was larger than the number of query reformulations.

As shown in Table 5, the most observed query reformation instances were conceptual changes to make the subsequent queries more specific, followed by those to make subsequent queries more general, then those to switch queries to a new topic, and closely tailed by those to make a parallel movement. For query iterations, re-execution occurred more than replacing concepts with synonyms.

### *4.3.4 Accessing and evaluating results*

In MedlinePlus, at the time of this study, the search results were presented as relevance-ranked lists, similar to most web search engines. On the left panel of the result page, beside results of retrieved documents, two filtering options allowed users to refine the results. That is, the searcher is given three different paths at the results page: (1) to access documents from the result list; (2) to limit results to a particular section, such as health topics, external health links, and news; (3) to limit results to a particular topic cluster, which was generated by the system based on the topic being searched.

In this study, eighteen participants clicked on the links to 155 documents directly from the result list. The two filtering options to refine results were used much less: eight participants used the collection option 31 times and six used topic cluster option seven times. Only two participants accessed the articles listed on the second and the third page of the results.

To evaluate results, some participants scanned the lists and selected links to follow based on the keywords in the headings. One commented in the post-session interviews: "*Initially [I] try to find the words that I typed in the heading […].*" When scanning the results, participants also evaluated the results based on various clues, such as who is behind the information and how the content was presented. Nine participants commented in the post-search interviews that the content was *basic*, but *comprehensive, reliable, professional, and current*. But two also pointed out that the information in MedlinePlus could be *biased* as the government provided much of information.

## 4.4 Browsing Behavior

Browsing, in this study, is specifically referred to the participants' behavior of following the structure of the information space laid out by certain organization schemas, such as alphabetical lists and subject hierarchies. The browsing strategy was used when the participants accessed a particular information resource type, such as encyclopedia, or collections of health topics.

### *4.4.1 Accessing different resources*

In MedlinePlus, entries in both the Drugs & Supplements and Encyclopedia sections were organized in alphabetical order. Articles in the News section, however, were ordered by both date and topic. For the Health Topics section, the next level schemas included: (1) alphabetical list, (2) body location/systems, (3) diagnosis and therapy, (4) disorders and conditions, (5) demographic groups, and (6) health and wellness. The user needed to select the appropriate schema to browse the topic of interest.

Across the three tasks, six participants accessed Drugs & Supplements and 12 accessed Encyclopedia through the alphabetical lists. Only one participant accessed the News by topic (marijuana). The participants' usage of the six schemas to access collections of Health Topics is detailed in Table 6. As shown in the table, thirteen participants (68.4%) accessed health topics for 44 times by navigating through these information organization schemas. The majority of these participants (11 out of 13) were accessed the schemas for Task 3. Correspondingly, the majority of the visits (30 out of 44) were for Task 3. Among these schemas, only four schemas were used. The most used one was the *Alphabetical list* with a total of 22 visits by 7 participants, accounting for 50% of the usage of all schemas. The *Body Location/System*, was used by 12 participants totaling 19 visits, accounting for about 43% of the usage. Both *Diagnosis and Therapy*, and *Health and Wellness* were used by one or two participants with one visit by each. Two schemas, *Disorders and Conditions*, and *Demographic Groups* were not used (thus not displayed in the table).

### *4.4.2 Accessing related topics*

In MedlinePlus, each health topic page contains a summary of the topic, in which related organs, conditions, or diseases are often mentioned. The page provides both in-text and related topics hyperlinks. Users can follow either the in-text links or related topics and categories links for additional resources. Table 7 shows participants' usage of these two means to access related topics.

**Table 4. Query reformulation: actions and frequencies**

| Actions | Type | Task 1 | Task 2 | Task 3 | *Total* |
|---|---|---|---|---|---|
| Related to concept | Add (A) | A (4); AA (1) | A (10) | A (20); AA (4) | 39 (28.9%) |
| | Delete (D) | D (2) | D (5) | D (6), DD (2) | 16 (11.9%) |
| | Replace (D…, A…) | DA (4); DDA (1) | DA (12); DDA (2); DAA (2) | DA (14); DDA(3); DDDAA (1); DAAA (1) | 40 (29.6%) |
| | Repeat | 1 | 3 | 9 | 13 (9.6%) |
| | Change to new concept(s) | | 5 | 23 | 28 (20.7%) |
| | *Total* | *13* | *39* | *83* | *135* |
| Related to form | Change the form of a term | | 2 | | |
| | Correct spellings | | | 1 | |
| Related to relationship | Change Boolean operators | | 2 | 2 | |

**Table 5. Query reformulation instances: semantic changes**

| | Task 1 | Task 2 | Task 3 | *Example* | *Total* |
|---|---|---|---|---|---|
| ***Conceptual reformulations*** | | | | | |
| Specification | 5 | 18 | 26 | hypertension and diabetes → hypertension and type 1 diabetes | 49 (35%) |
| Generalization | 3 | 9 | 16 | legalized marijuana → marijuana | 28 (20%) |
| Switching topic | | 3 | 19 | diabetes liver → insulin | 22 (15.7%) |
| Parallel movement | 3 | 4 | 13 | kidney insulin → liver insulin | 20 (14.3%) |
| ***Query iterations*** | | | | | |
| Re-execution | 3 | 3 | 9 | liver→ liver | 15 (10.7%) |
| Replacement with synonym | 1 | 5 | | diabetes hypertension → diabetes and high blood pressure | 6 (4.3%) |
| *Total* | *15* | *42* | *83* | | *140* |

**Table 6. Browsing to access a particular health topic**

| | No. of participants | | | | No. of visits | | | |
|---|---|---|---|---|---|---|---|---|
| | Task1 | Task2 | Task3 | Total | Task1 | Task2 | Task3 | Total |
| Alphabetical list | 4 | 3 | 5 | 7 | 4 | 4 | 14 | 22 (50.0%) |
| Body Location/System | | 3 | 10 | 12 | | 3 | 16 | 19 (43.2%) |
| Diagnosis and Therapy | 2 | | | 2 | 2 | | | 2 (4.5%) |
| Health and Wellness | 1 | | | 1 | 1 | | | 1 (2.3%) |
| *Total* | *5* | *5* | *11* | *13* | *7* | *7* | *30* | *44* |

**Table 7. Browsing to access related topics**

| | No. of participants | | | | No. of visits | | | |
|---|---|---|---|---|---|---|---|---|
| | Task1 | Task2 | Task3 | Total | Task1 | Task2 | Task3 | Total |
| In-text links in summary | 1 | 2 | 4 | 6 | 1 | 3 | 5 | 9 |
| “related topics” links | | 1 | 5 | 5 | | 1 | 8 | 9 |
| *Total* | *1* | *3* | *7* | ***9*** | *1* | *4* | *13* | ***18*** |

**Table 8. Patterns of use of searching and browsing strategies in completing search tasks**

| | Task 1 | Task 2 | Task 3 |
|---|---|---|---|
| Searching (S) | S (9) | S (10) | S (4) |
| Browsing (B) | B(HT) (2) | B(HT) (1) | |
| Sequential pattern of S and B (Abbreviations see footnotes) | B(DS) → S (2)<br>B(DS) → B(HT) →S<br>B(N) → S<br><br>S→ B(DS) →S (2)<br>S→ B(E) → S<br>S→ B(E) → B(HT) → S | B(E)→ S (2)<br>B(HT)→S (2)<br>B(E)→ S →S(D)→S<br>B(E)→ S→ B(HT)<br><br>S→ B(E) →S<br>S→ B(HT) →S | B(E) →S<br>B(HT) →S<br>B(HT) →S →B(HT)<br>B(HT) →S →S(D) →B(E) →S →B(RT) →B(RT)<br>B(HT) → S(D) →S →B(HT) → S → B(E) →S<br>B(HT) →B(E) → B(HT) → S →B(HT) →S(D) → B(HT) →S →B(HT) →B(E)<br><br>S(D) →S →B(HT)<br>S →B(HT) → S (2)<br>S →B(E) →S →B(HT)<br>S →B(RT) →S<br>S → B(E) →S(D) →B(E)<br>S →B(RT) → B(E) →S<br>S → B(HT) →B(DS) →B(HT) →B(E) → B(HT)<br>S → B(E) →S(D) →B(HT) →S → B(E) → B(HT) → S → B(E) →S |
| ***Total*** | ***19*** | ***19*** | ***19*** |

**Abbreviations**:
D for Dictionary, e.g. S(D) : Searching the dictionary
DS for Drugs and Supplements, e.g. B(DS): Browsing in the Drugs & Supplements section
E for Encyclopedia, e.g. B(E): Browsing in Encyclopedia
HT for Health topic, e.g. B(HT)Browsing by health topic
N for News; B(N), e.g. Browsing in the news section
RT for Related Topics, e.g., B(RT): Browsing by following related topics provided by MedlinePlus

As shown in the table, these hyperlinks were not used often: less than half of the participants (9 out of 19) looked for information by tracing related topics links, and the average usage was twice. Both the in-text links in topic summary and related topic links on the page were equally used, mainly for Task 3.

## 4.5 A Holistic View of Search Strategies

Generally speaking, MedlinePlus is a typical information-rich web space. It provides various types information in various formats and has an information architecture that supports both searching and browsing. Therefore, information search strategies in MedlinePlus include not only the deployment of searching and browsing techniques, but also the selection and use of different types of resources. This section provides a holistic view of the strategies that users use in searching for information in MedlinePlus from two aspects: use of resources and use of searching and browsing techniques.

### 4.5.1 Use of resources

As described in 3.2, MedlinePlus provides rich information organized by seven major types of resources with different methods to support easy access. The Health Topics section was the most accessed resource; all participants accessed at least one health topic page via browsing or selecting a link in search results. The second most used source was the medical encyclopedia; thirteen participants accessed it by directly navigating to the source or by clicking the link in search results. This source was used mainly for Task 2 and Task 3. Dictionary was accessed by 6 participants, mostly in their search for Task 3. The Drugs & Supplements section was accessed by four participants, three of whom used it for Task 1. Fewer participants accessed the News, Directories, Go Local, or Multiple Languages sections of MedlinePlus.

### 4.5.2 Use of searching and browsing strategies

As described in the previous sections, participants used both the searching and browsing strategies to accomplish the tasks. In terms of searching, in addition to the main site-wide search (S), participants also searched the Dictionary (D). In terms of browsing (B), they browsed to access information included in Health Topics (HT), Encyclopedia (E), Drugs & Supplements (DS), News (N), and Related Topics (RT). During the search sessions, these two strategies were used in combination to their benefits. Table 8 shows a paramount view of the deployment of the two strategies in each search session for each task by demonstrating the sequences in which they were used. The numbers in the parentheses are the number of participants who adopted the sequence.

As shown in the table, for Task 1 and Task 2, about half of the participants used only the searching strategy throughout the session (9 participants for Task 1 and 10 for Task 2), and a couple of them used only browsing (2 participants for Task 1 and 1 for Task 2). For Task 3, four participants (21.1%) used only searching and none of them was solely dependent on browsing to complete the task.

For both Task 1 and Task 2, eight participants (42.1%) used searching and browsing in combination and for Task 3, fifteen participants (78.9%) used the two strategies in combination. The patterns in which searching or browsing was deployed varied among the tasks. For Task 1, four participants started by browsing for the marijuana entry in Drugs & Supplements or in News, with one continuing to browse Health Topics. Nevertheless, all of them ended up using the main search function to search for information. The other four participants started with searching and moved onto browsing Drugs & Supplements, Encyclopedia, or Health Topics. These four participants also eventually searched for information. For Task 2, six participants started with browsing the Encyclopedia or Health Topics and then switched to searching, and one ended the session by going back to browse (Health Topics). Two participants started with searching, switched to browsing Encyclopedia and Health Topics, and ended the session by searching.

For Task 3, six participants started with browsing Encyclopedia or Health Topics and nine started with searching. Compared to those of Task 1 and Task 2, however, the patterns of Task 3 were more diverse, complex, exploratory, and iterative. For example, one participant performed searching at the beginning, and switched to browsing sequentially Health Topics first, then Drugs & Supplements, again Health Topics, then Encyclopedia, and finally Health Topics (S → B(HT) →B(DS) →B(HT) →B(E) → B(HT)).

## 5. DISCUSSION

Three research questions were introduced to guide this study of finding information in MedlinePlus: (1) How users search; (2) How users browse; (3) How tasks affect interaction strategy. The results were presented in the previous section according to these questions. This discussion will begin with the third question. The participants applied both browsing and searching strategies to complete three tasks of increased complexity as reflected by conceptual multiplicity. We observed that for tasks with relatively lower multiplicity (one or two concepts), single strategy, either browsing or searching alone, was used to complete the tasks; and searching was a dominant choice. When both searching and browsing were used in lower multiplicity questions, the interaction sequences were short and the last interaction before ending the session was searching except for one case. The previous studies also find that the majority of users primarily relied on the searching strategy to find information [e.g., 26].

As a contrast, for the high multiplicity task (Task 3 with three concepts), no participant used browsing alone; the majority searched long sequences of combined browsing and searching. The observed differences suggest that task complexity influences interaction strategies and patterns. It also must be noted: although the search engine in MedlinePlus at the time of data collection was very basic without advanced features (see 3.2), the participants relied more on searching than browsing. One explanation is that browsing a Web space to find information requires that the user is familiar with the system's information organizational structure or that the user's conceptual structure matches closely the system's organizational structure. In the situation of given descriptions of questions, the effort is much less for users to enter search terms than to understand the classification underlying navigational paths. On the other hand, the participants often used browsing strategies to access the encyclopedia and dictionary to understand the concept and health topic in order to have a preliminary understanding of the questions before searching.

The first research question was about searching behaviors. We observed that the participants tended to draw query terms from the task descriptions. In some cases, they also came up with terms associated with the terms used in task descriptions. The type of associations included conditions and organs. Similar to other studies [25, 32], the average query length in this study was 2-3 terms. The participants misspelled several common words in queries. This suggests that a function to detect misspellings and make suggestions is needed.

It has been recognized that formulating a search query is a challenging task, particularly for exploratory tasks [3, 8]. We observed that the mean session length was between 1.88 to 5.37 queries corresponding to complexity levels of the tasks. This is similar to HealthLink search session length, 2.34 queries [34]. Five actions were observed in query reformulations: to add concept(s), to delete concept(s), to replace (delete and add) concepts, to repeat concept(s), and to change to new concept(s). Replacing concepts occurred most, followed by adding concept(s) and changing to new concept(s). These actions led to two types of query reformulation: conceptual change and query iteration. The former resulted in a change to the meaning of the previous query; the latter the meaning remained the same as the previous query. Our observation extended the schema proposed by Reih and Xie [22] with a new type of query reformation: switching topic.

We observed that 85% of the query reformulations occurred as conceptual changes: specification was the most frequent reformation, followed by generalization, switching topic, and parallel movement. The popularity of query specification and generalization (64.7% of the conceptual changes) indicates that the participants had difficulties in determining the level of specificity of the information needed and the query reformulations were trial-and-error processes [26]. It also suggests that systems could provide a hierarchical terminology structure to help users in selecting query terms. The other two conceptual changes, switching topic and parallel movement, accounted for about 35.3% of the conceptual changes. The occurrence of these two changes might be due to the exploratory nature of the tasks. The tasks in the study involved multiple concepts or multiple facets of a concept, and MedlinePlus was limited in supporting the exploration of relationships between multiple conditions or health concepts, as commented by 7 participants in the post-search interviews. For example,

> It is pretty easy to find information on one disease, [but] it is hard to find information on articles that relate to different topics. […] It is harder to find information about how certain symptoms or organs are affected by other diseases, that's a little tricky.

Another lamented the lack of advanced search functions:

> It does not have any ability, like some websites, to say this, but not this, and this. [Although] you can still narrow it down by what kind of information you want, it is just more difficult to make relationships between like the kidney and liver.

Query iteration took two forms: re-executing the previous query (10.7%) and replacing terms with synonym(s) (4.3%). Why do participants re-execute a query? Like most websites, MedlinePlus at that time could not support searchers to backtrack to a specific previous search point. As a coping mechanism, the participants resubmitted the same query in order to return to these results. Using synonyms to reform a query was infrequent (4.3%), most of which substituted a medical term, e.g., “hypertension” with its lay-term counterpart, e.g., “high blood pressure” or vice versa. Our observation is consistent with the findings of the previous studies that general public have difficulties with medical terminologies [e.g., 32]

Literature suggests that access to search results has been largely ignored in user-centered IR research [20]. Our findings show that most participants accessed results by simply clicking on the links in the results; only a few used the functions provided by MedlinePlus to limit the results to a particular type or subject. It must be noted that some participants actually evaluated the quality of the results using certain criteria such as the source of the information. Previous studies either reported such behavior [e.g., 24, 26] or lack of such behavior [e.g., 11, 30]. Consistent with other studies [11], fewer participants in this study accessed links beyond the first result page.

Our second question was about browsing behavior. MedlinePlus organizes information by categories or alphabetic lists. Although searching was a dominant interaction behavior for less complex tasks, browsing was frequently used within a sequence of both interaction strategies. That is, participants switched between searching and browsing. The most used schema in Health Topics page was the alphabetical list, followed by the body location/systems. The use of these two schemas is not surprising because the task descriptions provided terms for which a list comes handy. Two of the questions involved organs that were easily located in body anatomy. However, the two paths leading to related topics (as related topics list or in-text links of summary) were not used much.

We observed that when performing the high complex task, the participants were more likely to browse the Encyclopedia and Health Topics, or follow links to related topics in addition to searching.

There are some limitations of the study. First, the participants were undergraduate students. This user group tends to have more versed skills in web searching but less often on health-related topics. Second, the tasks were predefined rather than users' real needs. The behavior reported in this study should not be generalized to the general public. Further studies should recruit users with real health information needs. The information architecture and interface of a web apace can influence interaction behavior, thus, it is valuable to observe the same user group using other consumer health information websites, particularly those with different structures to MedlinePlus.

# 6. CONCLUSIONS

Although MedlinePlus has strong support of browsing by categories and topics, participants in this study mostly used searching strategies to find information. Searching was used more for the task involving three concepts than the tasks involving one or two concepts. Some of our findings corroborate the reported findings of studies that observed general search engines and other websites. Although reformulation of queries is necessary for better results, most reformulations were moves along a conceptual hierarchy (generalization or specification). Despite the small percentage of occurrences, both synonyms and re-execution of previous queries should be improved by better design. Synonyms should be expanded automatically upon request. Search history should be accessible to allow backtracking to a specific point of the earlier search results.

Browsing strategy was mainly used in situations where the participants were familiar with the structure of the source, such as encyclopedia and dictionary, or the terms given in task description. However, most lay people will not be able to navigate through MedlinePlus's health topics due to a lack of medical conceptual structure. To reduce such difficulty, the system should design mechanisms to connect lay terms with medical terms.

One of the implications of the influences of search task complexity is to design cognitive assistance to help users at different stages of a search. For query construction, a display of terminological relationship can be useful; for query reformulation, soliciting reasons can suggest alternative moves. Other needed improvements include adding advanced search features, making information architecture visible to users, and encouraging evaluation of search results.

In light of the findings and the limitations of the study, we suggest some research questions for further investigation: Are the functions “related topics” and “Refined by …” in MedlinePlus underused? If so, why? How do different attributes of tasks affect searching behavior? How do real users interact with MedlinePlus to find information for their own topics? In addition, comparisons of MedlinePlus with other health information websites such as WebMD are also needed to better understand health information systems.

## 7. ACKNOWLEDGMENTS

The authors wish to thank our participants for their contributions to the study and thank Anthony Domina and Shannon Daily for their assistance with data analysis.


## 8. REFERENCES


[1] Arora, N.K., Hesse, B. W., Rimer, B.K., Viswanath, K., Clayman, M.L., and Croyle, R.T. 2007. Frustrated and confused: the American public rates its cancer-related information-seeking experiences. Journal of General Internal Medicine, 23(3), 223-228.

[2] Bates, M.J. 1979. Information search tactics. Journal of the American Society for Information Science, 30, 205-214.

[3] Belkin, N., J., Oddy, R. N., and Brooks, H. M. 1982. Ask for information retrieval: Part I. background and theory. Journal of Documentation, 38, 61-71.

[4] Boden, C. 2009. Overcoming the linguistic divide: a barrier to consumer health information. Journal of the Canadian Health Libraries Association, 30, 75-80.

[5] Capra, R., Marchionini, G., Velaso, J., and Muller, K. 2010. Tools-at-hand and learning in multi-session, collaborative search. In the Proceedings of CHI 2010, 951-960.

[6] Dillon, A. 2000. Spatial-semantics: How users derive shape from information space. Journal of the American Society for Information Science, 51(6): 521-528.

[7] Eckstrom, R. B., French, J. W., Harman, H. H., and Derman, D. 1976. Kit of Factor-referenced Cognitive Tests. Princeton, N.J.: Educational Testing Service.

[8] Efthimiadis, E.N. 1996. Query expansion. In M.E.Williams (Ed.). Annual Review of Information Systems and Technology, 31, 121-187. Medford, NJ: Information Today.

[9] Efthimiadis, E.N. 2009. How students search for consumer health information on the web. In Proceedings of the 42nd Hawaii International Conference on System Sciences, 1-8.

[10] Eysenbach, G. 2000. Consumer health informatics. BMJ, 320, 1713-1716.

[11] Eysenbach, G., and Kohler, C. 2002. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ, 324, 573-577.

[12] Fidel, R. 1985. Moves in online search. *Online Review, 9*(1), 61-74.

[13] Fox, S., and Jones, S. 2009. The Social Life of Health Information. Pew Internet & American Life project. Retrieved March 3, 2010, from http://www.pewinternet.org

[14] Hong, Y., Cruz. N., Marnas, G., Early, E., and Gillis, R. 2002. A query analysis of consumer health information retrieval. In the Proceedings of AMIA 2002, 1046.

[15] Keselman, A., Browne, A.C., and Kaufman, D.R. 2008. Consumer health information seeking as hypothesis testing. JAMIA, 15, 484-495.

[16] Kim, D., and Chang, H. 2007. Key functional characteristics in designing and operating health information websites for user satisfaction: An application of the extended technology acceptance model. International Journal of Medical Informatics, 76, 790-800.

[17] Lindberg, D. A. B. 2000. Internet access to the national library of medicine. Effective Clinical Practice, 3(5), 256-260.

[18] Luo, G., Tang, C., Yang, H., and Wei, X. 2008. MedSearch : A specialized search engine for medical information retrieval. In Proceedings of CIKM'08, 143-152.

[19] Marill, J. L., Miller, N., and Kitendaugh, P. 2006. The MedlinePlus public user interface: Studies of design challenges and opportunities. Journal of Medical Library Association, 94(1), 30-40.

[20] Marchionini, G., Capra, R., and Shah, C. 2008. Focus on results: personal and group information seeking over time. Poster at Workshop on HCIR 2008. Seattle, WA: October 23, 2008.

[21] McCray, A.T., and Tse, T. 2003. Understanding search failures in consumer health information systems. In the *Proceedings of the AMIA Symposium 2003*, 430-434.

[22] Rieh, S.Y., and Xie, H. 2006. Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information Processing and Management, 42*, 751-768.

[23] Shneiderman, B. 1994. Dynamic queries for visual information seeking. *IEEE Software, 11*(6), 70-77.

[24] Sillence, E., Briggs, P., Fishwick, L., and Harris, P. 2004. Trust and mistrust of online health sites. In the *Proceedings of CHI 2004*, 663-670.

[25] Spink, A., Yang, Y., Jansen, J., Nykanen, P., Lorence, D. P., Ozmutlu, S., and Ozmutlu, H.C. 2004. A study of medical and health queries to web search engines. *Health Information and Libraries Journal, 21*, 44-51.

[26] Tomes, E., and Latter C. 2007. How consumers search for health information. *Health Informatics Journal, 13*(3), 223-235.

[27] Wang, P. 2006. Modeling web searching behaviors and designing new effective interactions for digital libraries. Interim report 2. IMLS Grant: LG-06-05-0100-05 (http://web.utk.edu/~peilingw/pub/IMLSreport_2.pdf)

[28] White, R.W., Dumais, S., and Teevan, J. 2008. How medical expertise influences web search interaction. In *Proceedings of SIGIR'08*, 791-792.

[29] Wildemuth, B.M. 2004. The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology, 55*(3), 246-258.

[30] Williams, P., Nicholas, D., Huntington, P., and McLean, F. 2002. Surfing for health: user evaluation of a health information website. Part two: fieldwork. *Health Information and Libraries Journal, 19*, 214-225.

[31] Zeng, Q.T., Kogan, S., Ash, N., Greenes, R.A., and Boxwala, A.A. 2002. Characteristics of consumer terminology for health information retrieval. *Methods of Information in Medicine, 41*, 289-298.

[32] Zeng, Q.T., Kogan, S., Plovnick, R.M., Crowell, J., Lacroix, E.M., and Greenes, R.A. 2004. Positive attitudes and failed queries: an exploration of the conundrums of health information retrieval. *International Journal of Medical Informatics, 73*(1), 45-55.

[33] Zeng, Q.T., Crowell, J., Plovnick, R.M., Kim, E., Ngo, L., and Dibble, E. 2006. Assisting consumer health information retrieval with query recommendations. *Journal of American Medical Informatics Association, 13*, 80-90.

[34] Zhang, J., Wolfram, D., Wang, P., Hong, Y., and Gills, R. 2008. Visualization of health subject term analysis based on user queries. *Journal of the American Society for Information Science and Technology*, 59(12): 1933-1947.

[35] Zhang, Y. 2009. The construction of mental models of information-rich web spaces: The development process and the impact of task complexity. PhD. Dissertation, University of North Carolina at Chapel Hill.

[36] Zickuhr, K. 2010. *Generations 2010*. Pew Internet & American Life project. Retrieved June 20, 2011, from http://www.pewinternet.org/~/media//Files/Reports/2010/PIP_Generations_and_Tech10.pdf