Strategy Hubs: Domain Portals to Help Find Comprehensive Information

Suresh K. Bhavnani
School of Information, University of Michigan, Ann Arbor, MI 48109-1092. E-mail: bhavnani@umich.edu

Christopher K. Bichakjian
Department of Dermatology, University of Michigan Medical School and Comprehensive Cancer Center, Ann Arbor, MI 48109-0314. E-mail: chriskb@med.umich.edu

Timothy M. Johnson
Departments of Dermatology, Otolaryngology, and Surgery (Section of Plastic Surgery), University of Michigan Medical School and Comprehensive Cancer Center, Ann Arbor, MI 48109-0314. E-mail: timjohn@med.umich.edu

Roderick J. Little
Department of Biostatistics,University of Michigan, Ann Arbor, MI 48109-2029. E-mail: rlittle@umich.edu

Frederick A. Peck
School of Information, University of Michigan, Ann Arbor, MI 48109-1092. E-mail: peckf@umich.edu

Jennifer L. Schwartz
Department of Dermatology, University of Michigan Medical School and Comprehensive Cancer Center, Ann Arbor, MI 48109-0314. E-mail: jennschw@med.umich.edu

Victor J. Strecher
Department of Health Behavior and Health Education, School of Public Health, University of Michigan, Ann Arbor, MI 48109-0471. E-mail: strecher@umich.edu

Recent studies suggest that the wide variability in type, detail, and reliability of online information motivate expert searchers to develop *procedural search knowledge*. In contrast to prior research that has focused on *finding* relevant sources, procedural search knowledge focuses on how to *order* multiple relevant sources with the goal of retrieving comprehensive information. Because such procedural search knowledge is neither spontaneously inferred from the results of search engines, nor from the categories provided by domain-specific portals, the lack of such knowledge leads most novice searchers to retrieve incomplete information. In domains like healthcare, such incomplete information can lead to dangerous consequences. To address the above problem, a new kind of domain portal called a *Strategy Hub* was developed and tested. Strategy Hubs provide critical search procedures and associated high-quality links to enable users to find comprehensive and accurate information. We begin by describing how we collaborated with physicians to systematically identify generalizable search procedures to find comprehensive information about a disease, and how these search procedures were made available through the Strategy Hub. A controlled experiment suggests that this approach can improve the ability of novice searchers in finding comprehensive and accurate information, when compared to general-purpose search engines and domain-specific portals. We conclude with insights on how to refine and automate the Strategy Hub design, with the ultimate goal of helping users find more comprehensive information when searching in unfamiliar domains.

Received July 12, 2004; revised September 27, 2004; accepted September 29, 2004

 • Published online 11 October 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20238

Introduction

Numerous studies have attempted to analyze and identify the strategic knowledge acquired by expert searchers. These studies include the identification of strategies through experience (e.g., Bates, 1979; Drabenstott, 2000), through theoretical analysis (Belkin, 1995), through systematic observations of experts performing complex tasks (e.g., Fidel, 1991; O'Day & Jeffries, 1993; Xie, 2000), and through expert–novice comparisons to understand differences in search knowledge (e.g., Holscher & Strube, 2000; Hsieh-Yee, 1993; Lazonder, Biemans, Wopereis, 2000; Sutcliffe, Ennis, & Watkinson, 2000; Shute & Smith, 1993; Wildemuth, do Bliek, Friedman, & File, 1995). Such studies have shed light on the numerous and complex strategies useful in rapidly finding relevant sources of information.

An important focus of the above research has been to *find* relevant sources of information. However, the Web presents an extremely heterogeneous information environment, where sources vary widely along many dimensions including reliability, level of detail, and genre. Searching for information about a topic in domains such as healthcare typically requires more than knowledge of how to find relevant sources of information.

Consider the task of finding comprehensive information[1] about a search topic such as "Treatment options for Stage III melanoma." Getting comprehensive information for such a topic is difficult because there are many reliable and unreliable sources for such information, and to make matters worse, no single source contains all the information. For example, while the National Cancer Institute site provides primary treatment information about melanoma, supplemental treatments such as interferon for melanoma are described in other sources like the University of Michigan's Cancer site. To get comprehensive information about treatments for Stage III melanoma, users must therefore first have the subgoal to retrieve *primary* treatment information about the disease, followed by a second subgoal to look for *supplemental* treatments in a specialized source. Both subgoals are needed to obtain comprehensive information about treatments for Stage III melanoma.

Ordered subgoals, such as those described above, are neither spontaneously obvious from a list of relevant hits provided by Google, nor from the coarse-grained taxonomies in domain portals such as MEDLINEplus, a leading healthcare portal used by search experts (Bhavnani, 2001). For example, the melanoma page in MEDLINEplus provides three links under the heading "Treatment," none of which point to supplemental treatments. As a result, users often retrieve incomplete information (Bhavnani et al., 2003) when using general-purpose search engines or domain portals like MEDLINEplus. In critical domains such as healthcare, such searches can have dangerous consequences.

[1]Comprehensive information about a search topic contains all facts (e.g., claims or recommendations) considered important, for that topic, by experts in the field.

The above example demonstrates that the heterogeneity of the Web now requires users to know more than how to find relevant sources of information. In addition to finding relevant sources, users must also know which sources to visit in which order, particularly for search questions that require comprehensive information of a search topic. In comparison to research focusing on how to *find* sources of information, far less is known on the *order* to visit relevant sources of information once they have been found. Our goal in this article is to shed light on the rationale, nature, and generalization of such search procedures, and how they can be made available to users on the Web in a new kind of domain portal called a *Strategy Hub*.

We begin by describing the factors that make search procedures critical to find comprehensive information about a topic. We then describe how we developed and used an empirically based taxonomy of skin cancer questions to identify expert search procedures to find comprehensive information about melanoma, a deadly form of skin cancer. Next, we describe the design of the Strategy Hub to provide these search procedures, and an experiment to compare its performance with conventional search tools. We conclude with ideas of how to improve the design of the Strategy Hub, and ideas for automating some of the emergent ideas.

Prior Research on Ordering Information Sources

While there have been many researchers who have focused on identifying the strategic knowledge for finding relevant information, we have found relatively fewer studies that have focused on the knowledge to order sources of information.

Evidence on How Search Experts Order Information Sources

One of the earliest attempts at defining general search procedures was to guide library users to visit different types of sources in a particular order (Kirk, 1974). The author through self-reflection recommended that students who were working on their first scholarly paper in an unfamiliar subject domain should visit available sources in the following steps: (a) search a general source such as an encyclopedia to retrieve important references in the subject area, (b) search the card catalog to retrieve the important references, (c) search the retrieved references for keywords of interest, and (d) use keywords to search specific sources like the *Biological Abstracts* to find other targeted papers of interest. The goal of the above procedure (going progressively from general to more-specific sources) was intended to help students become more efficient and effective in their use of the library by using different classes of sources in a suggested sequence.

The importance of visiting sources in a particular sequence has been shown through two empirical studies of expert searchers. Florance and Marchionini (1995) reported that given a set of medical articles and a medical question, physicians either used an *additive* or *recursive* strategy to visit the sources. In the additive strategy, the physicians moved sequentially through a stack of articles by retrieving individual

DOI: 10.1002/asi

facts from each article. In contrast, the physicians used a recursive strategy where they moved back and forth among the articles, for example, to answer questions raised in one article, by referring to sections in another article. Although the study was inconclusive as to when each type of strategy was used, the authors suggest that the strategies were adaptations to the relevance of the articles: When the articles were very relevant (such as review articles), then the additive strategy was sufficient as the users could collect the information in the order presented. However, when the articles were overall only marginally relevant (such as when each article contained only pieces of information relevant to the task), then the users had to use the recursive strategy which involved revisiting pages to enable comprehension of the different pieces. This study suggests that users might benefit by new approaches that reduce such back and forth movement between sources, especially when information is scattered across those sources.

A more recent empirical study focused on the Web observed how domain experts searched within and outside their domains of expertise (Bhavnani, 2001). The study revealed that when the domain experts searched within their domains of expertise, they had recognizable *search procedures*. For example, the study identified the three-step search procedure followed by an expert healthcare searcher looking for flu shot information:

1. Access a reliable general-purpose healthcare portal to identify sources for flu-shot information.
2. Access high-quality sources of information to retrieve general flu-shot information.
3. Access a specific pharmaceutical Web site that sells a flu vaccine to verify the information.

Such search procedures enabled the domain experts to find comprehensive information quickly and effectively for the task within their domain of expertise, compared to when they performed a task outside their domains of expertise, where they relied on the order of hits provided by the search engine, Google (Bhavnani, 2001).

The above study suggests that users who have acquired search expertise in a particular domain know more than just how to use query-based search engines, and the names of high-quality sources; they have also acquired the *procedural search knowledge* to determine which sites to visit in which order when searching for comprehensive information. This procedural search knowledge typically consists of three components:

1. The subgoals to organize a search in a particular domain. For example, the healthcare search expert knew the critical subgoal of *verifying* healthcare information by visiting a pharmaceutical source.
2. The order in which to satisfy those subgoals. For example, the expert knew to first visit a general domain portal to get broad and general information, before visiting a specific pharmaceutical company to verify that information.
3. The selection knowledge to decide which sites or pages will satisfy a particular subgoal, such as to visit MEDLINEplus to obtain reliable healthcare sources.

Attempts to Make Procedural Search Knowledge Explicit to Users

The notion of providing procedural knowledge to access existing sources of information dates back to Bush (1947). In his seminal paper, Bush sketched out a system called a *memex*, where users could store large amounts of information, and provide annotated *trails* through that information for different tasks. Such trails would provide reusable and sharable paths through existing sources of information. The goal of such a system was to go beyond indexing information, and attempt to capture the associative nature of human knowledge by making semantic connections between chunks of information to perform different tasks.

Several researchers interested in improving the use of hypertext systems have attempted to implement the above idea of a memex. For example, Zellweger (1987, 1988, 1989) built a system that allowed users to specify simple or complex branching "directed paths," Halaz, Moran, and Trigg (1987) used "guided tours" to guide a reader's traversal along a path of existing sources, and Shipman, Furuta, Brenner, Chung, and Hsieh, (2000) developed "Walden Paths" to help teachers order relevant Web sites for classroom instruction. The above research[2] attempts have all focused on providing authors a way to specify a metastructure in the form of a path that is independent of the sources being connected.

While the above studies have drawn attention for the need to make procedural search knowledge explicit in a hypertext environment such as the Web, none of them have focused on the nature of the search procedures themselves, how they generalize, and how they affect the search behavior of users in comparison to general-purpose search engines and domain portals. The goal of our research was not to build an authoring tool to enable users to input their search procedures (as Bush suggested), but rather to identify search procedures from experts, and explore how such search procedures could be made available on domain portals.

But what makes the study of procedural search knowledge so important on the Web? Despite the use of modern search engines and extensive domain portals, what are the factors that make procedural search knowledge to visit *many* different sources in a particular *order* so important?

Importance of Procedural Search Knowledge

Our research has identified three reasons that have made procedural search knowledge important, especially for users searching for comprehensive information in unfamiliar domains.

[2]The above research is different from the notion of *Library Pathfinders,* which are an "organized introductory checklist of various types of English language sources of information on a specific topic" (p. 292, Canfield, 1972). Library Pathfinders are therefore similar to categories of links currently provided by most domain portals such as MEDLINEplus.

DOI: 10.1002/asi

Information Scatter

While it has been well known that relevant articles for most topics are scattered across a wide number of journals (Bradford, 1948), and databases (Hood & Wilson, 2001) little is known about how facts are scattered across Web pages. In a recent study, we analyzed how facts about a healthcare topic were distributed *across* high-quality healthcare sites (Bhavnani, 2005). We first asked two physicians to identify facts (e.g. blue eyes increases your chance of getting melanoma) that they believed to be important for a comprehensive understanding by patients of five common melanoma topics (e.g. melanoma risk and prevention). We then analyzed how those facts were distributed across the top-10 healthcare sites with melanoma information.

The study revealed that for each of the five melanoma topics, the distribution of facts across the relevant pages was skewed towards few facts, with no single page or single Web site that provided all the facts. The above study shed light on the complex environment often encountered when searching for comprehensive information: Searchers must often visit a combination of pages and Web sites to find all relevant facts about a topic.

While the scatter of information across sites motivates users to visit many *different* pages and Web sites, the scatter often leads to a suggested *ordering* of sources when the information they contain has a prerequisite association. For example, if the definition of a term like *Stage III melanoma* is on page X, while its treatment is on page Y, then such scatter suggests that users would benefit by first visiting page X, before visiting page Y so that they understand the full rationale behind the treatment.

Information Density

While the above study probed the distribution of facts, a subsequent study explored the density of information within those pages (Bhavnani, 2005). The results suggested that page authors create pages in different information densities leading to the existence of three types of pages: (1) *general* pages that contain many, but not all, facts in medium amounts of detail about the topic; (2) *specific* pages that contain a few facts in high amounts of detail about the topic; and (3) *sparse* pages that contain a few facts in low amounts of detail about related topics.

The above three types of pages suggest that users can maximize the retrieval and comprehension of the information that they contain by visiting the pages in a general–specific–sparse order. Because the general pages contain breadth information spanning many facts, such pages are better to visit first to enable an overview of all the relevant facts. After a user has obtained an overview of the topic, the user could then dig deeper into specific pages to find detailed information about particular facts. Having obtained the detailed information, the user might then find information about related topics. This approach is similar to the general to specific search recommended by Kirk (1974) for novice searchers searching in an unfamiliar domain.

The danger of the alternate approach (where users find and access specific pages first) is that without an overview of the topic, they can either (a) become overwhelmed with a few details and end their searches prematurely resulting in incomplete information, or (b) have to go back and forth between the sources to make sense of all the information similar to the recursive strategy identified by Florance and Marchionini (1995) discussed earlier. Both of the above situations could lead users to be inefficient and ineffective in retrieving comprehensive information about a topic.

Information Specialization

Another reason why ordering of sources has become critical when searching for comprehensive information about a topic is how sites within a domain have specialized into various site genres (Crowston & Williams, 1997). For example, in the healthcare domain, sites range from *ask-a-doc* sites (that provide answers to healthcare questions from a real doctor), to sites that provide *risk calculators* (that calculate your risk for suffering from a particular disease). E-commerce sites range from *review* sites (that provide reviews by consumers for different products), to *price comparison* sites (that provide a list of online vendors that sell a product, ranked by price).

Experts in various domains know about the existence of these site genres and form subgoals to exploit this specialization when searching for information. For example, in the study mentioned earlier (Bhavnani, 2001), when looking for three low prices for a new digital camera, shopping experts first visited review sites (e.g., cnet.com) to learn which cameras were highly rated, followed by finding low prices through a price comparison site (e.g., mysimon.com), followed by looking for discounts in yet another set of sites that advertised online discounts (e.g., staples.com).

Once again, the experts had procedural knowledge of how to sequence their search through genres of sites leading to comprehensive information about digital cameras and their prices. Such searches led to lower prices for high-quality cameras when compared to novices who relied on Google for their searches, and who did not infer the critical subgoals known by the experts from the links provided by Google. The specialization of content therefore also implies an ordering where some sites need to be accessed before others because of the content they provide.

The Need to Make Procedural Search Knowledge Explicit to Users

Although experts have identifiable search procedures that appear to improve search efficiency and effectiveness, such knowledge is not easily inferred from the information provided by conventional search tools. General-purpose search tools like Google provide a ranked list of URLs that are relevant to the query based on link analysis (Brin & Page, 1998), and the occurrence of the query in the pages. While the ranking algorithm attempts to give higher ranks to pages that are most pointed to, there is no explicit guarantee for the reliability of such highly ranked sites, nor any indication of the critical subgoals in a domain to guide which pages to visit in which order. Furthermore, our current research has shown that Google provides general, specific, and sparse pages in no particular order. Domain portals such as MEDLINEplus, while providing reliable sources of information, also do not provide the procedural knowledge to organize visiting different sources of relevant information.

The focus of the above systems is to provide relevant URLs either as hits as a result of a query, or as manually constructed categories. Neither of them provides the critical procedural knowledge to guide users on which links to visit in which order. The lack of search procedures directly affects the performance of users when they search in unfamiliar domains because they have to infer the important subgoals, order them, and determine which URLs to visit to satisfy each subgoal. This, as previous research has shown, can be error-prone and time-consuming even for expert searchers when they look for information outside their domain of expertise (Bhavnani, 2001).

It is critical to note that we are fully aware that search is not "algorithmic" in nature, and that most search strategies are heuristic or rules of thumb. Our analysis of the procedural search knowledge is not an attempt to identify inflexible paths that do not appreciate user's changing information needs (Bates, 1989). Rather, the search procedures we explore are *suggested* paths to assist users visit pages that can help comprehensive coverage, and that enable a more effective comprehension of information, especially when searching for information in vast and unfamiliar domains such as healthcare.

As discussed earlier, a few researchers have suggested the importance of search procedures, and several others have developed authoring applications that enable users to add search paths through hypertext systems. The research presented in this article contributes to the above in the following ways:

- We systematically identified expert search procedures for topics in a domain.
- We analyzed the components of the expert search procedures.
- We made the search procedures available to novice users on the Web in a new kind of domain portal called a Strategy Hub.
- We analyzed how novice users performed search tasks using the Strategy Hub compared to conventional search approaches like Google and MEDLINEplus.

Identification of Search Procedures to Find Comprehensive Information About Melanoma

We chose to focus our research on the healthcare domain with a focus on melanoma (a deadly form of skin cancer) for three reasons:

1. **Desire for comprehensive and accurate information.** Several studies have shown that a majority of patients desire comprehensive and accurate information about their illness to help them achieve important coping outcomes. Such outcomes include promoting self-care and treatment compliance, reducing anxiety, and learning the language of their disease (Hinds, Streater, & Mood, 1995; Ream & Richardson, 1996; for a review see Mills & Sullivan, 1999).
2. **Difficulty in finding comprehensive and accurate information.** While patients desire comprehensive and accurate healthcare information, they have difficulty in achieving their goal. Several studies have shown that novice searchers begin their search by entering a few query terms in search engines like Google (Eysenbach & Kohler, 2002, Fox & Fallows, 2003), access the resulting hits in the order presented (Bhavnani, 2001), do not check the reliability of their sources (Eysenbach & Kohler, 2002), and end their searches prematurely without accessing sources that in combination provide comprehensive information (Bhavnani, 2001).
3. **Access to domain experts.** We focused on the disease melanoma because we had access to two skin cancer physicians who not only specialized in treating melanoma, but also had experience in studying the information needs of patients (Johnson, 2003) and searching the Web for melanoma information (Bichakjian et al., 2002).

To begin a systematic identification of search procedures, we used an existing, empirically-based taxonomy of real-world skin-cancer questions (Bhavnani et al., 2002). This taxonomy, developed by two skin cancer physicians through an interrater study, was based on real-world questions, and had high interrater reliability. It is similar to Pratt, Hearst, and Fagan's *query-type* taxonomy for healthcare searches (1999), although it includes one additional top-level category (terminology). The first column in Figure 1 shows this skin cancer taxonomy (see Bhavnani et al., 2002 for a definition for each node in the taxonomy).

To assist the physicians in identifying search procedures for leaf nodes in the taxonomy, we generalized each leaf node in the taxonomy to the form, *My question relates to <topic in the taxonomy> for <disease in the taxonomy>*. This generalized form is referred to as a *question type*. For example, the melanoma questions in the Risk/Prevention Qualitative category were generalized to the question type: *My question relates to <qualitative information on risk factors and prevention> for <melanoma>*. Examples of these question types are shown in the second column of Figure 1.

Two skin cancer physicians were given 15 question types based on the 15 leaf-nodes in the taxonomy, and were asked to pool their past experience to describe explicitly the steps they would take to answer each question type. Furthermore, they were encouraged to access the Web to identify Web pages that were appropriate to retrieve information for each step. The two physicians identified the search procedures and associated links by working together and reaching a consensus. This resulted in the identification of 15 search procedures, one for each leaf node in the taxonomy (three of which are shown in the third column of Figure 1). Analysis of the 15 search procedures to find comprehensive melanoma information provided the following insights

DOI: 10.1002/asi

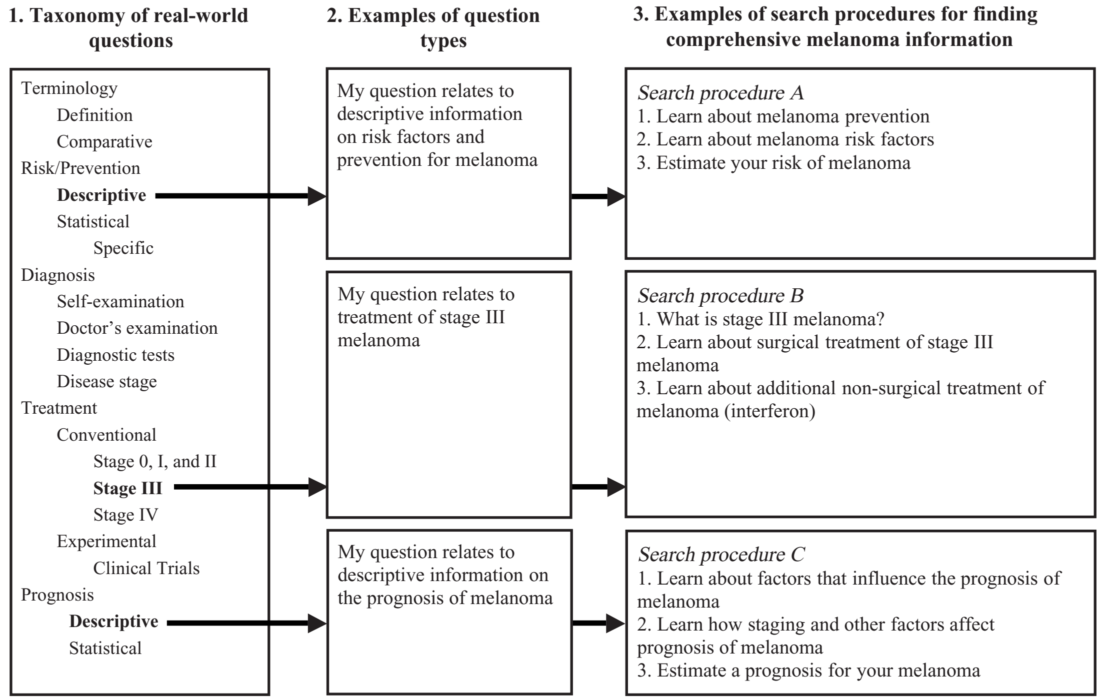

FIG. 1. A taxonomy of real-world questions (Column 1), was abstracted to question types (Column 2). The question types were used by skin cancer physicians to systematically identify search procedures (Column 3), which were subsequently generalized into templates (see Appendix A).

about subgoals, links, and generalizations of the search procedures:

1. **Subgoals.** Each search procedure consists of between two to four steps, which represent critical subgoals to find information about a topic. These subgoals were motivated by two reasons: (a) They attempt to provide perquisite overview information about a topic before providing more detailed information. For example, search procedure B in Figure 1 has the subgoal, *What is stage III melanoma?* before the subgoal, *Learn about surgical treatment of stage III melanoma.* (b) Some search procedures address the variability in the specialization. For example, Step 3 (estimate your risk of melanoma) for search procedure A in Figure 1 addresses the existence of a melanoma risk calculator on Harvard's cancer-prevention Web site. However, because this site does not provide a description of the risk factors that are used in the estimate, users must therefore first visit other Web pages that provide this prerequisite information. This is reflected in steps 1 and 2 of the same procedure, each of which requires the user to visit different links pointing to different genres of healthcare sites.
2. **Links.** Many of the subgoals had more than one link reflecting the scatter of information. For example, the second subgoal of search procedure B in Figure 1 (*Learn about surgical treatment of stage III melanoma*) had two links to cover all the information about surgical treatment of melanoma. A closer analysis of the links across search procedures revealed that the content of sets of pages for a search procedure varied in the amount of overlapping information. For example, search procedure B in Figure 1 had the information repeated in three pages, whereas search procedure C had no repeating information. These differences could be caused because we were working from a taxonomy that had a much finer level of granularity compared to the existing Web pages that were selected by the physicians.
3. **Generalization of search procedures.** Although the procedures identified by the skin cancer physicians were focused on searching for melanoma information, they followed patterns that generalized to two levels as shown in Appendix A. At the lowest level of generalization, the search procedures generalized to four *templates* that appear to be useful for other diseases within healthcare. For example, the *specialization* template could be used to identify search procedures for other diseases such as HIV/AIDS: (a) understand the nature of the HIV retrovirus, (b) learn about antiretroviral drug therapy, and (c) learn about immune-based therapies). We are currently attempting to use these generalized templates to rapidly identify search procedures for other diseases in the healthcare domain.

At the next level of generalization, we identified two templates: (a) General to Specific, and (b) Basic to Related. These templates appear to generalize across domains. For example, the General to Specific template is useful such as when shopping for a digital camera. A recommended procedure is to visit a Web page that describes general information about a digital camera, before visiting a Web page that describe attributes of a specific camera such as its price and features.

While the search procedures and templates appear to generalize, it is currently not clear how to select templates for specific topics. Our hunch is that the templates provide a starting point from which an analyst can generate search procedures using the following process: (a) identify an empirically based taxonomy of questions

DOI: 10.1002/asi

asked in a domain, and (b) use the question taxonomy to elicit search procedures from domain experts using the templates as a starting point. Alternatively, as we will discuss later, an automated approach might be used to approximate the search procedures by using the highest level in the taxonomy.

Having identified 15 expert search procedures to find comprehensive information for melanoma, we were motivated to explore how they could be made available on the Web in a new form of domain portal called a Strategy Hub.

Design and Development of the Strategy Hub

As described earlier, neither search engines, nor domain portals provide the search procedures that we have identified from search experts. Therefore, users searching for comprehensive information in an unfamiliar domain have the difficult task of inferring these search procedures from a list of ranked hits, or from coarse-grained selection categories typically provided by domain portals. We therefore designed a new kind of domain portal called a Strategy Hub to address this issue.

Overview of the Strategy Hub Design

As shown in Figure 2, the home page of the Strategy Hub guides the user to select a disease from a disease hierarchy. The hierarchy behaves similarly to a directory structure in Windows where nodes can be opened up to the leaf nodes. When the cursor moves over any of the nodes in the taxonomy, an explanation of that node is provided in the *Brief explanation* box. The selection of a disease results in the display of a disease topic taxonomy related to that disease. As shown in Figure 3, selecting *Melanoma* results in the display of the skin cancer taxonomy (shown in Figure 1.) Similar to the home page, when the cursor is moved over any of the nodes in the disease topic taxonomy, an explanation of that node is provided in the *Brief explanation* box. When a leaf node is selected, the associated search procedure is displayed as shown in the left-hand window in Figure 4. Each step has an associated link or set of links. Selection of a step in the search procedure results in a new window displaying the page associated with the selected link, as shown in the lower right-hand window in Figure 4.

Strategy Hubs have two characteristics that distinguish them from conventional portals:

1. They provide selection categories that are defined at a finer-grained level to enable users to learn more precisely how information in the domain is organized, and to select appropriate topics of interest. For example, while "Treatment" is a leaf node in the categories provided by MEDLINEplus, our implementation of a Strategy Hub provides two more levels of specificity below Treatment (conventional and Stage III) as shown in Figure 3.
2. They provide explicit search procedures consisting of ordered subgoals, in addition to reliable links to satisfy each subgoal to find comprehensive information about a selected topic. For example, selection of the node Stage III in Figure 3 will provide the search procedure and links shown in Figure 4.

The above representations are aimed to provide users with a useful way to organize melanoma information, and a systematic way to conduct a search. Such representations

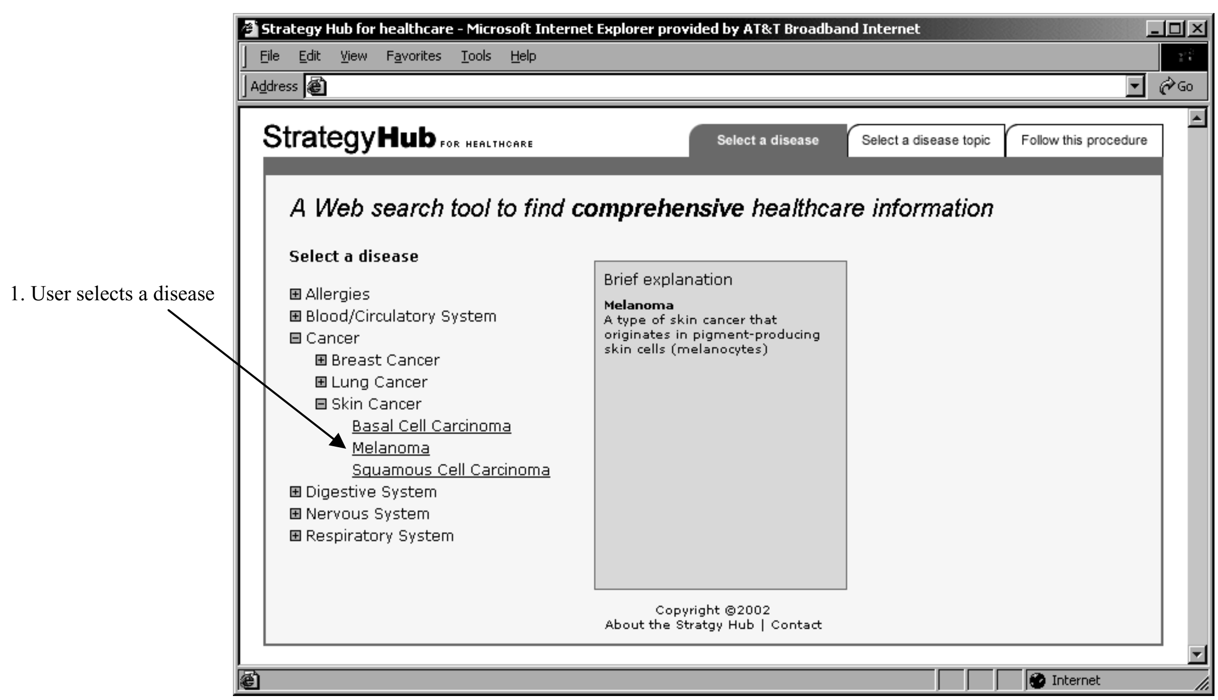

FIG. 2. The homepage of the Strategy Hub guides the selection of a disease. The *Brief explanation* box provides explanations when the cursor is moved over a particular node.

DOI: 10.1002/asi

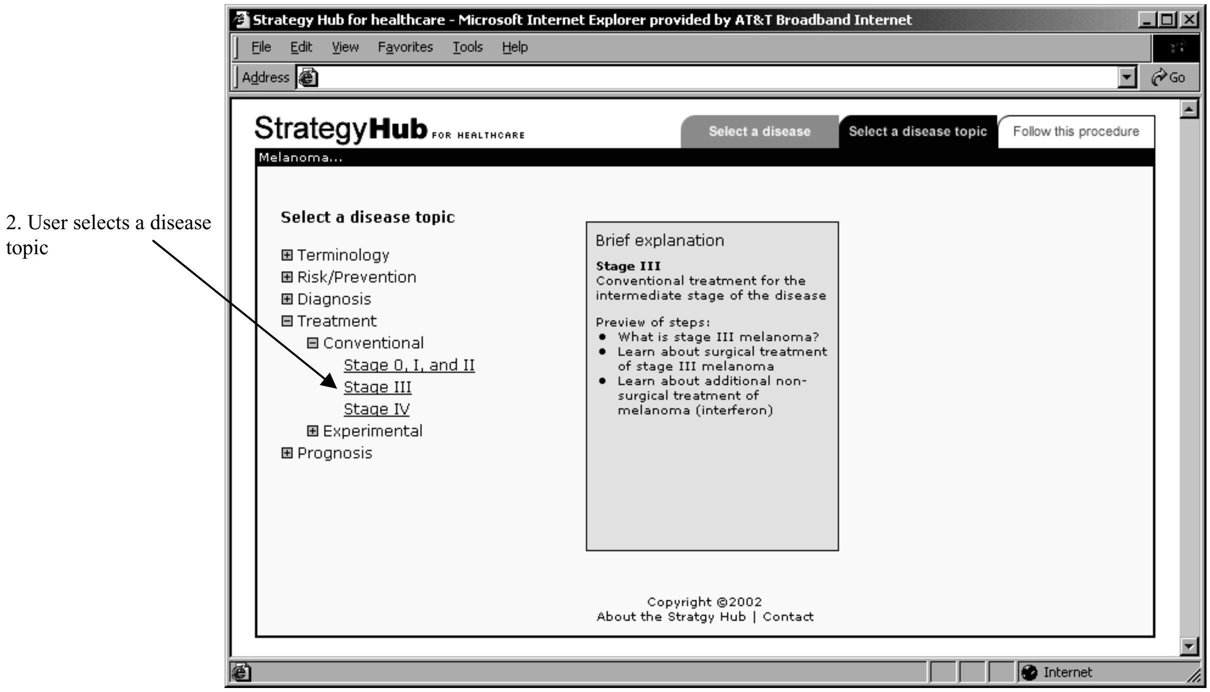

FIG. 3. After selecting a disease, the user is presented with a taxonomy of disease topics. The *Brief explanation* box provides explanations when the cursor is placed above a particular node, in addition to a preview of the steps of a search procedure to find information related to that node.

can help users reduce the cost of *sensemaking* (finding a representation and encoding information in that representation to perform tasks [Russell, Stefik, Pirolli, & Card, 1993]) when searching for information in an unfamiliar domain.

One might argue that instead of spending resources on eliciting search procedures and links from physicians, we could have asked them to answer the questions directly, and used the answers to build a healthcare content site, rather than a domain portal. There are three reasons why we focused on the latter approach: (a) There already exist excellent content pages for all the question types we identified, and therefore the problem is not the lack of accurate information, but rather how to find it; (b) rapidly changing healthcare information is much more likely to be updated in existing authoritative pages, a fact which can be leveraged by linking to them; and (c) by analyzing how experts select and order sources, we have the opportunity to understand how to automate the approach for other topics where we might not have such expert input.

As discussed in Bhavnani et al. (2003), the design of the Strategy Hub prototype was guided by design principles critical for search interfaces that have been suggested by Furnas, Landauer, Gomez, and Dumais (1987), Egan, Remde, Landauer, Lochbaum, and Gomez (1989), and Shneiderman, Byrd, and Croft (1997). Below, we focus on three principles that were critical for the success of the Strategy Hub.

Address Vocabulary Problem

As shown in Figures 2 and 3, the home page guides the user to select a disease, and a disease topic from a hierarchy based on the empirically-based taxonomy developed by the experts. The hierarchies behave similarly to a directory structure in Windows where nodes can be opened up to the leaf nodes. While such hierarchies are intuitively clear, several studies have shown the difficulty that users have in mapping real-world goals to interface elements such as icons on an interface. Furnas et al. (1987) refer to this as the “vocabulary problem,” which is based on the observation that users differ substantially in the terms they use to describe a goal or object, and the overlap of the terms between users is small. The vocabulary problem therefore leads to the difficulty of providing short descriptions of interface elements, such as in a taxonomy, which serves all users.

Furnas et al. (1987) suggest that the vocabulary problem can be addressed by providing multiple synonyms for interface elements. As shown in Figures 2 and 3, this problem is addressed by adding the *Brief explanation* box in the interface of the Strategy Hub, which provides explanations in grammatical sentences that increase the overlaps between the terms people use to describe concepts. In addition to providing explanations, the *Brief explanation* box also provides the steps of the procedure to give the user some “information scent” (Chi, Pirolli, & Pitkow, 2000) about the steps to expect. The brief explanation box reduces the chance of users selecting the wrong nodes.

Provide Focus Plus Context

Once the user selects a disease and disease topic, the system responds by providing a sequence of recommended steps to search for information related to the disease topic, in

DOI: 10.1002/asi

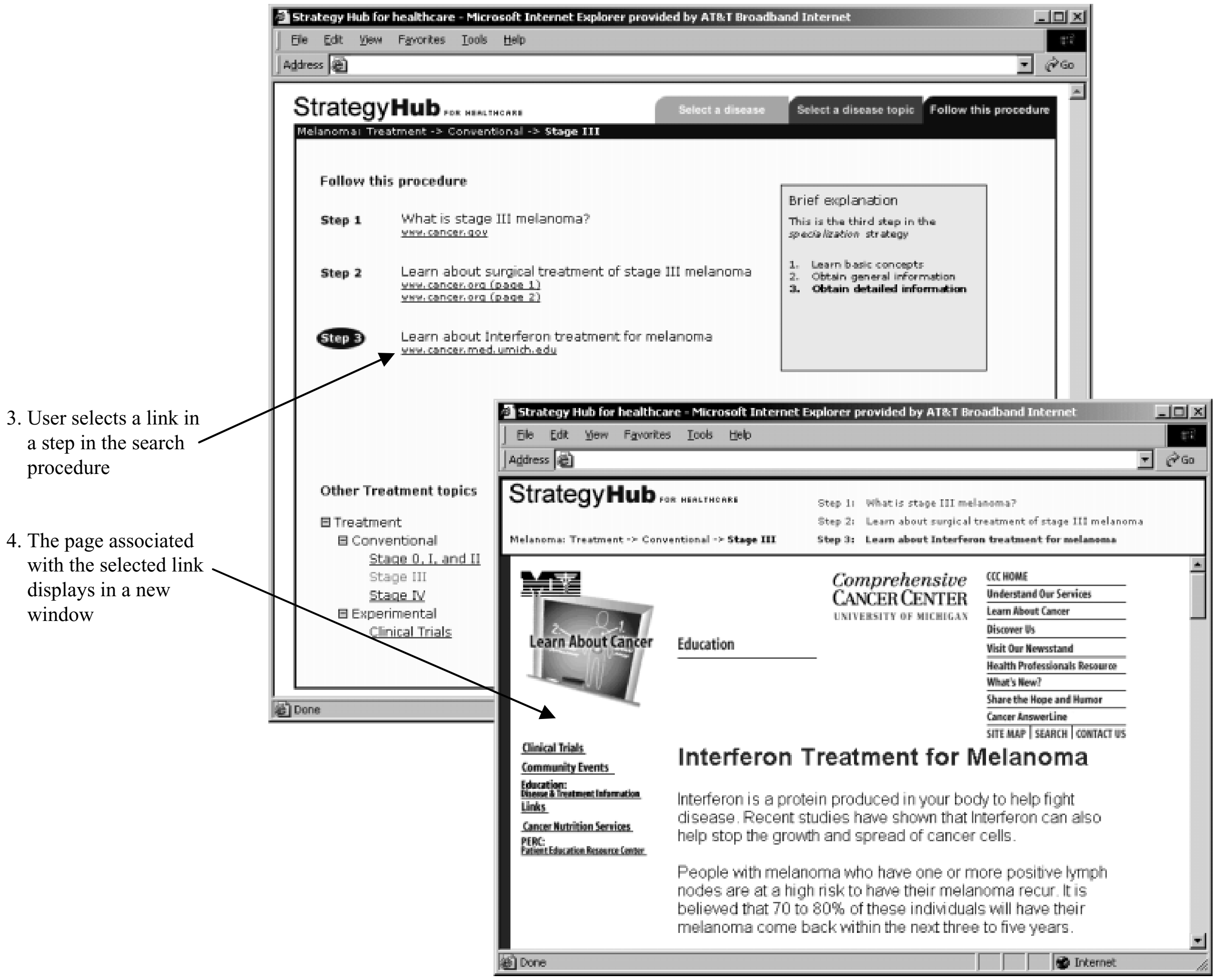

FIG. 4. After “melanoma” and “Stage III” are selected, the Strategy Hub (shown in the upper left) displays the recommended steps of a search procedure, with reliable links for each step. When a link is selected, the associated page is displayed in a new window (shown in the lower right-hand corner). The entire search procedure is also visible in the upper frame of the new window, with the appropriate step bolded.

addition to links at each step. When a link is selected, the corresponding page is displayed in a new window. For example, as shown in Figure 4, the link to the University of Michigan site has been selected in the third step, and the associated page is displayed in the lower right-hand window. This dual-window design is important because it is easy to forget the overall steps in a plan unless it is visible at all times. As shown in Figure 4, the entire search procedure is further emphasized by displaying a copy of it in the upper frame of the new window. The dual-window design therefore provides a combination of a context view, which shows you where you are in the procedure, and a focused view of the content. This combination of focus plus context has been found to be critical for interfaces related to search (Egan et al., 1989).

Provide User Control

The interface is designed to provide the user with a large amount of control to navigate through the system. The user can abandon a search procedure at any time by either returning to the disease or disease topic page through the tabs (shown in the top window in Figure 4), or directly accessing other nodes within a disease topic. For example, as shown in the lower part of the top Strategy Hub window in Figure 4, the user can jump to any other node within the Treatment hierarchy, at which point the appropriate search procedure will be displayed.

To better understand the advantages of providing search procedures through the Strategy Hub, and to identify problems with the interface design, we conducted an experiment to compare the Strategy Hub with conventional search tools.

Experiment Comparing the Strategy Hub to Conventional Search Tools

We designed an experiment to test the Strategy Hub based on two goals: (a) to analyze the efficiency, effectiveness, and satisfaction of Strategy Hub users, when compared to users of conventional search tools, and (b) to analyze the usefulness of the search procedures to help find comprehensive information about melanoma topics.

DOI: 10.1002/asi

Hypothesis

We hypothesized that subjects who had little or no experience in searching for healthcare and melanoma information on the Web, would be more effective, efficient, and satisfied in retrieving complete and accurate information from the Web for comprehensive questions when using the Strategy Hub, compared to similar subjects who use conventional search tools. We chose to focus our experiment on freshman nursing students who had high motivation to learn about healthcare topics, but had low experience in searching for healthcare information. Our goal was to refine the Strategy Hub design with a population of nonpatients highly motivated to find healthcare information, before we tested it with real patients in a future study. Furthermore, our goal was to compare the Strategy Hub as a whole (which included the fine-grained topic taxonomy, search procedures, and interface design) to the other tools, before we committed to more-detailed experiments that analyzed the role of the different parts.

Experimental Design

The hypothesis was tested in a 3×2 between-subject design as shown in Table 1. The first factor consisted of the following three tool conditions:

- **Tool Condition 1: Any Search Tool.** In this condition the subjects were instructed to use any search tool (e.g., Google) or set of search tools. This represents a realistic condition of how most people currently search the Web for healthcare information (Fox & Rainie, 2000). For this condition, the browser window was set up with a blank page displayed.
- **Tool Condition 2: MEDLINEplus Homepage.** In this condition, the subjects were instructed to use only MEDLINEplus (a reputed healthcare portal, which provides links to many reliable Web sites) to perform the search task. For this condition, the browser window was set up to display the home page of MEDLINEplus. This condition represents a realistic situation of how a subject with knowledge of healthcare portals would find information on the Web.
- **Tool Condition 3: Strategy Hub Homepage.** In this condition, the subjects were instructed to use only the Strategy Hub to perform the search task. For this condition the browser window was set up to display the home page of the Strategy Hub (as shown in Figure 2). This condition represents a hypothetical situation where a patient is advised by her physician to use the Strategy Hub to retrieve information about melanoma.

The second experimental factor consisted of two task conditions related to learning about melanoma treatment and diagnosis. These tasks represent two of the major subtopics in our empirically-based taxonomy of melanoma questions discussed earlier. The wording for the two tasks (shown in Table 1) was determined by the skin cancer physicians to provide a realistic search context for nonpatients looking for melanoma information. Because the melanoma physicians were experienced in studying the information needs of patients (Bichakjian et al., 2002), the questions reflected the real information needs of skin-cancer patients.

The above between-subject design (with each subject answering only one question in each tool condition) was necessitated to control for order effects. This was because users typically visit many pages during a search and may inadvertently find answers to a later question.

Subjects

Freshmen from the School of Nursing at the University of Michigan were recruited to take part in the study in return for $25. The recruitment yielded 79 subjects, of whom only 59 actually attended the experiment. Most (94.9%) of the subjects were female. The mean age of the subjects (except three whose age data were not available) was 18.32 years. All subjects were between 17–20 years old, except one who was 32 years old.

None of the subjects had attended any formal training on searching for healthcare information, but all had received formal training in the preceding weeks on how to use Microsoft Word to create documents using the Windows operating system.

Instruments

We used two instruments to balance subjects on important characteristics, and six instruments to measure outcomes.

1. *Instruments to balance subjects across conditions.*
   Search experience. Search experience was assessed through the question, "How often do you conduct a search on the World Wide Web?" and the associated 5-point Likert scale: 1 (*Never*), 2 (*Once or twice a year*), 3 (*Once or twice a month*), 4 (*Once or twice a week*), 5 (*Once or more times a day*).

TABLE 1. The pilot experiment had a 3×2 between subject design that varied by tool and task type. Users in each tool condition were given only one task to guard against order effects. The slightly uneven distribution of the number of subjects in the cells was caused by subjects not showing up for the experiment.

	Any tool chosen by user	MEDLINEplus homepage	Strategy Hub homepage
Your friend has a type of skin cancer called melanoma. Please try and learn all you can about the *treatment of melanoma* so you can describe the treatment of melanoma to your friend.	8 subjects	10 subjects	12 subjects
Your friend thinks that she has a type of skin cancer called melanoma. Please learn all you can about the *diagnosis of melanoma* so you can describe to your friend how she and her doctor can diagnose if she has melanoma.	8 subjects	11 subjects	10 subjects

DOI: 10.1002/asi

Melanoma search experience. Prior melanoma search experience was assessed through the question, "How often do you conduct a search for melanoma information on the World Wide Web?" and the associated 5-point Likert scale: 1 (*Never*), 2 (*Once or twice a year*), 3 (*Once or twice a month*), 4 (*Once or twice a week*), 5 (*Once or more times a day*).

2. *Instruments to measure outcomes.*
Search effectiveness was measured by the accuracy of an essay answer to the search question, and score on a multiple-choice test on melanoma knowledge. In addition, we measured satisfaction, trust, perception of correctness, and value of search procedures.

Accuracy of essay answer. The subjects' essay answer was judged by two independent raters who assessed the natural language answers based on a list of criteria for each question. The weighted criteria were English statements that a skin cancer expert identified as important statements related to each question. For example, the statement "Treatment for melanoma is based on the stage of the disease" was considered important to be present in an answer for treatment (weight = 5 on a scale of 1–5), while the statement, "Radiation therapy is the use of high-energy rays to kill the melanoma tumor" was considered less important (weight = 3). To account for natural language variations, many of these statements had optional synonyms for descriptive terms. For example, "entire skin surface" could be replaced by "full body." The treatment and diagnosis questions had 33 and 13 criteria, respectively. Appendix B shows the entire set of criteria and their weighted importance for the treatment and diagnosis questions.

Score on multiple-choice test. The subjects' knowledge of their task was also assessed by a 10-question multiple-choice test. These questions were identified by a skin cancer expert. For example, the question, "What is the primary purpose of a sentinel lymph node biopsy?" tested the subject's knowledge of an important diagnostic procedure for melanoma.

Satisfaction. Search satisfaction was assessed through the question, "Please indicate how satisfied you are with your search by circling the appropriate number below" and the associated 5-point Likert scale: 1 (*Extremely unsatisfied*), 2 (*Somewhat unsatisfied*), 3 (*Neutral*), 4 (*Somewhat satisfied*), 5 (*Extremely satisfied*).

Trust. The subjects' assessment of how trustworthy the sites they had visited was measured through the question, "Please indicate how much you trust the sites that you visited by circling the appropriate number below" and the associated 5-point Likert scale: 1 (*Extremely untrustworthy*), 2 (*Somewhat untrustworthy*), 3 (*Neutral*), 4 (*Somewhat trustworthy*), 5 (*Extremely trustworthy*).

Certainty. The subjects' assessment of the answer correctness was measured through the following question, "Please indicate how sure you are about the correctness of your answer by circling the appropriate number below" and the associated 5-point Likert scale: 1 (*Extremely uncertain*), 2 (*Somewhat uncertain*), 3 (*Neutral*), 4 (*Somewhat certain*), 5 (*Extremely certain*).

Value of search procedures. The subjects in the Strategy Hub condition were asked to answer two[3] pairs of questions about the search procedures. The first pair of questions was about the entire search procedure: "To what extent was the order of the steps in the procedures unhelpful or helpful in the Strategy Hub?" 1 (*Extremely unhelpful*), 2 (*Somewhat unhelpful*), 3 (*Neutral*), 4 (*Somewhat helpful*), 5 (*Extremely helpful*). Explanations for the above rating was solicited by: "Please explain reasons for the above rating of how unhelpful or helpful was the order of the steps."

The second pair of questions was about the information in each step: "To what extent was the information in each step of the procedures (e.g., 'Step 1: Learn about melanoma prevention') unhelpful or helpful in the Strategy Hub?" 1 (*Extremely unhelpful*), 2 (*Somewhat unhelpful*), 3 (*Neutral*), 4 (*Somewhat helpful*), 5 (*Extremely helpful*). Explanation for the above rating was solicited by: "Please explain reasons for the above rating of how unhelpful or helpful was the information in each step."

In addition to the above outcome measures, we also measured efficiency using the recorded time on the screen capture videos. A more detailed description of how we measured efficiency is described in the *Analysis* subsection below.

Experimental Method

The subjects were balanced across the six conditions by prior experience in searching, and prior melanoma search experience (i.e., the subjects were randomly assigned to the different conditions, with the constraint that each condition contained the same proportion of experience levels). The slight uneven distribution in the number of subjects across the conditions was the result of scheduled subjects not showing up for the experiment.

Subjects in each condition were asked to watch a digital video that provided instructions to perform one of two tasks provided in a Word document on their computer. The users were instructed to search for answers using only the approach in their condition, within a maximum of 35 minutes (this time limit was determined from a pilot experiment where 25 minutes was too short for many to complete similar tasks, whereas longer times resulted in subjects widely varying in their interpretations of what was expected from the task). The subjects were instructed to cut and paste into the Word document the paragraphs from the Web that they felt contained the answer to the task in their condition. This was done to distinguish the retrieval of information from the

[3]These two questions were part of a longer questionnaire that consisted of five other questions about the overall experience of using the Strategy Hub. We present results from only two questions because those were directly related to the search procedures and the current study. The other five questions elicited generally positive responses about the overall experience of using the Strategy Hub.

interpretation and construction of the final essay answer. The Word document also had a section that instructed the subjects to write out their essay answer. For example, the subjects in the treatment condition had the following instruction: "Please type out the description of melanoma treatment that you would give to your friend." This section is where the subjects' typed out their final essay answer in their own words.

Regardless of completion, users were asked to stop searching and writing after 35 minutes. A screen capture tool was used to record the interactions on the screen, and a Web logger developed by Xerox PARC (Reeder, Pirolli, & Card, 2000) was used to record the time and occurrence of keystrokes related to clicking a link, scrolling, and using the back button.

After the subjects completed the task, they were instructed to turn off their monitors, and answer the 10-question multiple-choice melanoma test that was specific to the task in their condition. The subjects were also instructed to rate how satisfied they were with their search, how much they trusted the sites they visited, and how certain they were about the correctness of their answer. The Strategy Hub subjects were asked to complete an additional questionnaire, two questions of which were about search procedures.

Analysis

Search effectiveness was measured by (a) the accuracy of the essay answer to the search question, and (b) a score on the multiple-choice melanoma test. The accuracy of the essay answer was judged by two independent raters (blinded to condition, and not connected with the project), who assessed the answers based on a list of criteria for each question discussed earlier and shown in Appendix B. The agreement between the two judges was calculated, and a consensus agreement reached for the ratings where they disagreed. The second measure for search effectiveness was the score on the multiple-choice test.

Search efficiency was measured by total search time [total task time − (cut and paste time + writing time)]. Satisfaction was measured by the responses of how satisfied users were with their search on a 5-point Likert scale. The secondary measures of trust and certainty were similarly measured on the corresponding 5-point Likert scales.

Results

Search effectiveness (essay answer). Interrater agreement between the two judges for essay answers to the treatment question was at 96.86% (Cohen's $\kappa = 0.859$) where they agreed on 959 out of 990 categories. Agreement for answers to the diagnosis question was at 90.98% agreement (Cohen's $\kappa = 0.795$) where they agreed on 343 out of 377 categories. The judges arrived at a consensus score for the answers where they disagreed.

A two-way ANOVA on mean accuracy score for the essay answer revealed a significant main effect for tool condition,

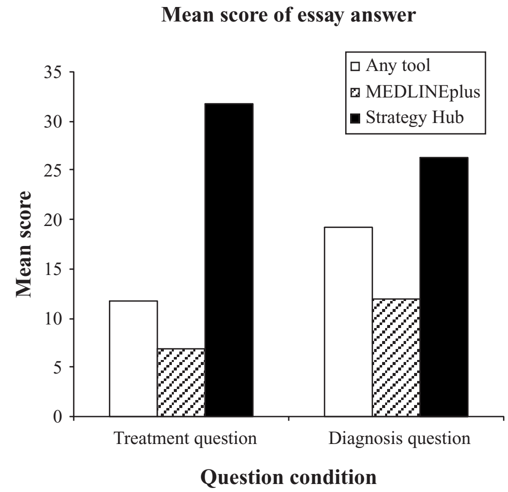

FIG. 5. Mean accuracy scores of the essay answers for the treatment and diagnosis questions, across the three tool conditions. For the treatment question, Strategy Hub scores were significantly higher than Any tool ($p < .05$) and MEDLINEplus ($p < .01$). For the diagnosis question, Strategy Hub scores were significantly higher than MEDLINEplus ($p < .001$).

$F(2,53) = 12.58$, $p < .001$. There was neither a significant effect for question, $F(1,53) = .51$, $p = .480$, nor for the interaction of question and tool condition, $F(2,53) = 1.35$, $p = .268$.

To further assess the significant main effect for tool condition, we performed pair-wise t tests between each tool condition. As shown in Figure 5, Strategy Hub subjects had a significantly higher mean accuracy score for their treatment answers when compared to the MEDLINEplus, and Any tool subjects. (Strategy Hub > MEDLINEplus [31.67, 6.9, $p < .01$], Strategy Hub > Any tool [31.67, 11.75, $p < .05$]). The same was true for the Diagnosis question but only for the MEDLINEplus condition. (Strategy Hub > MEDLINEplus [26.4, 12, $p < .001$]). Although the Strategy Hub answers were better than the Any tool condition, the difference was not significant (Strategy Hub = Any tool [26.4, 19.25, $p = .18$]). However, in general, the results provide evidence that the Strategy Hub users produced more comprehensive and accurate essay answers compared to the other conditions.

Search effectiveness (multiple-choice answer). We conducted a similar analysis for the multiple-choice score. A two-way ANOVA on the mean multiple-choice score revealed a significant main effect for tool condition, $F(2,53) = 4.32$, $p < .05$, and a significant main effect for question, $F(1, 53) = 52.71$, $p < .001$. There was not a significant interaction effect, $F(2,53) = .36$, $p = .703$).

To further assess the significant main effect for tool condition, we performed pair-wise t tests. As shown in Figure 6, the Strategy Hub users did significantly better than the other two conditions in the multiple-choice test for the diagnosis question (Strategy Hub > MEDLINEplus [8.6, 7.45, $p < .05$], Strategy Hub > Any tool [8.6, 6.75, $p < .01$]). However, although the Strategy Hub users did better than the

DOI: 10.1002/asi

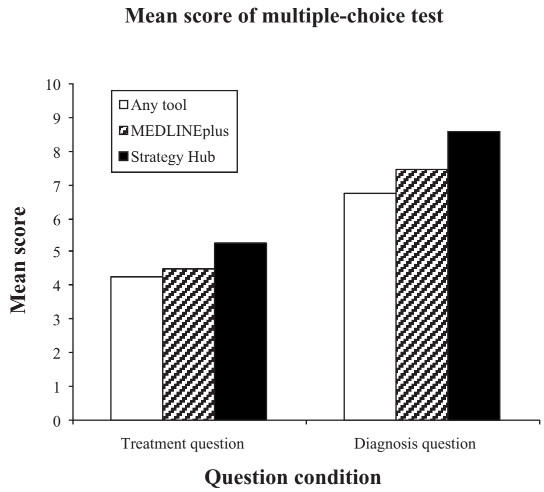

FIG. 6. Mean multiple-choice test score for the treatment and diagnosis questions, across the three tool conditions. Strategy Hub users had significantly higher scores than the other two conditions for the diagnosis question ($p < .01$ for Any tool, and $p < .05$ for MEDLINEplus).

other conditions for the treatment multiple-choice test, the differences were not significant (Strategy Hub = MEDLINEplus [5.25, 4.5, $p = .30$], Strategy Hub = Any tool [5.25, 4.25, $p = .23$]).

A closer analysis of the multiple-choice scores for the treatment question suggested an explanation. As shown in Figure 6, none of the conditions got a mean score that exceeded 55%. Furthermore, as shown by the two-way ANOVA result discussed above, there was a main effect of question type. Both these suggested that, in comparison to the diagnosis multiple-choice test the treatment multiple-choice test was too difficult, limiting its ability to differentiate between the conditions. The multiple-choice test results therefore show that the Strategy Hub users performed better then the other tool conditions only when the test was not too difficult.

Search efficiency. As discussed earlier, search time was defined as total task time − (copy/paste + writing time). A two-way ANOVA on search time revealed a significant main effect for tool condition, $F(2,53) = 6.90, p < .01$. There was not a significant effect for question, $F(1, 53) = .34$, $p = .564$, nor for the interaction of question and condition, $F(2,53) = 2.58$, $p = .086$).

To further assess the significant effect for tool condition, we performed pair-wise *t* tests. For the treatment question, there was no significant difference in search time between the Strategy Hub and the other tool conditions (Strategy Hub = MEDLINEplus [10:48, 9:55, $p = .62$], Strategy Hub = Any tool [10:48, 11:52, $p = .58$]). For the diagnosis question, Strategy Hub users were significantly faster than any tool users (9:30, 14:38, $p < .05$), and significantly slower than MEDLINEplus users (9:30, 6:35, $p < .05$).

Because of the significant main effect for search time, one might argue that the Strategy Hub users were more effective than the other conditions because they searched longer than the users in the other tool conditions. This is similar to the notion of a speed-accuracy trade-off, where the faster the user, the poorer the accuracy.

To test this rival hypothesis, we performed a regression analyses on our effectiveness measures, controlling for search time. The results revealed that, in general, the efficiency effects were unchanged when controlling for search time. For the mean accuracy score, Strategy Hub continued to have significantly higher scores than both tool conditions for the treatment question (Strategy Hub > MEDLINEplus, $p < .001$, Strategy Hub > Any tool, $p < .01$) and for the diagnosis question (Strategy Hub > MEDLINEplus, $p < .01$, Strategy Hub > Any tool, $p < .01$). For the mean multiple choice score, there was still not a significant difference between conditions on the treatment question (Strategy Hub = MEDLINEplus, $p = .29$, Strategy Hub = Any tool, $p = .24$). For the diagnosis question, the Strategy Hub users had marginally higher scores than MEDLINEplus ($p < .10$), and significantly higher scores than Any tool ($p < .05$).

Search satisfaction. There was no significant difference between the Strategy Hub and the other conditions for search satisfaction (Strategy Hub = MEDLINEplus [4.36, 4.60, $p = .23$], Strategy Hub = Any tool [4.36, 4.00, $p = .21$]). However, there was a marginally significant difference between MEDLINEplus and Any tool (MEDLINEplus > Any tool, $p < .06$). Furthermore, all tool/question conditions except one (Any tool/Diagnosis condition had a mean score of 3.88) had a mean score of 4 and above. This result suggests that subjects in Any tool and MEDLINEplus conditions were quite satisfied, *despite* the fact that they had fairly mediocre scores on search effectiveness. This, in turn, suggests that they were poor judges of their own search behavior, a dangerous situation when searching for healthcare information.

Trust and certainty. Instruments for trust and certainty were included in the experiment in an exploratory mode to investigate how certain subjects were of their answer, and how much they trusted the sites they visited. As shown in Figure 7, MEDLINEplus had the significantly higher score for trust for both tasks (MEDLINEplus > Strategy Hub [4.76, 4.10, $p < .001$], MEDLINEplus > Any tool [4.76, 4.06, $p < .01$]). This could be because unlike MEDLINEplus, the Strategy Hub interface does not make salient its author or sponsor. For example, there is no mention of the doctors or the University being involved in the development of the site, an important consideration to gain trust. This was done to eliminate any bias in favor of the University of Michigan, from where the subjects were recruited. Furthermore, unlike MEDLINEplus that has links annotated by the organization, the Strategy Hub provides only the URL of the site. Either or both of these reasons could have caused the Strategy Hub users to rate the links that they visited significantly lower on trustworthiness. Finally, the mean scores for certainty are high across the

DOI: 10.1002/asi

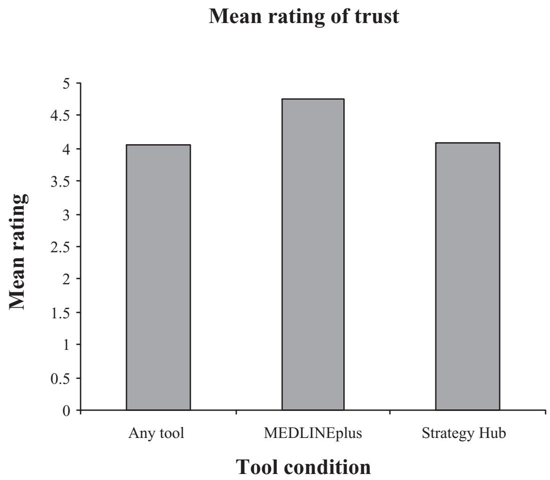

FIG. 7. Mean trust rating collapsed over questions for the three tool conditions. Mean trust rating for MEDLINEplus was significantly higher than mean ratings for Any tool ($p < .01$), and for Strategy Hub ($p < .01$).

board with no significant differences between the means (Strategy Hub = MEDLINEplus [4.32, 4.33, $p = .94$], Strategy Hub = Any tool [4.32, 4.13, $p = .46$]), suggesting an over-confidence on the part of all users. Because of the different layers of meanings possible in the issue of trust and certainty, future research should use a combination of qualitative and quantitative measures.

Value of search procedures. While the overall measures showed an improvement in the search outcomes, the questionnaire data (shown in Table 2) provide direct evidence that the search procedures played an important role in achieving those results. For the question, "To what extent was the order of the steps in the procedures unhelpful or helpful in the Strategy Hub?," the majority of the subjects[4] ($n = 17$, 80.9%) gave a rating of ≧4 on the 5-point Lickert scale. Explanations (some had more than one) for the above ratings fell in the following categories: helped in navigation/research ($n = 9$); provided a narrowing process leading to the information being searched ($n = 5$); was organized and helpful ($n = 5$); helped to answer the question quickly and to retain the knowledge ($n = 1$); order of the steps was constraining because one could not jump to any step ($n = 1$); more information could have been provided besides the links ($n = 1$).

Four (19%) subjects gave the search procedures a helpful rating of ≦3. Their comments were mixed and fell under the following categories: steps were not really helpful ($n = 2$); information would have been more helpful to have everything "right after another" ($n = 1$); search procedures were helpful and made information easier to find ($n = 1$).

For the question, "To what extent was the information in each step of the procedures (e.g., 'Step 1: Learn about melanoma prevention') unhelpful or helpful in the Strategy Hub?" the majority of the subjects ($n = 15$, 68%) gave a rating of ≧4 on the 5-point Lickert scale. Their comments fell into the following categories: Helped organize an unfamiliar topic ($n = 9$); made the search more efficient ($n = 7$); helped target relevant information in the content page ($n = 3$); helped to know what steps were coming up ($n = 2$); information was repetitive, or not relevant ($n = 2$).

Seven (32%) subjects gave the steps in the search procedures a helpful rating of ≦3. Their comments fell into the following categories: Helped organize an unfamiliar topic ($n = 2$); The steps were too restrictive ($n = 2$); information was repetitive, or not relevant ($n = 2$); helped to know what steps were coming up ($n = 1$); steps had awkward wording ($n = 1$); did not remember the steps ($n = 1$).

The overall results therefore suggest that the subjects in the Strategy Hub condition perceived the search procedures as being helpful, with several explicitly noting that the order provided a structure that narrowed and guided the search. This was an important goal for providing search procedures in the Strategy Hub.

Post Hoc Analysis of Interface Use

The Strategy Hub used in the experiment incorporated many design features in addition to the search procedures. For example, Strategy Hub had a finer-grained taxonomy compared to MEDLINEplus, and an interface design that was based on principles for good search interfaces. Our goal was to analyze if the Strategy Hub design as a whole provided any improvements over conventional search tools *before* we conducted further studies to analyze which parts

[4]One subject who rated the search procedures as somewhat helpful was dropped from the analysis because her explanation revealed that she had misunderstood the question to mean steps of the experiment.

TABLE 2. Total number of responses to the two questions on value of search procedures. The responses to the questions indicate that most subjects found the search procedures to be helpful.

	1 (Extremely unhelpful)	2 (Somewhat unhelpful)	3 (Neutral)	4 (Somewhat helpful)	5 (Extremely helpful)
To what extent was the *order of the steps* in the procedures unhelpful or helpful in the Strategy Hub?	0	0	4	8	10
To what extent was the *information in each step* of the procedures (e.g., "Step 1: Learn about melanoma prevention") unhelpful or helpful in the Strategy Hub?	0	1	6	7	8

DOI: 10.1002/asi

of the Strategy Hub accounted for which outcomes. Each of these changes needs to be controlled in future studies to understand their role. The following analysis of how the Strategy Hub interface was used is provided only as post hoc analysis to inform future experiments.

Our post hoc analysis of interface use focused on two observations that we had made during our pilot studies (Bhavnani et al., 2003).

1. **Low stickiness.** During the pilot on an earlier version of the interface, we noticed that many Strategy Hub users followed links on content pages that led them to pages *outside* the Strategy Hub. Many such users never returned to the Strategy Hub. Subsequent interviews revealed that the interface design had misled subjects into concluding that the Strategy Hub was a content provider that provided cancer information instead of a portal with links pointing to content providers. Because these subjects did not understand the distinction between the Strategy Hub and the content pages, they did not realize that they were "leaving" the Strategy Hub by clicking on the content links.
2. **Under-use of search procedures.** During our pilot study, we also observed that a few subjects did not follow the suggested steps in the search procedure, and hence did not use the Strategy Hub as intended. Subsequent interviews showed that the interface design did not make clear the sequential nature of the steps leading subjects to interpret the steps as *categories* of pages. This led users to miss critical steps in the search procedures leading to the retrieval of incomplete information.

Our post hoc analysis therefore probed the following two questions: (a) What percentage of users remained close to the Strategy Hub? (b) Was there a difference in the performance of users who visited critical subgoals in the Strategy Hub versus those that did not?

We addressed the above questions by developing and analyzing *navigation graphs* of all 22 Strategy Hub users in the experiment. As shown in Figure 8, a navigation graph consists of a table where each row represents a relevant subnode in the skin cancer topic taxonomy (shown earlier in Figure 1), and columns represent steps in the search procedure for that node. This table is superimposed by a graph that represents which subnodes and steps were visited by a subject. Solid circles indicate that the subject visited a particular step, and empty circles indicate that the subject visited a particular step but then left the Strategy Hub by clicking on a link inside a content page. A step in the search procedure was considered visited if the subject spent more than 15 seconds on the link, or if the subject copy/pasted information from that link. This was done to distinguish users who clicked on a link just to get to the next step, but did not read the contents. Figure 8A shows the navigation graph of a user who visited many steps, left the Strategy Hub, and then returned. In contrast, Figure 8B shows the navigation graph of a user that visited fewer steps, left the Strategy Hub, but never returned.

The navigation analysis revealed that 16 out of 22 subjects left the Strategy Hub at least once by clicking on a content link within a suggested page, which took them outside the Strategy Hub. Of those 16 subjects, 12 returned, and 4 did not. Therefore, the Strategy Hub successfully retained 18 of 22 subjects (82%), which we believe is an acceptably high level of stickiness.

Next, we probed the relationship between visiting relevant steps in a procedure, and the effectiveness of search results. We first had to identify the steps in a procedure that were necessary to obtain comprehensive information. Analysis of the necessary steps revealed an interesting complication. First, we realized that the *links* provided by the treatment steps had high overlap of content. This made it difficult to determine which steps were necessary and which were not. (This problem is, of course, inherent to the nature of a portal that only provides URLs to other pages whose content it cannot control.) For example, many high-quality melanoma treatment pages are quite long and contain several subtopics. This caused the physicians to select the same pages for several Strategy Hub subtopics. However, in contrast to the Treatment node, the Diagnosis node did not

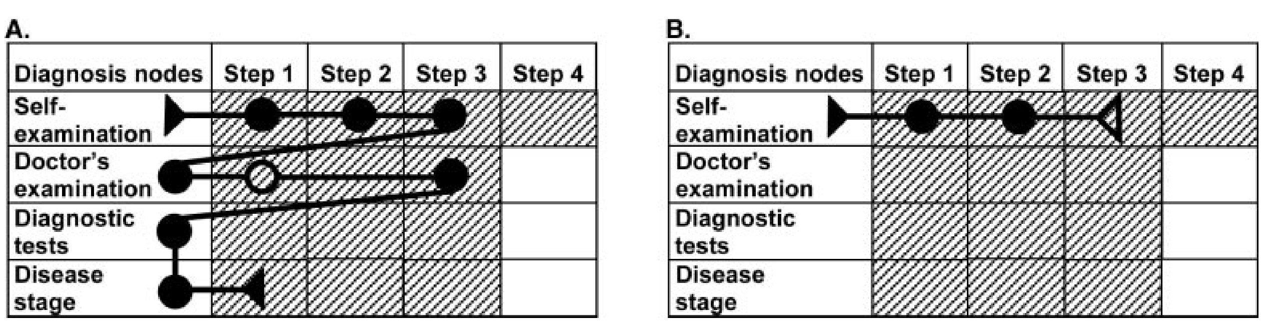

FIG. 8. Two examples of a navigation graph that represent the nodes and steps visited in order by a subject. Hatched cells show which steps were provided in the Strategy Hub for each subnode. Right-facing triangles indicate the first step visited by a subject, left-facing triangles indicate the last step visited by a subject, and circles indicate that the intermediate steps visited by a subject. Solid shapes indicate that the subject visited a particular step, and empty shapes indicate that the subject visited a particular step but then left the Strategy Hub by clicking on a link inside a content page. (A) Shows the navigation graph of a user who visited many nodes and steps, leaving the Strategy Hub, but returning to resume her search. (B) Shows the navigation graph of a user who visited few nodes and steps, and left the Strategy Hub without returning.

DOI: 10.1002/asi

TABLE 3. The Diagnostic node had four subnodes (Self-examination, Doctor's Examination, Diagnostic Tests, and Disease Stage). Subgoals that were necessary to obtain comprehensive information about the Diagnostic task are set bold face.

Diagnosis nodes	Step 1	Step 2	Step 3	Step 4
Self-examination	**Learn about the ABCDs of melanoma detection**	**Learn how to distinguish between a mole and a melanoma**	**Learn how and how often to do a self-examination**	Confirm your self diagnosis by locating a dermatologist near you
Doctor's examination	**When should I consult a doctor for a diagnosis**	**Learn about a doctor's examination of the skin, and how often to get one**	**Learn about different types of diagnostic procedures done in a doctor's exam**	
Diagnostic tests	**What is a biopsy?**	**Learn about biopsies for melanoma**	**Learn about tests to determine if melanoma has spread to other parts of the body**	
Disease stage	Learn the basics of staging	Learn how melanoma is staged	Determine your stage of melanoma	

have as much overlap, leading to more- or less-independent content pages. We therefore focused on analyzing the behavior of users performing the Diagnosis task.

Second, we realized that not all steps under the diagnosis *node* were relevant for the diagnosis task. Therefore, the Diagnosis task required only a subset of the information that was provided by the Strategy Hub. Table 3 shows all the subnodes and respective steps for the search procedures under the diagnosis node. The bolded text represent the nodes that were required to get comprehensive information (based on the list of facts provided to us by the physicians). As shown, only 9 out of 13 steps were necessary to visit to get a comprehensive information of the diagnosis topic.

A qualitative analysis of each user's navigational graphs suggested that users fell into two groups of navigation behaviors. Group 1 ($n = 5$) visited all relevant steps in sequence for at least one node, and Group 2 ($n = 5$) did not. Furthermore, Group 1 had a much higher overlap with the necessary steps compared to Group 2 (51%, 33%). Group 1 might therefore represent subjects who used the Strategy Hub as intended; that is, those subjects who read the subgoals before visiting a link. Analysis of the multiple-choice scores revealed that Group 1 had a higher mean score on the multiple-choice test then Group 2 (9.2, 8.0).[5]

The above two analyses of navigation behavior therefore suggests that the redesigned Strategy Hub had an acceptably high level of stickiness, and that visiting relevant steps in a search procedure increased a subject's multiple-choice score. However, the results are confounded by the overlap of information between pages. Future experiments should explicitly control for such variables, perhaps by modifying real Web pages for the purposes of the experiment so that they do not contain overlapping information.

[5]We focused on the multiple-choice scores (instead of the essay answer) as it represented a more objective comparison of melanoma knowledge across the subjects. Although this difference is significant ($p < .05$), it should be interpreted descriptively rather than inferentially, as these are exploratory analyses that are not testing a specific hypothesis through a controlled experiment.

Discussion

Our results have revealed important strengths and limitations of the Strategy Hub over conventional search tools. Furthermore, the experiment has provided additional insights into the design of our instruments, and the design of future experiments.

Arguably, the most important goal is for a user to find accurate and comprehensive information about a healthcare topic in a reasonable amount of time. Based on the analysis of the essay answers, the results suggest that, in general, the Strategy Hub does in fact significantly improve the quality of answers to a search question in comparison to other search tools, within the constraints of searching and writing out an answer in 35 minutes. However, while the Strategy Hub users were significantly better in their essay answers, they had mixed results when performing the multiple-choice test. This, we believe, is because the multiple-choice test for the treatment task was too difficult, yielding a low score across all conditions. An analysis of the scores revealed that the answers for 3 out of the 10 questions were subtle in nature, requiring only a trained medical expert to know the correct answer. Furthermore, the analysis revealed that 2 out of the 10 questions required very precise knowledge about specific treatments and definitions, which might not have been motivated by the broad nature of the tasks. Finally, the result on search efficiency revealed that the improved search effectiveness neither came at the expense of efficiency, nor was a result of a speed–accuracy trade-off.

It is important to note, that similar to many information retrieval studies, each of the above two measures of search effectiveness combines retrieval *and* comprehension of relevant content. While future studies might control for retrieval and comprehension, we believe variations in comprehension did not create any bias because the subjects were randomly assigned across the conditions. Furthermore, all the subjects were native English speakers, and the material that they retrieved was from consumer healthcare sites written for a lay audience (vs. medical sites written for medical professionals).

DOI: 10.1002/asi

The result on search satisfaction is also a critical result because it demonstrates that users in the Any tool and MEDLINEplus conditions have high satisfaction with their searches, *despite* the quality of their answers being much lower compared to those in the Strategy Hub condition. This in hindsight should not be surprising because by using most available Internet search tools, users can get an answer fairly quickly. However, because users do not have an idea of the scope of a health topic, they could perceive their answers to be complete, and therefore could be ending their searches too early. In contrast, the search procedures in the Strategy Hub provide the essential subgoals, which explicitly show the scope of the topic and when to end their search. We believe this is causing their answers to be of a higher quality. The scores related to certainty also appear related to the above explanation.

The result on trust provided an important feedback on how the Strategy Hub might be perceived in its entirety. As discussed earlier, the Strategy Hub interface has no description of its authors or institution mentioned anywhere, a conscious omission to avoid bias as we were testing the portal on students at the University of Michigan. We hypothesize that users perceived the links they visited through MEDLINEplus to be more trustworthy because (a) the home page clearly displays the sponsoring organizations (the National Library of Medicine, and the National Institute of Health), and (b) the links that are provided also state the sponsoring organizations. Our future interface designs will test if making sponsoring organizations explicit on the interface can improve the trustworthiness of the Strategy Hub.

The questionnaire on search procedures provided evidence that the subjects valued the search procedures. Most subjects stated that the search procedures provided a structure that narrowed and guided their search, whereas very few complained that the search procedures were constraining. Finally, our post hoc analysis of navigation behavior suggested that the current design does retain a high percentage of users, and the users do benefit by following the search procedures. While our experiment focused on analyzing the overall strengths and limitations of using the Strategy Hub (which included a fine-grained taxonomy, interface design based on HCI principles, and search procedures), future studies should specifically probe the value of search procedures through more controlled experiments.

One could argue that the above experimental design was unfair because the Strategy Hub was manually created by physicians with a focus on a single disease, whereas Google and MEDLINEplus contain a lot more information. However, it is important to note that MEDLINEplus is also manually created by experts, and users in that condition surprisingly did not fare very well compared to users in the Any tool condition. Manual creation of portals therefore does not seem to guarantee higher performance. Furthermore, our experimental tasks were focused on a single disease and therefore the breadth of diseases that comparative systems contain should not affect the experimental results.

To address the issue of scalability, we have begun to explore an approach to automate the Strategy Hub. This approach assumes that physicians will pool their knowledge to create a database of facts that they believe patients must know for a comprehensive understanding of specific healthcare topics. When a user selects a topic such as melanoma risk/prevention, the system will (a) extract the corresponding list of facts for that topic from the database, (b) retrieve relevant pages for that topic using Google, (c) use content analysis tools such as latent semantic analysis (LSA; Dumais, Furnas, Landauer, & Deerwester, 1988) to dynamically determine pages with different fact densities such as pages that contain general, and specific information, and (d) automatically generate the search procedures, for example, by ordering the presentation of the general and specific pages. This approach will approximate the majority of the search procedures that we have discussed in this article, and make the Strategy Hub scalable to other domains. Our initial studies have revealed that LSA performed reasonably well compared to a human judge in determining fact-depth and fact-breadth (Peck, Bhavnani, Blackmon, & Radev, 2004), and we are exploring more sophisticated natural language analyses to improve the results. Future research should explore if this scalable version continues to enhance the search effectiveness of users searching in unfamiliar domains.

The experiment also adds another dimension to our understanding of domain knowledge and its effects on search outcomes. While several studies on domain knowledge (see Wildemuth, 2004 for a review) have consistently shown that domain experts use different and more search terms (e.g. Marchionini, 1989, Shute & Smith, 1993, Vakkari, 2002), our study suggests that domain experts also have acquired search procedures that enable them to find comprehensive information about a topic. These search procedures consist of domain-specific subgoals, a recommended order to visit those subgoals, and links to Web pages that contain information about those subgoals. Furthermore, the results suggest that when such domain knowledge is provided in a domain portal, it enables novice searchers without domain knowledge to find more comprehensive information compared to conventional search tools.

Summary and Conclusion

Our research was motivated by two observations.

- Expert searchers have acquired effective and efficient search procedures that guide them to retrieve comprehensive information about a topic from different sources. Such search procedures, we hypothesized, were necessary given the wide scatter of information across sources, and the wide variability of detail and specialization within sources.
- Novices, searching in unfamiliar domains find it difficult to infer such search procedures from conventional search tools, often leading to the retrieval of incomplete information.

To address the above situation, we collaborated with healthcare search experts to systematically identify search procedures to find comprehensive information for a specific disease. Analysis of the search procedures showed that they

DOI: 10.1002/asi

indeed did address the wide scatter of information across sources, and the wide variability of detail and specialization within sources. Furthermore, the search procedures could be generalized into templates at two levels of generality, and therefore could be used to identify search procedures in other domains. We then showed how the search procedures could be made available on the Web through a new form of domain portal called a Strategy Hub using principles of user-centered design.

A controlled experiment demonstrated that the Strategy Hub, in general, could improve the efficiency, effectiveness, and satisfaction of users when attempting to answer comprehensive questions in an unfamiliar domain. The questionnaire data suggests that the Strategy Hub users did find the search procedures helpful in narrowing their search. A post hoc evaluation of the navigation within the Strategy Hub revealed that despite pointing to many pages, the Strategy Hub retained a high percentage of users. This provided evidence that our interface design was an improvement over earlier iterations. Furthermore, there was some evidence that users who visited many steps in sequence provided by the Strategy Hub were more effective than those who did not. Finally, the analysis of the search procedures provided insights on how to automate the Strategy Hub, and make it scalable to other topics and domains.

Although we have shown how search procedures can be provided in a new form of domain portal, we believe the notion of providing such procedural knowledge is much more general. Search procedures can be useful within any large site where there does not exist a one-to-one mapping between a task and a page. In such cases, the retrieval of information from the site would require the user to infer which pages to visit in which order to get comprehensive information about the topic. This can be time-consuming and error-prone. In such situations, search procedures like those we have described, could guide users to appropriate pages in the right order leading to more comprehensive results. Furthermore, we believe search engines (e.g., Vivisimo) that provide automatic categorization of links (Dumais, Cutrell, & Chen, 2001) could also provide search procedures to guide users to find more comprehensive information.

The notion of providing search procedures, their generalization within healthcare and beyond, and the interface design related to how to provide them are therefore the important contributions presented in this article. Besides providing a new direction in the research for search interfaces, search procedures should lead users to be more effective, efficient, and satisfied when finding comprehensive information in unfamiliar domains.

Acknowledgments

This research was supported in part by NSF Award # EIA-9812607, and NIH grants P50CA101451 and P30CA46592. The authors thank M. Bates, N. Belkin, S. Mathan, F. Reif, B. Suhm, G. Vallabha, and the anonymous reviewers for their contributions.

References

Bates, M.J. (1979). Information search tactics. Journal of the American Society for Information Science, 30(4), 205–214.

Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review, 13(5), 407–424.

Belkin, N., Cool, C., Stein, A., & Thiel, U. (1995). Cases, scripts, and information-seeking strategies: on the design of interactive information retrieval systems. Expert Systems with Applications, 9(3), 379–395.

Bhavnani, S.K. (2001). Important cognitive components of domain-specific search knowledge. In E.M. Vorhees & D.K. Harman (Eds.), NIST Special Publication 500-250: The Tenth Text Retrieval Conference (TREC ’01) (pp. 571–578). Washington, DC: NIST.

Bhavnani, S.K. (2005). Why is it difficult to find comprehensive information? Implications of information scatter for search and design. Journal of the American Society for Information Science and Technology, 56(9), 989–1003.

Bhavnani, S.K., Bichakjian, C.K., Schwartz, J.L., Strecher, V.J., Dunn, R.L., Johnson, T.M., et al. (2002). Getting patients to the right healthcare sources: From real-world questions to strategy hubs. In Proceedings of the AMIA 2002 Symposium (AMIA ’02) (pp. 51–55).

Bhavnani, S.K., Bichakjian, C.K., Johnson, T.M., Little, R.J., Peck, F.A., Schwartz, J.L., et al. (2003). Strategy hubs: Next-generation domain portals with search procedures. In Proceedings of the Conference on Human Factors in Computing Systems (CHI ’03) (pp. 393–400).

Bichakjian, C., Schwartz, J., Wang, T., Hall J., Johnson, T., & Biermann, S. (2002). Melanoma information on the Internet: Often incomplete-a public health opportunity? Journal of Clinical Oncology, 20, 1, 134–141.

Bradford, S.C. (1948). Documentation. London: Crosby Lockwood.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual (Web) search engine. Computer Networks, 30, 107–117.

Bush, V. (1945). As We May Think. The Atlantic Monthly, 6, 1, 101–108.

Canfield, M. (1972). Library pathfinders. Drexel Library Quarterly, 8(3), 287–300.

Chi, E., Pirolli, P., & Pitkow, J. (2000). The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a Web site. In Proceedings of the Conference on Human Factors in Computing Systems (CHI ’00) (pp. 161–168).

Crowston, K., & Williams, M. (1997). Reproduced and emergent genres of communication on the World-Wide Web. In R.H. Sprague (Ed.), Proceedings of the Thirtieth Hawaii International Conference on System Sciences, Volume VI (pp. 30–39). Washington, DC: IEEE Computer Society Press.

Drabenstott, K. (2000). Web search strategies. In W.J. Wheeler (Ed.), Saving the User’s time through subject access innovation: Papers in honor of Pauline Atherton Caochrane (pp. 114–161). Champaign, IL: University of Illinois.

Dumais, S., Cutrell, E., & Chen, H. (2001). Optimizing search by showing results in context. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’01) (pp. 272–284).

Dumais, S.T., Furnas, G.W., Landauer, T.K. & Deerwester, S. (1988). Using latent semantic analysis to improve information retrieval. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’88) (pp. 281–285).

Egan, D.E., Remde, J.R., Landauer, T.K., Lochbaum, C.C., & Gomez, L.M. (1989). Behavioral evaluation and analysis of a hypertext browser. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Wings for the Mind (CHI ’89) (pp. 205–210).

Eysenbach, G., & Köhler, 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, British Medical Journal, 324, 573–577.

Fidel, R. (1991). Searcher’s selection of search keys. Journal of the American Society for Information Science, 42(7) 490–500.

Florance, V., & Marchionini, G. (1995). Information processing in the context of medical care. In Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ’95) (pp. 158–163).

Fox, S., & Fallows, F. (2003). Health searches and email have become more commonplace, but there is room for improvement in searches and overall Internet access. Pew Internet and American live project: Online life report. Retrieved July 21, 2005, from http://www.pewinternet.org/report_display.asp?r=95

Fox, S., & Rainie, L. (2000). The online health care revolution: How the Web helps Americans take better care of themselves. Pew Internet and American live project: Online life report. Retrieved July 21, 2005, from http://www.pewinternet.org/report_display.asp?r=26

Furnas, G.W., Landauer, T.K., Gomez, L.M., & Dumais, S.T. (1987). The vocabulary problem in human-system communication. Communications of the Association for Computing Machinery, 30(11), 964–971.

Halasz, F., Moran, T., & Trigg, R. (1987). NoteCards in a nutshell. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '87) (pp. 42–52).

Hinds, C., Streater, A., & Mood, D. (1995). Functions and preferred methods of receiving information related to radiotherapy: Perceptions of patients with cancer. Cancer Nursing, 18(5), 374–384.

Holscher, C., & Strube, G. (2000). Web search behavior of internet experts and newbies. Computer Networks, 33(1), 337–346.

Hood, W., & Wilson, C. (2001). The scatter of documents over databases in different subject domains: How many databases are needed? Journal of the American Society for Information Science, 52(14), 1242–1254.

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

Johnson, T.M (2003). Excellence in oncologic care: Passion, priority, and performance. Archives of Dermatolology, 139, 1130–1132.

Kirk, T. (1974). Problems in library instruction in four-year colleges. In: J. Lubans, Jr. (Ed.), Educating the library user (pp. 83–103). New York: R.R. Bowker.

Lazonder, A.W., Biemans, H.J.A., & Wopereis, I.G.J.H. (2000). Differences between novice and experienced users in searching information on the world wide web. Journal of the American Society for Information Science, 51(6), 576–581.

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. Journal of the American Society for Information Science, 40(1), 54–66.

Mills, M., & Sullivan, K. (1999). The importance of information giving for patients newly diagnosed with cancer: a review of the literature. Journal of Clinical Nursing, 8, 631–642.

O'Day, V., & Jeffries, R. (1993). Orienteering in an information landscape: How information seekers get from here to there. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '93) (pp. 438–445).

Peck, F.A., Bhavnani, S.K., Blackmon, M.H., & Radev, D.R. (2004). Exploring the use of natural language systems for fact identification: Towards the automatic construction of healthcare portals. In Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIST '04) (pp. 327–338).

Pratt, W., Hearst, M., & Fagan, L. (1999). A knowledge-based approach to organizing retrieved documents. In Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI/IAAI '99) (pp. 80–85).

Ream E., & Richardson A. (1996). Fatigue: A concept analysis. International Journal of Nursing Studies, 33(5), 519–529.

Reeder, R., Pirolli, P., & Card, S.K. (2000). WebLogger: A data collections tools for web-use studies (UIR Technical report UIR-R-2000–06). Palo Alto, CA: Xerox PARC.

Russell, D.M., Stefik, M., Pirolli, P., & Card, S.K. (1993). The cost structure of sensemaking. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '93) (pp. 269–276).

Shneiderman, B., Byrd, D., & Croft, W.B. (1997). Clarifying search: A user interface framework for text searches. DLIB Mag, 3, 1. Retrieved July 21, 2005, from http://www.dlib.org/dlib/january97/retrieval/01schneiderman.htm

Shipman, F., Furuta, R., Brenner, D., Chung, C., & Hsieh, H. (2000). Guided paths through web-based collections: Design, experiences, and adaptations. Journal of the American Society of Information Science and Technology, 51(3), 260–272.

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

Sutcliffe, A.G., Ennis, M., & Watkinson, S.J. (2000). Empirical studies of end-user information searching. Journal of the American Society of Information Science and Technology, 51(13), 1211–1231.

Vakkari, P. (2002). Subject knowledge, source of terms and terms selection in query expansion: An analytical study. In Proceedings of the 24th European Colloquium on IR Research (pp. 110–123). Berlin: Springer.

Wildemuth, B.M. (2004). The effects of domain knowledge on search tactic formulation. Journal of the American Society of Information Science and Technology, 55, 246–258.

Wildemuth, B.M., do Bliek, R., Friedman, C.P., & File, D.D. (1995). Medical students' personal knowledge, search proficiency, and database use in problem solving. Journal of the American Society of Information Science and Technology, 46, 590–607.

Xie, H. (2000). Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. Journal of the American Society for Information Science, 51, 841–857.

Zellweger, P. (1987). Directed paths through collections of multi-media documents. In Proceedings of the ACM Conference on Hypertext (Hypertext '87) (unpublished position paper).

Zellweger, P. (1988). Active paths through multimedia documents. In J.C. van Vliet (Ed.), Proceedings of the International Conference on Electronic Publishing, Document Manipulation, and Typography (pp. 19–34). Cambridge, England: Cambridge University Press.

Zellweger, P. (1989). Scripted documents: A hypertext path mechanism. In Proceedings of the ACM Conference on Hypertext (Hypertext '89) (pp. 1–26).

DOI: 10.1002/asi

Appendix A

Generalizations

Search procedures to find comprehensive information

General to specific template
1. Get a general overview of relevant facts
2. Get specific details for particular facts

Basic to related template
1. Learn the basic concepts about the topic
2. Learn more general information about the topic

See box below

Specialization template
1. Learn basic concepts
2. Obtain general information
3. Obtain detailed information

Detection template
1. Understand typical symptoms, and non-symptoms
2. Learn how to do a self-examination
3. Find a doctor for a professional opinion

Estimation template
1. Understand the factors involved in an estimate, and how they affect the estimate
2. Calculate the estimate

Instantiations for melanoma information

Topic: Treatment-Conventional—Stage 0, I, & II
1. What are Stage 0, I, and II melanomas?
2. Learn about surgical treatment of Stage 0, I, and II melanomas
3. Learn about sentinel lymph node biopsy

Topic: Risk/Prevention-specific
1. What are the effects of ultraviolet radiation on the skin?
2. Learn about how to reduce exposure to ultraviolet radiation
3. Learn about the risks of tanning booths

Topic: Treatment-Conventional—Stage III
1. What is Stage III melanoma?
2. Learn about surgical treatment of Stage III melanoma
3. Learn about additional nonsurgical treatment of melanoma (interferon)

Topic: Treatment-Conventional—Stage IV
1. What is Stage IV melanoma?
2. Learn about treatment options for Stage IV melanoma
3. Learn about important supportive services

Topic: Diagnosis—Diagnostic tests
1. How is melanoma detected?
2. Learn about biopsies for melanoma
3. Learn about additional tests for melanoma

Topic: Risk/Prevention—Statistical
1. Understand general cancer statistics
2. Obtain general melanoma statistics
3. Obtain detailed melanoma statistics

Topic: Diagnosis—Doctor's-exam
1. When should I consult a doctor?
2. Learn about a doctor's examination of the skin, and how often to get one
3. Learn about different types of diagnostic procedures done in a doctor's exam

Topic: Treatment-Experimental—Clinical trial
1. What is a clinical trial?
2. Understand key issues related to taking part in a clinical trial
3. Find a clinical trial

Topic: Prognosis—Statistical
1. Understand general prognosis statistics for cancer
2. Obtain general prognosis statistics for melanoma
3. Obtain detailed prognosis statistics for melanoma

Topic: Diagnosis—Self-examination
1. Learn about the ABCDs of moles and melanoma
2. Understand the difference between moles and melanoma
3. Learn how and how often to do a self-skin examination
4. Locate a dermatologist near you

Topic: Diagnosis—Disease stage
1. Learn the basics of staging
2. Learn how melanoma is staged
3. Determine you stage of melanoma

Topic: Risk/Prevention—Descriptive
1. Learn about melanoma prevention
2. Learn about melanoma risk factors
3. Estimate your risk of melanoma

Topic: Prognosis—Descriptive
1. Learn about factors that influence the prognosis of melanoma
2. Learn how staging and other factors affect prognosis of melanoma
3. Estimate a prognosis for your melanoma

Elaboration template
1. Learn the definition or distinguishing feature of the main term
2. Learn related concepts relevant to the main term

Topic: Terminology—Definition
1. What is melanoma?
2. Learn about terms related to melanoma (i.e., lymph node, metastasis)

Topic: Terminology—Comparative
1. Learn about the difference between melanoma and other common skin cancers
2. Learn about other types of cancer (i.e., breast, lung)

Two of the above search procedures (Treatment-Conventional—Stages 0, I, and II, and Risk/Prevention—Specific) were originally under a template called the *Problem–Solution Template,* which was deleted after the experiment to achieve greater consistency in the taxonomy.

DOI: 10.1002/asi

Appendix B

Weighted criteria to evaluate answers to the treatment question.

	Criteria	Weighted importance (on a scale of 1–5)
1.	Treatment for melanoma is based on the stage [or level; or severity; or tumor size/thickness] [*not* type or form] of the disease	5
2.	There are 5 stages of melanoma	2.5
3.	Stage 0 is when the melanoma tumor is in the outer layer of skin only [or has not spread]	5
4.	Stage 0 is treated by surgical excision	5
5.	Patients with Stage 0 melanoma have a 5 year survival rate of 97–100% [any number between 97% and 100% is acceptable]	4
6.	Stage I is when the melanoma tumor is less than 1.5 mm thick	5
7.	Stage I is when the melanoma tumor has not spread beyond the skin [or has not spread to the lymph nodes]	5
8.	Stage I is treated by surgical excision [or surgery, or removal]	5
9.	Stage I patients might have a sentinel lymph node biopsy [or lymph nodes tested] [*not* all lymph nodes removed]	4
10.	Patients with Stage I melanoma have a 5 year survival rate of over 80% [any number over 80%]	3
11.	Stage II is when the melanoma tumor is greater than 1.5 mm thick.	5
12.	Stage II is when the melanoma tumor has not spread beyond the skin [or has not spread to the lymph nodes]	5
13.	Stage II is treated by surgical excision	5
14.	Stage II patients might have a sentinel lymph node biopsy [or lymph nodes tested] [*not* all lymph nodes removed]	5
15.	Patients with Stage II melanoma have a 5 year survival rate of 45–80% [any number between 45% and 80% is acceptable]	3
16.	Stage III is when the melanoma tumor has spread to the lymph nodes	5
17.	Stage III is treated by surgery to remove the melanoma from the skin and lymph nodes [or lymph node dissection]	5
18.	Stage III may also be treated by immunotherapy [or biological therapy, or vaccine therapy, or Interferon]	3
19.	Stage III may also be treated by radiation therapy	3
20.	Stage III may also be treated with clinical trials [or testing with humans]	3
21.	Patients with Stage III melanoma have a 5 year survival rate of 10–60% [any number between 10% and 60% is acceptable]	3
22.	Stage IV is when the melanoma tumor has spread to distant organs in the body	5
23.	Stage IV is treated by systematic therapy [i.e. chemotherapy]	4
24.	Stage IV may also be treated by radiation therapy	4
25.	Stage IV may also be treated by surgery	4
26.	Stage IV may also be treated with clinical trials [or testing with humans]	4
27.	Patients with Stage IV melanoma have a 5 year survival rate of 5–20% [any number between 5% and 20% is acceptable]	3
28.	Clinical trials test experimental treatment methods with human subjects [or human testing]	3
29.	Sentinel lymph node biopsy is used to determine whether the tumor has spread to the lymph nodes	4
30.	Immunotherapy [or biological therapy, or vaccine therapy, or Interferon, or cytokine] is given to boost your immune system [or make cancer cells more visible to T-cells] to treat melanoma	3
31.	Chemotherapy is an anti-cancer drug	3
32.	Radiation therapy is the use of high-energy rays [or x-rays] to kill the melanoma tumor	3
33.	Surgery is the removal of the melanoma from the part of the body that it has affected [or after surgery, the melanoma will be gone]	5

Weighted criteria to evaluate answers to the diagnosis question

	Criteria	Weighted importance (on a scale of 1–5)
1.	Self-examination is used to find potential melanoma tumors on the skin	3
2.	During a self examination, you should check the entire skin surface [or entire body] [or everywhere on body]	4
3.	During a self examination, you use the ABCDs [or ABCs; or size, shape, color] of melanoma to identify moles that might be melanoma	5
4.	In the ABCDs of melanoma, "A" stands for Asymmetry [also acceptable: look for asymmetrical moles; or look for moles where one half is different from the other half]	5
5.	In the ABCDs of melanoma, "B" stands for Border irregularity [Also acceptable: look for irregular border; or look for a scalloped border]	5
6.	In the ABCDs of melanoma, "C" stands for Color variance [or multi-colored, or 2 or more colors, or mentions multiple colors] [also acceptable: look for moles with color variance; or look for moles with 2 or more colors; or look for moles with multiple colors]	5
7.	In the ABCDs of melanoma, "D" stands for Diameter $>$ 6.0 mm [or width of a pencil eraser] [also acceptable: look for moles with diameter $>$ 6.0 mm; or look for moles with a width of a pencil eraser]	5
8.	During a self examination, you are looking for moles that have changed [or new moles that have appeared] since your last examination.	5
9.	During a self-examination, you are looking for itching [or bleeding; or tender] moles	5
10.	Experts recommend that you should perform a self examination every month	2
11.	A doctor's [or health care professional's; or nurse practitioner's] examination determines if a biopsy should be done to test a mole for melanoma.	4
12.	During a doctor's [or health care professional's; or nurse practitioner's] examination, the doctor will examine the entire skin surface [or entire body; or everywhere on body] of the patient	3
13.	A sentinel lymph node biopsy determines if melanoma has spread to the lymph nodes	3

DOI: 10.1002/asi