ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Int. J. Human-Computer Studies 59 (2003) 755–776

International Journal of Human-Computer Studies

www.elsevier.com/locate/ijhcs

Testing a Cancer Meta Spider

Hsinchun Chen*, Haiyan Fan, Michael Chau, Daniel Zeng

Department of Management Information Systems, The University of Arizona, 1130 E. Helen St., Room 430Z, Tucson, AZ 85721, USA

Received 19 March 2002; received in revised form 9 January 2003; accepted 2 June 2003

Abstract

As in many other applications, the rapid proliferation and unrestricted Web-based publishing of health-related content have made finding pertinent and useful healthcare information increasingly difficult. Although the development of healthcare information retrieval systems such as medical search engines and peer-reviewed medical Web directories has helped alleviate this information and cognitive overload problem, the effectiveness of these systems has been limited by low search precision, poor presentation of search results, and the required user search effort. To address these challenges, we have developed a domain-specific meta-search tool called Cancer Spider. By leveraging post-retrieval document clustering techniques, this system aids users in querying multiple medical data sources to gain an overview of the retrieved documents and locating answers of high quality to a wide spectrum of health questions. The system presents the retrieved documents to users in two different views: (1) Web pages organized by a list of key phrases, and (2) Web pages clustered into regions discussing different topics on a two-dimensional map (self-organizing map). In this paper, we present the major components of the Cancer Spider system and a user evaluation study designed to evaluate the effectiveness and efficiency of our approach. Initial results comparing Cancer Spider with NLM Gateway, a premium medical search site, have shown that they achieved comparable performances measured by precision, recall, and *F*-measure. Cancer Spider required less user searching time, fewer documents that need to be browsed, and less user effort.

*Corresponding author.

E-mail addresses: hchen@bpa.arizona.edu (H. Chen), fan@bpa.arizona.edu (H. Fan), mchau@bpa.arizona.edu (M. Chau), zeng@bpa.arizona.edu (D. Zeng).

1071-5819/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S1071-5819(03)00118-6

1. Introduction

In the healthcare profession, potentially significant decisions often depend on the availability of reliable and up-to-date information, although health-related data, especially those that are Web-based, are highly distributed, of varying quality, and difficult to locate. For instance, clinical information is often mingled with non-clinical information, and consumer information is undistinguished from clinical or research information (Hersh, 1996). Documents with different amounts of technical detail and varying quality are often mixed together in an unstructured way and it has become increasingly difficult to judge the quality and credibility of a piece of Web-based medical information. As a result, medical professionals and the general public increasingly experience the information and cognitive overload problem (Bowman et al., 1994) when seeking medical information.

The development of medical information retrieval (IR) systems such as medical search engines and peer-reviewed medical Web directories has helped alleviate this problem. However, the effectiveness and usefulness of these systems have been limited by low search precision and poor presentation of the retrieved documents. More specifically the following should be noted.

- Finding specific answers to user questions can be time-consuming and expensive, in part because of the amount of effort required to browse through large collections of returned documents and to identify the relevant ones. A search query in a general search engine like Google often returns thousands of results.
- Traditional search engines, including medical search engines, present search results as ranked lists, ordered by estimated relevance to the query. A major drawback of this presentation is that it fails to give users a quick overall "feel" for the retrieved documents and requires often significant manual browsing effort from users to locate documents of interest.

To address these problems with existing medical IR systems, we have developed Cancer Spider, a meta-search engine that performs post-retrieval document clustering and semantics-based visualization. In the post-retrieval phase of the system operation, we apply a linguistic-based noun phrasing technique to extract key concepts from documents, aiming to improve search precision. Semantics-based visualization using the Self-Organizing Map algorithm enables users to summarize easily the subject areas covered by the retrieved documents and navigate among them. From the user's perspective, Cancer Spider allows a user to easily access multiple medical literature databases, gain an overview of the retrieved documents, and to locate quality answers to a wide spectrum of health questions.

The rest of the paper is structured as follows. Section 2 surveys the current status of IR in the healthcare domain and some IR techniques, including meta-searching, document indexing, and post-retrieval clustering and visualization. In Section 3, we present the architectural design and major components of Cancer Spider. Section 4 discusses the design of a user study conducted to evaluate the proposed approach, and Section 5 reports and discusses the findings of this user study. We conclude the paper in Section 6 with a discussion of future research and system development.

2. Research backgrounds

2.1. Information retrieval in healthcare

Healthcare is an information-intensive business. Hersh (1996) classified textural health information into two main categories. The first category is patient-specific information. The second category is knowledge-based information, which can be further divided into the following three layers. Primary knowledge-based information contains original research reported in academic journals, books, technical reports, and other sources. Secondary knowledge-based information consists of indexes that catalog primary literature. The best known of these is *Index Medicus*. MEDLINE, a widely-used medical computer database, is the computerized version of *Index Medicus*. Tertiary literature consists of summaries of research in review articles, books and so on.

The data volume of these information resources is overwhelming. According to a *Time* magazine special issue on on-line health (March 2001), more than 26,000 Web sites provide health-related information. MEDLINE itself contains over 11 million references to journal articles in health-related sciences. LOCATOR*plus*, the on-line catalog of the National Library of Medicine (NLM), includes over 800,000 catalog records for books, audiovisuals, journals, computer files, and other materials in the Library's collections.

Kiley (1999) summarized two main ways of searching for medical information on the Web: (1) using a free-text search engine or Web directory service to query a database of Internet resources, and (2) browsing/searching through human-evaluated sources of information. The first approach refers either to browsing Web sites or to utilizing an Internet search tool. Some examples of reputable medical Web sites are *MayoClinic.com* (www.mayoclinic.com), *WebMD* (www.webmed.com), *Inteli-Health* (www.intelihealth.com), and *DrKoop* (www.drkoop.com). A few Web portals also have emerged as popular sources of health information. Among these are *Yahoo Health* pages (www.yahoo.com/Health) and *CBSHealthWatch* (www.cbshealthwatch.com). Kiley pointed out that the weakness of free-text searching lies in the indiscriminate method for document retrieval and that content is not evaluated professionally. Both of them lead to potential information overload and low information quality. An important consideration of IR in healthcare is information quality control and assurance. A substantial number of Web-based information sources are not suitable for direct clinical application because of poor data organization, questionable validity and uncertain reliability (Westberg and Miller, 1999). It also has been shown that many medical search engines return web pages that are not even related to healthcare related topics (Bin and Lun, 2001).

The second approach refers to evaluated subject catalogs or peer-reviewed Web directories, which deal with these information quality issues as they are compiled by domain experts and provide points of access to relevant and authoritative sources. Some well-known examples are NLM's MEDLINE-based literature databases; Medical Matrix (www.medicalmatrix.org), compiled by the Internet Working Group of the American Medical Informatics Association; Organising Medical Networked

Information, or OMNI (omni.ac.uk), created by a core team of information specialists and subject experts based at the University of Nottingham Greenfield Medical Library; and CliniWeb (www.ohsu.edu/cliniweb/) developed by the Oregon Health Sciences University.

Although evaluated subject catalogs have gained enthusiastic acceptance in practice, new problems have arisen with respect to the effectiveness of retrieving information from them. Studies have shown that MEDLINE and its derivatives can be helpful in answering clinical questions, but finding specific answers can be time-consuming, in part because of the manual effort required to browse through large collections of relevant publications (Westberg and Miller, 1999).

2.2. Meta-searching and search result filtering

Search engines, such as Google and AltaVista, have been developed to help users locate and retrieve relevant information from the Web. Most search engines rely on traditional IR approaches, such as the vector space model and the TF*IDF measure (Salton, 1989). However, because of the large size of the Web, a search engine can only index a portion of the entire Web due to limited computational resources. It has been shown that in 1999, the best search engine covered only about 16% of Web sites (Lawrence and Giles, 1999). By relying solely on one search engine, users could miss a significant number of documents they would find most relevant and no single search engine is likely to return more than 45% of relevant results (Selberg and Etzioni, 1995).

One approach to the problem is meta-searching. A meta-search system does not maintain its own search index; instead, it forwards search queries to several other search engines and then combines their search results, achieving higher coverage. MetaCrawler, which provides a single interface to allow users to search simultaneously from six different search engines, was the first meta-search system reported (Selberg and Etzioni, 1995, 1997). Other examples of meta-search systems include SavvySearch (Howe and Dreilinger, 1997) and Profusion (Gauch et al., 1996). Some meta-search systems, in addition to getting a list of URLs and summaries returned by other search engines, also fetch and analyse the documents in the result set. Inquirus, also known as the NECI meta-search engine, downloads the content of all result pages and generates a new summary of each page based on the search query. Pages which are no longer available (i.e. dead links) or do not contain the search terms are removed from the search results (Lawrence and Giles, 1998a, b). The MetaSpider system also performs filtering on the search result pages and such filtering has been demonstrated to improve search performance (Chau et al., 2001; Chen et al., 2001). One potential drawback with such approaches is that it may take up to a few minutes to download all the documents from the Web (Zamir and Etzioni, 1999; Chen et al., 2001).

2.3. Document indexing

Automatic indexing algorithms have been developed to extract key concepts from text and it has been shown that automatic indexing can be as effective as human indexing (Salton, 1986). Automatic indexing can be based on either single words or phrases. The extracted terms (words or phrases) are then used to form a vector to represent the document of interest, based on each term's frequencies or other scores. This method is referred to as the *vector space model* (Salton et al., 1975). The TF*IDF score, calculated by multiplying the term frequency by the inverse document frequency, is the most popular score used in the IR literature (Salton, 1989).

Single word indexing allows users to search for documents that contain the search keyword(s) and has been widely adopted in IR systems. Usually, stop-word removal and stemming are performed by IR systems before the indexing step. In stop-word removal, each word is checked against a pre-defined list of non-semantic bearing, high-frequency words such as *a*, *the* and *of*. These stop words are then removed from the index to reduce storage space and increase retrieval speed. Stemming is the process of removing the common morphological and inflectional endings from words (e.g., Lovins, 1968; Porter, 1980). For example, the words *computer*, *computers*, *computing*, *computation*, and *computational* will all be reduced to the same word stem *comput*. Using stemming, users do not need to enter the exact search keyword to retrieve documents that contain the keyword in another morphological or inflectional form. However, some researchers suggest that stemming does not improve retrieval performance in English and the usefulness of stemming is still under debate (Harman, 1991; Krovetz, 1993; Hull, 1996).

Phrase indexing also has been used in IR systems. In many cases, phrases can convey and represent more precise meaning than single words. For example, the phrase *lung cancer* is a term more specific than just the single-word term *cancer*. In term-based searching, phrases may lead to lower recall rate because users may need to enter the exact phrase that appears in a potentially relevant document in order to retrieve it. However, because the meaning of phrases is more specific and precise, phrase indexing is highly suitable for applications such as document clustering or automatic thesaurus generation. The Arizona Noun Phraser (AZNP) is an example of phrase extraction tools (Tolle and Chen, 2000). It extracts all the noun phrases from each document, based on part-of-speech tagging and linguistic rules. AZNP has three components. The *tokenizer* takes documents as text input and creates output that conforms to the UPenn Treebank word tokenization rules by separating all punctuation and symbols from text without interfering with textual content. The *tagger* module assigns every word in the document a part-of-speech designation based on the Brill's tagger (Brill, 1995). The last module, called the *phrase generation* module, converts the words and associated part-of-speech tags into noun phrases by matching tag patterns to a noun phrase pattern established by linguistic rules. For example, the phrase *preventative measurement* would be considered a valid noun phrase because it matches the rule that an *adjective–noun* sequence forms a noun phrase.

2.4. Document classification, clustering, and visualization

Manually browsing through Web pages to locate useful information can be mentally exhausting and time-consuming. To address this problem, much IR research has been devoted to developing techniques and tools to analyse and visualize large collections of Web documents in an automated or semi-automated manner (e.g., Zamir and Etzioni, 1999; Chen et al., 2001).

After key concepts in the form of words or phrases are extracted from documents, they can be used for further analysis such as document classification, document clustering, and visualization. Classification and clustering tools automatically group documents into different topics, which in turn can be visually presented to users to facilitate understanding. The difference between document classification and clustering is that while categories are predefined (usually manually) in document classification, they are generated dynamically and automatically in document clustering. Because categories are not predefined, document clustering can often result in more specific categories and descriptive labels related to user queries. In general, document classification, clustering, and visualization techniques can be used to help users better understand the retrieved document set, identify interesting documents more effectively, and gain a quick overview of the documents' contents.

Web document clustering techniques can be classified into two broad categories. The first approach aims to provide additional information about the retrieved documents, such as query-term distribution, document size, source, topic, author, etc., and to cluster documents according to those pre-defined attributes. For example, if query-term distribution is used as the clustering attribute, the algorithm can be applied to show how the retrieved documents relate to each term used in the query, and how the internal subtopic structure of the documents relates to the query (Veerasamy and Belkin, 1996; Hearst, 1995).

The second approach is based on inter-document similarities and attempts to reduce the multidimensional vector space (used to represent all related documents) to a two-dimensional or three-dimensional space by aggregating similar documents under the same theme, thus providing users with a quick overview of the entire collection. In this approach, cluster labels are determined based on the phrases that have appeared in the documents collected. This approach usually includes some machine learning components. For example, the Kohonen self-organizing map (SOM) approach clusters documents into different topics that are automatically constructed on-the-fly using neural network algorithms (Kohonen, 1995; Lin, 1997; Chen et al., 1996; Chen et al., 1998). These clusters then are mapped into different regions on a two-dimensional map displayed to the user. Each region contains similar documents, and regions that are conceptually related are located close to each other on the map. The Grouper system, which clusters search results from a meta-search engine using the suffix tree clustering algorithm, is another example of document clustering (Zamir and Etzioni, 1999).

Document clustering and visualization, although presenting great potential for improving user Web search experience, are relatively new to general users; very few commercial search engines have these capabilities. Most document clustering techniques discussed above are used only in prototype research systems. While the results are promising, it has been reported that training and system assistance may be needed to enhance the effectiveness of clustering and visual user interfaces for IR (Sutcliffe et al., 2000).

2.5. IR system evaluation

The evaluation of IR systems is of significant importance in healthcare applications. The goal of an evaluative study is to determine whether a system helps the users for whom it is intended (Hersh, 1996).

Relevance-based recall and precision are the most widely used performance measures in the IR literature. Recall (R) is the proportion of relevant documents retrieved from the data sources, while precision (P) is the fraction of the retrieved documents which are relevant to the user query. F-measure, a single measure that tries to balance recall and precision, is defined as

$$F = \frac{2 \times R \times P}{R + P}.$$

In our work, we adopt an evaluation method based on these three relevance measures as well as several new user-oriented measures proposed to capture users' cognitive effort needed in the information search process. A detailed description is presented in Section 4.6.

3. Cancer Spider system architecture

In this section, we present the architectural design of Cancer Spider, focusing both on its use as a meta medical search tool and as a document clustering and visualization tool. Although the current implementation of Cancer Spider focuses on cancer-related topics, it can be easily adapted for other medical areas, since the core technologies are domain-independent.

Intended end-users of Cancer Spider are cancer researchers, physicians and medical librarians. The design goal of Cancer Spider is to offer value-added capabilities that assist users in effectively weeding out irrelevant Web pages and locating useful information. Shneiderman (1997) developed a four-phase framework for information searching on the Web that guides the design and analysis of search systems. The four phases are (1) *formulation* (expressing the search), (2) initiating *action* (launching the search), (3) review of *results* (reading messages and outcomes) and (4) *refinement* (formulating the next step). Our tool is designed to assist users in the first three phases.

Cancer Spider was implemented based on the MetaSpider system developed in our previous research (Chen et al., 2001). The Cancer Spider system incorporates functions that have been shown useful in other information retrieval and Web retrieval systems. Such functions include meta-searching, document filtering, phrase indexing, and document clustering. The architecture of the Cancer Spider is shown in Fig. 1. The major functions of the Cancer Spider are: (1) searching, (2) filtering, (3) phrase selection, and (4) visualization using self-organizing map (SOM). We briefly discuss each functionality in the following sections. Readers are referred to Chen et al. (2001) for a more detailed technical discussion.

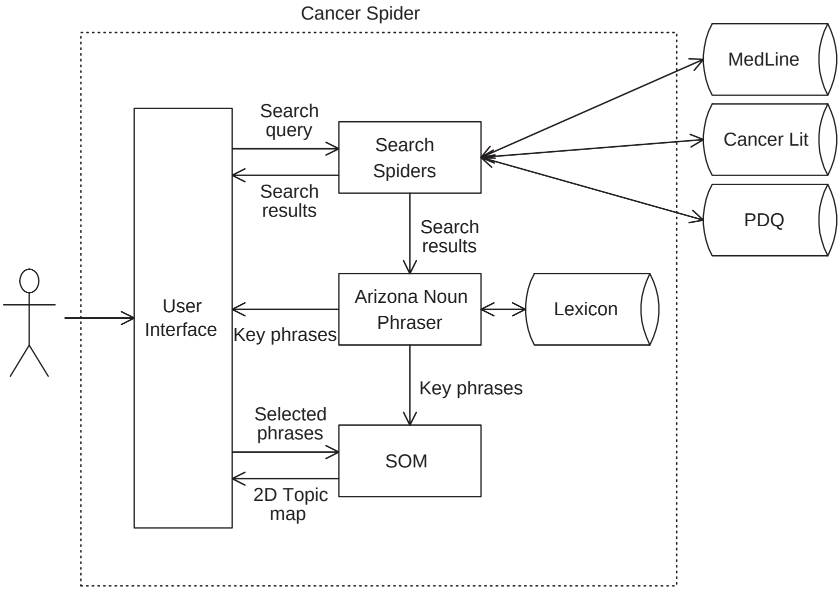


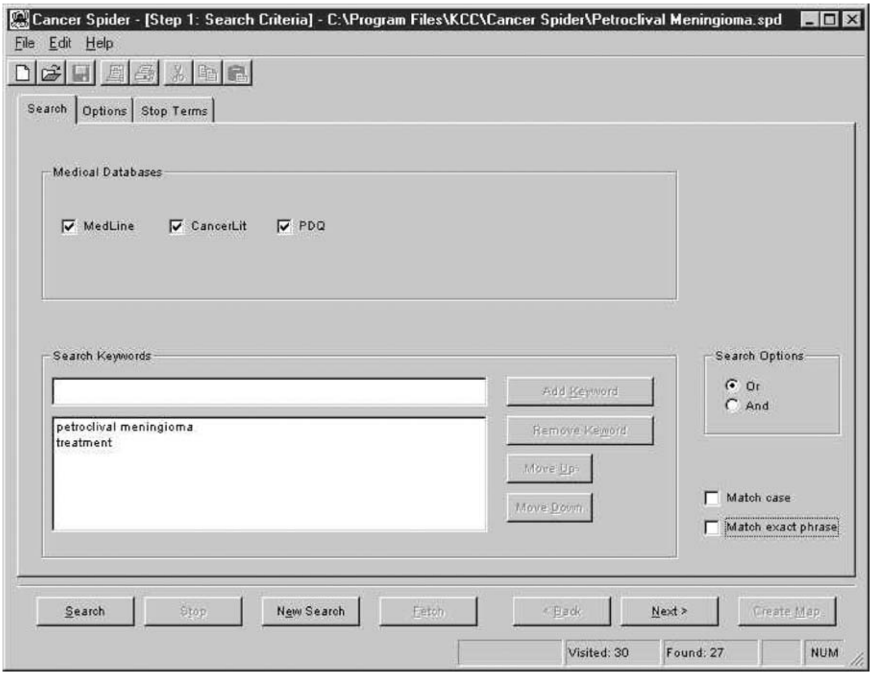
Fig. 1. Cancer Spider architecture.

3.1. Meta-searching

Meta-searching, which leverages the capabilities of multiple Web search engines and other types of information sources, provides a simple, uniform user interface that promises significant advances in coping with information overload and low-precision issues (Selberg and Etzioni, 1997; Chen et al., 2001). Meta-search engines can improve search performance by sending queries to multiple search engines or other data sources and collating only the highest-ranking subset of the returns from each data source.

Cancer Spider provides meta-searching capabilities and uses multiple data sources to locate documents that are relevant to the user query. The information search part of Cancer Spider's user interface is comprised of three panels. The *Search* panel provides users with an array of medical literature databases to choose from, as shown in Fig. 2. Users can customize their searches on the *Options* panel and specify stop-terms on the *Stop Terms* panel.

Currently Cancer Spider connects to three databases: MEDLINE, CANCERLIT and PDQ. MEDLINE, or Medical Literature, Analysis, and Retrieval System Online (www.ncbi.nlm.nih.gov/entrez/query.fcgi) is the National Library of Medicine's (NLM) main bibliographic database that contains over 11 million references to journal articles in life sciences with a concentration on biomedicine. Published by the National Cancer Institute (NCI), CANCERLIT (www.cancer.gov/search/cancer_literature/) contains references to the vast realm of cancer literature published from the 1960s to the present. The PDQ (Physician Data Query) database (www.nci.nih.gov/cancerinfo/pdq/cancerdatabase/), NCI's comprehensive cancer



Cancer Spider - [Step 1: Search Criteria] - C:\Program Files\KCC\Cancer Spider\Petroclival Meningioma.spd

Medical Databases

☑ MedLine ☑ CancerLit ☑ PDQ

Search Keywords

petroclival meningioma
treatment

Search Options

● Or
○ And

□ Match case
□ Match exact phrase

Visited: 30 Found: 27 NUM

Fig. 2. Cancer Spider user interface.

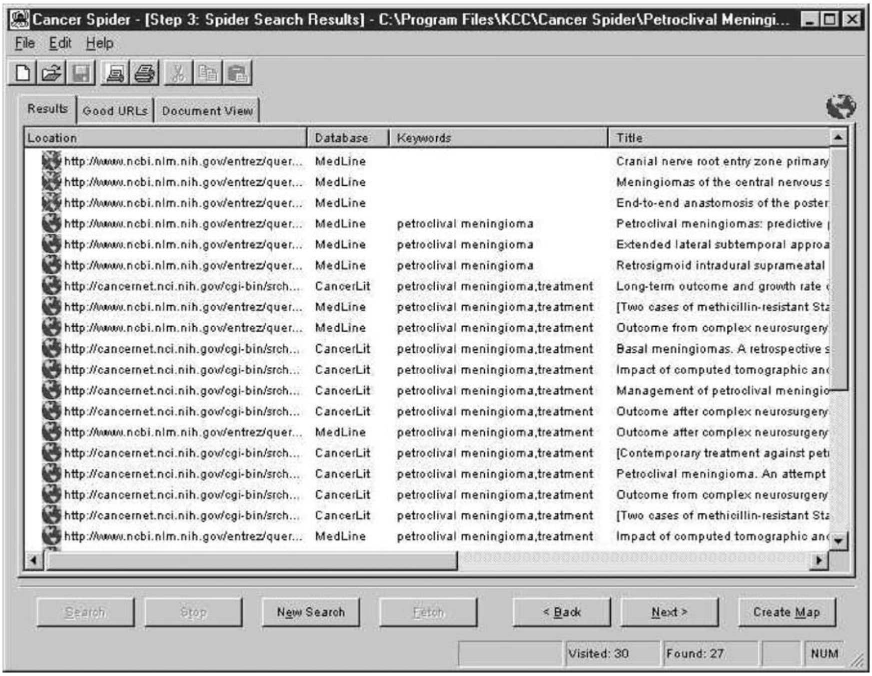
database, contains peer-reviewed summaries on cancer treatment, screening, prevention, genetics, and supportive care.

3.2. Search result filtering

Cancer Spider sends out queries to the multiple data sources selected by the user and collects the top *n* results, *n* being specified by the user at the beginning of the search. Unlike other meta-search tools, which show only the URLs and page summaries to the user, Cancer Spider will fetch the full text of the URLs returned by the underlying data sources and perform post-retrieval filtering and analysis.

On the *Search Result* panel illustrated in Fig. 3, the user can see each returned document's URL address, its ranking in the original data source, and the title of the document. Cancer Spider performs a weeding routine to eliminate "poor" pages from the returned document set. Poor pages are those that are no longer available or do not contain the search term(s) anymore. The ranking of the remaining pages is based on the relevance scores provided by the source databases; no re-ranking operations are performed by the Cancer Spider system.

The user can sort the document list so that documents containing more search terms are grouped together and presented first. These search terms provide the most convenient and direct way for users to evaluate the relevance of the documents in the list. With the help of document titles, users can access the documents easily at this stage.


Cancer Spider - [Step 3: Spider Search Results] - C:\Program Files\KCC\Cancer Spider\Petroclival Meningi...
File
Edit
Help
Results
Good URLs
Document View
Search
Stop
New Search
Fetch
< Back
Next >
Create Map
NUM

Location	Database	Keywords	Title
http://www.ncbi.nlm.nih.gov/entrez/quer...	MedLine		Cranial nerve root entry zone primary
http://www.ncbi.nlm.nih.gov/entrez/quer...	MedLine		Meningiomas of the central nervous s
http://www.ncbi.nlm.nih.gov/entrez/quer...	MedLine		End-to-end anastomosis of the poster
http://www.ncbi.nlm.nih.gov/entrez/quer...	MedLine	petroclival meningioma	Petroclival meningiomas: predictive
http://www.ncbi.nlm.nih.gov/entrez/quer...	MedLine	petroclival meningioma	Extended lateral subtemporal approa
http://www.ncbi.nlm.nih.gov/entrez/quer...	MedLine	petroclival meningioma	Retrosigmoid intradural suprameatal
http://cancernet.nci.nih.gov/cgi-bin/srch...	CancerLit	petroclival meningioma,treatment	Long-term outcome and growth rate
http://www.ncbi.nlm.nih.gov/entrez/quer...	MedLine	petroclival meningioma,treatment	[Two cases of methicillin-resistant Sta
http://www.ncbi.nlm.nih.gov/entrez/quer...	MedLine	petroclival meningioma,treatment	Outcome from complex neurosurgery
http://cancernet.nci.nih.gov/cgi-bin/srch...	CancerLit	petroclival meningioma,treatment	Basal meningiomas. A retrospective s
http://cancernet.nci.nih.gov/cgi-bin/srch...	CancerLit	petroclival meningioma,treatment	Impact of computed tomographic and
http://cancernet.nci.nih.gov/cgi-bin/srch...	CancerLit	petroclival meningioma,treatment	Management of petroclival meningio
http://cancernet.nci.nih.gov/cgi-bin/srch...	CancerLit	petroclival meningioma,treatment	Outcome after complex neurosurgery
http://www.ncbi.nlm.nih.gov/entrez/quer...	MedLine	petroclival meningioma,treatment	Outcome after complex neurosurgery
http://cancernet.nci.nih.gov/cgi-bin/srch...	CancerLit	petroclival meningioma,treatment	[Contemporary treatment against pet
http://cancernet.nci.nih.gov/cgi-bin/srch...	CancerLit	petroclival meningioma,treatment	Petroclival meningioma. An attempt
http://cancernet.nci.nih.gov/cgi-bin/srch...	CancerLit	petroclival meningioma,treatment	Outcome from complex neurosurgery
http://cancernet.nci.nih.gov/cgi-bin/srch...	CancerLit	petroclival meningioma,treatment	[Two cases of methicillin-resistant Sta
http://www.ncbi.nlm.nih.gov/entrez/quer...	MedLine	petroclival meningioma,treatment	Impact of computed tomographic anc

Visited: 30 Found: 27

Fig. 3. Cancer Spider search results.

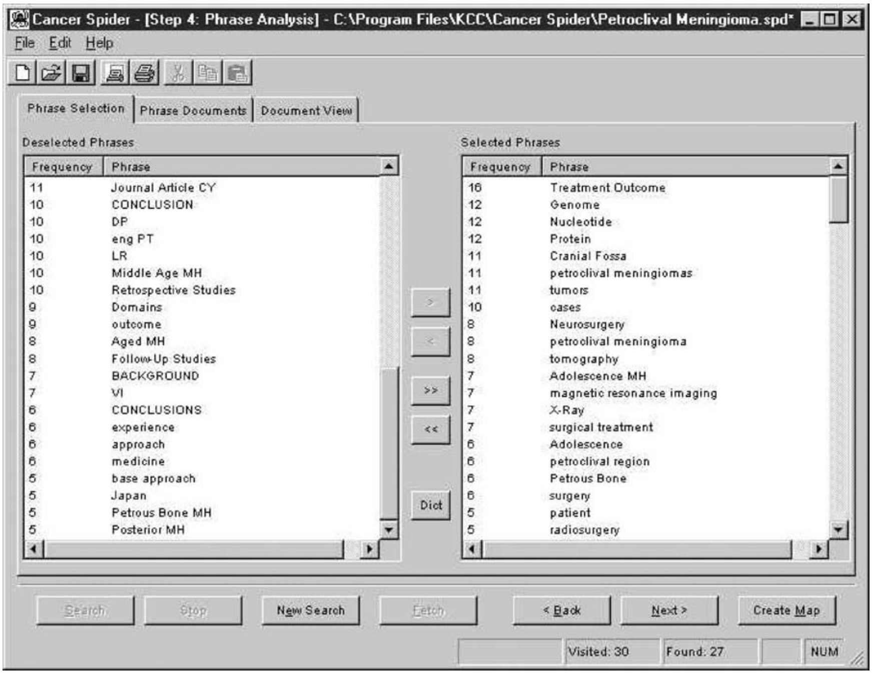
3.3. Phrase selection

The Arizona Noun Phraser (AZNP) discussed earlier is used in Cancer Spider to extract the key phrases that appear in the documents retrieved and filtered by the system. The frequencies of occurrences of the phrases are recorded and sent to the User Interface.

Fig. 4 presents the list of all noun phrases extracted from the documents retrieved. Clicking on any phrase, the user can view from the entire document set a list of documents that contain the phrase. Grouping documents by key noun phrases helps the user narrow down search results within the broader search context. For example, Fig. 4 corresponds to a search scenario where the user enters *petroclival meningioma* and *treatment* as the search terms. Noun phrases such as *Treatment Outcome*, *Neurosurgery*, and *Tomography* are then identified by AZNP as related concepts. Clicking on any of the noun phrases, the subset of the documents relating to that specific concept will be presented to the user.

3.4. Visualization using self-organizing map (SOM)

In order to give users an overview of the set of documents collected, Cancer Spider employs the self-organizing map (SOM) to automatically cluster the Web pages



Deselected Phrases

Frequency	Phrase
11	Journal Article CY
10	CONCLUSION
10	DP
10	eng PT
10	LR
10	Middle Age MH
10	Retrospective Studies
9	Domains
9	outcome
8	Aged MH
8	Follow-Up Studies
7	BACKGROUND
7	VI
6	CONCLUSIONS
6	experience
6	approach
6	medicine
5	base approach
5	Japan
5	Petrous Bone MH
5	Posterior MH

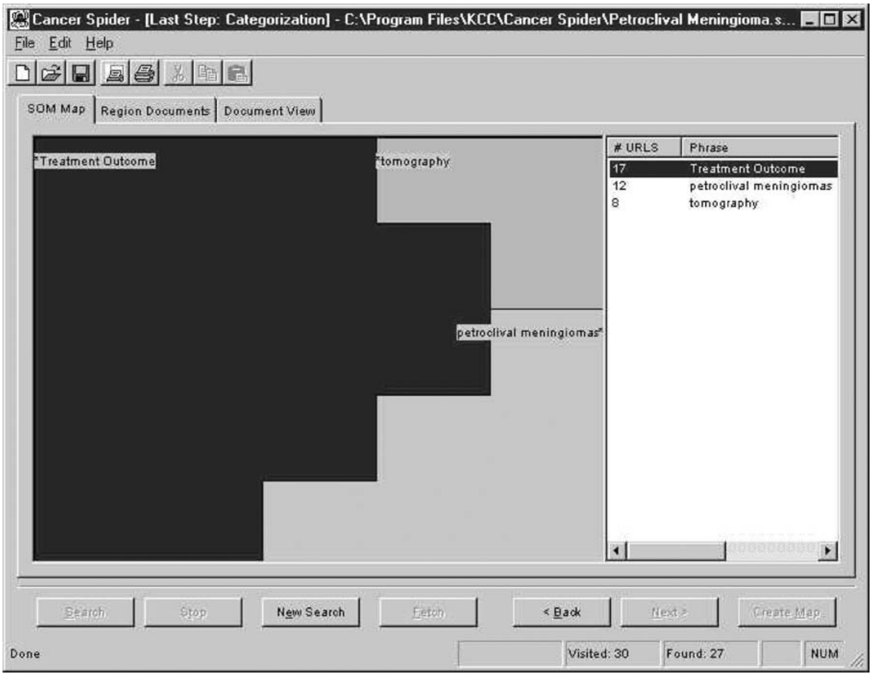
Selected Phrases

Frequency	Phrase
16	Treatment Outcome
12	Genome
12	Nucleotide
12	Protein
11	Cranial Fossa
11	petroclival meningiomas
11	tumors
10	cases
8	Neurosurgery
8	petroclival meningioma
8	tomography
7	Adolescence MH
7	magnetic resonance imaging
7	X-Ray
7	surgical treatment
6	Adolescence
6	petroclival region
6	Petrous Bone
6	surgery
5	patient
5	radiosurgery

Fig. 4. Cancer Spider phrase selection.

collected into different regions on a two-dimensional map (Fig. 5). The SOM algorithm creates an intuitive, graphic display of important concepts contained in textual information (Lin et al., 1991; Orwig et al., 1997). Each document is represented as an input vector of noun phrases extracted by AZNP and a two-dimensional grid of output nodes is created (Chen et al., 1998). After the network is trained through repeated presentation of all inputs, each region in the map is labeled by the phrase best describing the key concept most representative of the cluster of documents in that region (e.g. *treatment outcome*). The size of the color block indicates the relative significance of the term to the documents collected. The relative proximity reveals the distance between the two concepts presented by the respective term. The map helps users understand the topics related to the search query by providing an overview of the Web pages retrieved.

Sophisticated users can tailor the Noun Phraser by deselecting some of the trivial phrases to make the most sense of the map. For example, on the initial map created, frequently appearing terms such as *treatment outcome* and *petroclival meningioma* take up large areas. To allow other words or concepts to be seen, users can go back to the Noun Phraser page and deselect such frequently appearing terms to permit inclusion of terms such as *CT scan* and *gamma rays* as shown in Fig. 6. Users can click on any of the color blocks to go to the list of Web sites that contain the corresponding terms and phrases.



Cancer Spider - [Last Step: Categorization] - C:\Program Files\KCC\Cancer Spider\Petroclival Meningioma.s...

# URLS	Phrase
17	Treatment Outcome
12	petroclival meningiomas
8	tomography

Visited: 30
Found: 27

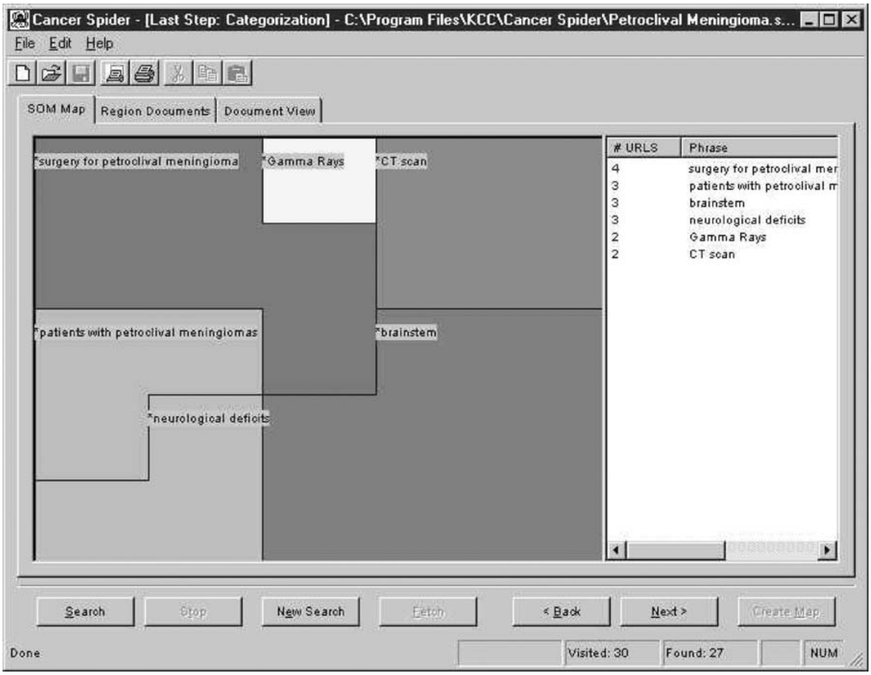
Fig. 5. Documents clustered into different categories in SOM.

4. Evaluation methodology

4.1. Comparison base

We conducted a user study to evaluate the proposed approach of meta-searching, clustering, and visualization implemented in the Cancer Spider system. In our experiment, Cancer Spider was compared with NLM Gateway (http://gateway.nlm.nih.gov/gw/Cmd?GMBasicSearch), the portal search engine to NLM's multiple literature databases.

NLM Gateway is a Web-based system that lets users search simultaneously multiple retrieval systems at the US National Library of Medicine (NLM) through a unified Web interface. The Gateway connects users with multiple NLM retrieval systems including MEDLINE/PubMed, OLDMEDLINE, LOCATOR*plus*, AIDS Meetings, HSR Meetings, HSRProj, MEDLINE*plus*, and DIRLINE. Released to the public in October 2000, NLM Gateway provides advanced search capabilities such as terminology suggestion and search history archiving. For users who are not familiar with medical terminology, the system is particularly helpful in providing definitions and related terms for the targeted search, based on NLM's MeSH (Medical Subject Heading) and the UMLS (Unified Medical Language Systems) Metathesaurus. The purpose of the UMLS is to develop knowledge sources that can be used by a wide variety of applications programs to overcome retrieval problems



Cancer Spider - [Last Step: Categorization] - C:\Program Files\KCC\Cancer Spider\Petroclival Meningioma.s...

# URLS	Phrase
4	surgery for petroclival mer
3	patients with petroclival m
3	brainstem
3	neurological deficits
2	Gamma Rays
2	CT scan

Done | Visited: 30 | Found: 27 | NUM

Fig. 6. Iterative customization of SOM.

caused by differences in terminology and the scattering of relevant information across many databases.

4.2. Theme-based evaluation

Because Cancer Spider has been designed to facilitate and integrate both document retrieval and automated clustering, traditional evaluation methodologies that treat document retrieval and clustering completely separately are not directly applicable. We have developed a new evaluation framework based on theme identification. Instead of asking the subject to find specific information, we give the subject relatively open-ended information search tasks and ask him or her to summarize search results in the form of a number of themes. This is similar to the soft queries created by the National Institute of Standards and Technology (NIST) for the TREC-6 ad hoc task. The TREC (Text Retrieval Conference) series is sponsored by the NIST and the Defense Advanced Research Projects Agency (DARPA) to encourage research in information retrieval from large text collections (Voorhees and Harman, 1998).

In our experiment, a theme was defined as "a short phrase that summarizes a specific aspect of the search results." Phrases like *stereo tactic radio surgery such as gamma knife* and *diagnostic and therapeutic use of radioactive iodine as treatment* are examples of such themes. Within this theme-based framework, we have designed protocols to permit evaluation of the extent to which combined document retrieval and clustering facilitate users' identification of major themes related to a certain topic. Section 4.5 presents a detailed discussion of how we used this theme-based framework in our user study.

4.3. Experiment hypotheses

Our overall research goal is to examine how effectively and efficiently Cancer Spider supports users in locating useful information and understanding retrieved documents. The specific hypotheses examined in our user study are the following.

H1 Compared to NLM Gateway, Cancer Spider achieves higher precision when assisting the user in theme identification.
H2 Compared to NLM Gateway, Cancer Spider achieves higher recall when assisting the user in theme identification.
H3 Compared to NLM Gateway, Cancer Spider achieves better overall performance as evaluated by *F*-measure.
H4 Compared to NLM Gateway, Cancer Spider requires less time for users to locate useful information and to understand retrieved documents.
H5 Compared to NLM Gateway, Cancer Spider requires less manual browsing effort from its users.

4.4. Experiment tasks

In our user study, we used six search questions selected from a list of over 100 real-world research questions compiled by medical librarians at the Arizona Health Sciences Library. These questions include questions asked by library users at the reference desk at the medical library, as well as questions raised by physicians or cancer researchers in their practice. The six search questions used in our experiment were as follows.

1. What role do antioxidants play in reducing cancer risk and which neoplasms are most affected?
2. How can chemotherapy be used to cure ALL (acute lymphocytic leukemia)?
3. What are the treatment options for petroclival meningioma?
4. What are the connections between breast cancer and iodine?
5. What are the symptoms of pancreatic neoplasm?
6. What is the role of folic acid in the prevention of colon cancer?

These questions were chosen because they allowed subjects to explore a given topic, as opposed to locating specific answers for a narrow question. They represent a balanced combination of general, exploratory, and focused searching. From the perspective of cancer medicine, these questions represent a good mixture of both clinically oriented questions and research questions at the molecular level.

4.5. Experimental subjects and expert evaluators

Thirty cancer researchers, including graduate students of cancer-related majors, medical students, and lab technicians from the Arizona Cancer Center were recruited to participate in the experiment. This subject body represented a good combination of cancer research and clinical background. Subjects were assigned information search tasks and required to jot down the themes they had identified after searching on a given IR system. For each IR system, each subject was assigned a search task and was instructed to perform a search on the system. To avoid a potential fatigue effect, we rotated the order in which each IR system was evaluated. Although subjects were not given a specific time frame within which to perform the searches, they were encouraged to stop after 20 min (most subjects took less than 20 min to finish the tasks).

Two senior Ph.D. students at the Arizona Cancer Center majoring in Cancer Biology and Molecular and Cellular Biology, respectively, were recruited as content experts to evaluate all subjects' answers. These two experts first performed all six searches on the Web and identified themes for each search task independently. Then they compared their search results with each other's and worked together to come up with a final standard answer set, against which all subjects' answers were evaluated. To ensure the comprehensiveness of the standard answer set, these two experts were not restricted to using only the two IR systems under investigation but could choose to utilize any type of available information resources they felt most suitable for answering a particular question. Finally, the experts worked together to evaluate all subjects' answers.

It is worth mentioning that the evaluation was based on semantic meanings rather than exact phrasing of the answers. For example, Question 1 was "What role do antioxidants play in reducing cancer risk and which neoplasms are most affected?" The experts recognized that antioxidants protect cells from varying kinds of oxidative damage. Answers that mentioned only protection of lipid oxidation were recognized as a valid theme, since lipid oxidation is one type of damage that can be prevented by antioxidants. However, answers that listed several types of oxidative damage, e.g. lipid per oxidation, protein cross-linking and DNA modification, each as a separate theme, were considered to be one valid theme, as all these represent different forms of oxidative damage preventable by antioxidants.

4.6. Experiment measurements

We collected and examined both quantitative and qualitative data in our study. For quantitative data, our primary interests were in the performance and efficiency of the IR systems under investigation. Performance was evaluated by theme-based precision, recall and F-measure, whereas efficiency was measured by searching time, precision effort and the total number of documents browsed.

Precision rate was computed as the number of correct themes identified by the subject divided by the total number of themes in the subject's answer set. Recall rate was calculated as the number of correct themes identified by the subject divided by the total number of themes in the standard answer compiled by the experts. Searching time was recorded as number of minutes spent on the searching, including both the response time of the system and subjects' browsing time. Time elapsed while subjects wrote their answers on the answer sheet were not included.

Precision effort, a new user-oriented measure we have developed, is defined as the ratio between the number of relevant documents the user found helpful and the number of documents examined in an attempt to find the useful documents. The number of documents browsed records the number of articles subjects viewed in a search session seeking answers to the search questions.

Qualitative data were drawn from user search logs and questionnaires. The search log created for each subject recorded major observations of user behaviors, as well as the user's *think aloud* discourse during searching. To enable qualitative comparison of the two systems, subjects were required to fill out questionnaires at the end of their search sessions. The questionnaire investigated subject's experiences on five different dimensions: (1) user interface, (2) usefulness of the information retrieved in answering the search questions, (3) subjects' level of certainty about their answers, (4) user satisfaction of the search experience and (5) the amount of knowledge obtained after the search.

5. Experimental results and analysis

5.1. Performance

Measured by theme-based precision and recall, the performances of the two IR systems, Cancer Spider and NLM Gateway were comparable. The main statistics are summarized in Table 1. Both the mean precision and the mean recall of Cancer Spider were comparable with those of NLM Gateway. On average, NLM Gateway seemed to be slightly better than Cancer Spider, but the difference was not statistically significant, as suggested by the p-value of the pair-wise t-test. The F-measure, which was the combined measure of precision and recall, also reveals the same result. These findings indicated that hypotheses H1, H2, and H3 were not confirmed.

The relative comparability of the two IR systems can be explained by the fact that each system has its unique, differentiating features. First, NLM Gateway, being the portal site to many NLM administered medical literature databases, has the advantage of comprehensive data coverage (higher recall rate). In comparison, the current version of Cancer Spider connects to only three databases, considerably fewer than NLM Gateway. On the other hand, the vast amount of data available in NLM Gateway may result in lower search precision, more search time and more browsing effort.

Second, the use of UMLS Metathesaurus in NLM Gateway greatly improved the system's usability and performance. Of the 30 subjects in our experiment, 20 utilized UMLS Metathesaurus to find related terminology and perceive its usefulness

Table 1
System performance

	Sample size	Cancer Spider		NLM Gateway		*p*-Value of *t*-test
		Mean	Variance	Mean	Variance	
Precision	30	0.803	0.117	0.826	0.122	0.7572
Recall	30	0.533	0.112	0.539	0.121	0.9523
F-measure	30	0.612	0.105	0.622	0.110	0.9056

favorably. Cancer Spider does not provide domain-specific, high quality ontological assistance to its users.

Third, the traditional Web-enabled interface of NLM Gateway provides a user-friendly navigating environment. The document summary panel presents to the users not only full-text titles, but also authors, resources and publication dates of the retrieved articles. In contrast, Cancer Spider, as a research prototype, provides limited graphical options and the navigating function is different from typical Web-based interfaces. This unfamiliar user interface and lack of certain contextual information (e.g. Cancer Spider does not show author and other meta-level resource information to the user) may have affected searching performance.

Despite these disadvantages, Cancer Spider's performance was comparable to that of NLM Gateway. We hypothesize that this was largely due to Cancer Spider's value-added capabilities in post-search processing and clustering. First of all, we found that users liked the key phrase extraction feature of Cancer Spider. This feature provides immediate feedback to the user regarding whether the given document contains the search terms or not. As the system supports multiple search terms, the user can first roughly judge the relevance of a returned article by eyeballing the number of search terms it contains. Then, by clicking on the *Rank* bar on the top of the *Search* panel, the user can rank all the returned documents sorted by the number of search terms they contain.

Second, the clustering function of Cancer Spider helps users narrow down the search scope by focusing on the noun phrases (key topics) in which a user is interested. At the *Phrase Selection* stage, the user can click on any noun phrases of interest that have been extracted from the full-text documents to find a subset of articles that focus on the particular topic. For example, one subject saw the value of the Cancer Spider clustering function when working on the topic *using aspirin in skin cancer prevention*. He was quoted as saying "it helps me to focus on different aspects of the topic, e.g. how aspirin is related to this particular gene in the molecular pathway."

Third, the interaction between the system and users is comparatively active and dynamic, unlike traditional search services in which users passively take whatever the system generates. Subjects indicated that they liked the flexible, interactive design of Cancer Spider, which gives them a sense of "user-in-control." It has been shown that users perform better and have increased subjective satisfaction when they can view and control the search process (Koenemann and Belkin, 1996). This benefit is illustrated by the following comment made by one of the subjects: "I like the system because I have the control over how the documents can be arranged by keywords, as opposed to other search engines, where the user cannot manipulate the result." At the searching stage, users can rank the returned documents by the number of search terms contained so that they can directly visit those most relevant without sifting blindly through document summaries. At the clustering stage, users can sort noun phrases by the frequency of their appearance in the returned documents or simply in the alphabetic order. Such handy tools give users a sense of control and are helpful in assisting them to focus their search effort.

5.2. Efficiency

In our experiment, we used search time, the number of documents browsed, and precision effort to measure the efficiency of the two medical IR systems under investigation. The key experimental findings are summarized in Table 2.

5.2.1. Time

The mean search time for Cancer Spider (10.22 min) was significantly shorter than that of NLM Gateway (14.00 min), with a *p*-value of 0.0003. This has confirmed hypothesis H4 at the 1% significance level that Cancer Spider requires less time than NLM Gateway to locate useful information and to understand retrieved documents.

We believe a main reason for this result is that Cancer Spider provides document clustering. First of all, based on search terms, Cancer Spider filters out all irrelevant documents. Second, by grouping a large set of retrieved documents into small subtopics, clustering helps users understand the structure of the topic and expedites the process of reading documents and understanding the topic as a whole.

5.2.2. Number of documents browsed and precision effort

The mean number of documents browsed using Cancer Spider (4.533) was significantly lower than that using NLM Gateway (6.233), with a *p*-value of 0.0067. The mean precision effort of Cancer Spider (0.648) is significantly higher than that of NLM Gateway (0.436), with a *p*-value of 0.0009. Both of these measures represent the amount of manual browsing effort. For both of them, the differences between the two systems are significant at the 1% significance level. This confirms hypothesis H5.

Table 2
Searching time and effort

	Sample size	Cancer Spider		NLM Gateway		*p*-Value of *t*-test
		Mean	Variance	Mean	Variance	
Time (in min)	30	10.22	15.64	14.00	23.18	0.0003*
No. of documents browsed	30	4.53	3.57	6.23	12.46	0.0067*
Precision effort	30	0.65	0.05	0.43	0.07	0.0009*

*The difference is statistically significant at the 1% level.

In other words, compared to NLM Gateway, Cancer Spider requires less manual browsing effort from its users.

The key phrase extraction feature of Cancer Spider allows users to quickly judge the relevance of a given document. It saves users a significant amount of time since it eliminates the need for them to go through all the document summaries. Furthermore, Cancer Spider groups all returned documents into sub-topics based on a clustering algorithm, letting users focus only on sub-topics of interest to them, as opposed to opening and reading every single document. As a result, users need to browse fewer documents. Similarly, the higher precision effort associated with Cancer Spider is closely related to Cancer Spider's key phrase extraction and clustering capability. As commented by one of the subjects, "It (Cancer Spider) returned the highest amount of usable citations compared to other search engines, which usually only return 20–30% of what I need. It is the best search engine I've ever used."

5.3. Questionnaire results

The questionnaire was designed primarily to uncover users' subjective experience with the medical IR systems under study. It was comprised of questions relating to five different dimensions of the user experience: user interface, usefulness of the information retrieved in answering the search questions, users' level of certainty about their answers, user satisfaction of the search experience, and the amount of knowledge obtained after the search.

On a scale of 1–5 (5 being the most desirable), the result of the questionnaire is shown in Table 3. The data show that NLM Gateway scored higher in user interface, but the difference was not significant. Cancer Spider scored higher in all the other four categories. The differences in three of these categories were statistically significant at the 5% significance level, and the remaining one (Usefulness) is significant at the 10% level. The smaller variances in the evaluation of Cancer Spider also show that Cancer Spider performs more consistently for subjects with different backgrounds.

Table 3
Subjects' ratings of the two systems

	Sample size	Cancer Spider		NLM Gateway		*p*-Value of *t*-test
		Mean	Variance	Mean	Variance	
Interface	30	3.55	0.68	3.72	1.14	0.5018
Usefulness	30	3.93	0.71	3.41	1.04	0.0831**
Certainty	30	4.07	0.57	3.72	1.06	0.0479*
Satisfaction	30	3.79	0.74	3.14	0.91	0.0079*
Knowledge obtained	30	3.72	0.64	3.07	1.50	0.0139*

*The difference is statistically significant at the 5% level.

**The difference is statistically significant at the 10% level.

The interface of Cancer Spider is one of the dominant themes in our verbal protocol analysis. Although subjects showed strong interest in the interactive interface, they made many valuable suggestions for improvement. For instance, tabs such as *Good URLs* and *Document View* were designed to cut through different windows within a panel. Subjects who were accustomed to simple Web browser navigation found the distinction between within-panel and inter-panel switching confusing.

6. Conclusions and future directions

Both Cancer Spider and NLM Gateway have strengths and weaknesses. NLM Gateway, as a portal site to many high-quality NLM databases, is a conventional medical search engine. Cancer Spider's data coverage is not as comprehensive as that of NLM Gateway. However, Cancer Spider draws its strength from post-retrieval processing and clustering capability.

With respect to system performance measured by theme-based precision and recall, Cancer Spider and NLM Gateway achieved comparable results. When measured by system efficiency, Cancer Spider demonstrated statistically significant superiority in search time, the number of documents browsed, and precision effort. Cancer Spider's key phrase extraction feature and clustering tools have been proven helpful in assisting users to locate useful information. Overall, our experimental results show that when compared with NLM Gateway, users of Cancer Spider were able to retrieve the same amount of relevant information and achieve the same effectiveness (as measured by precision, recall and F-measure) in a shorter time and with less manual effort. One should note that Cancer Spider did not have access to as many high-quality data sources as NLM Gateway. As both systems achieved similar performance, it is possible that Cancer Spider may perform even better if it connects with more relevant data sources.

While the reported user study focused on the post-retrieval processing capabilities of Cancer Spider, a separate evaluation of the visualization component will be an interesting research issue. We plan to evaluate how users perceive search results clustered into a topic-map compared to a traditional ranked-list. It is also interesting to study how the different visualization metaphors affect a user's search effectiveness and efficiency.

Another interesting research area is UMLS-enhanced semantic parsing. Semantic parsing involves natural language processing capabilities that go beyond simple part-of-speech tagging and try to achieve a deeper understanding of semantic types and relationships between concepts that appear in the queries or documents. User queries will not be limited to the exact key phrases that a user enters. By including other semantically relevant search terms, documents containing closely related concepts also will be retrieved, thus solving the problem of vocabulary differences, a major issue in medical IR. We are planning to enhance Cancer Spider's performance through leveraging ontological knowledge from UMLS to provide high-quality domain knowledge and to suggest search terms to the user.

Acknowledgements

The Cancer Spider project is supported in part by the following research grants:

- NSF Digital Library Initiative-2, "High-Performance Digital Library Systems: From Information to Knowledge Management," IIS-9817473, April 1999–March 2002.
- NIH/NLM, "UMLS Enhanced Dynamic Agents to Manage Medical Knowledge," 1 R01 LM06919-1A1, February 2001–January 2004.

We would like to thank the researchers at the Arizona Cancer Center who participated in the user study and the librarians at the Arizona Health and Science Library who provided us with the necessary experiment facilities. We would also like to thank the members of the University of Arizona Artificial Intelligence Lab, particularly Bill Oliver, Wojciech Wyzga, Andy Lowe, Jared Phipps, Gondy Leroy, Wai-Ki Sung, Jennifer Embrich, Maryan Mikhael, Xiaoyun Sun, Hui Liu and Ann Lally, for their involvement in the project.

References

Bin, L., Lun, K.C., 2001. The retrieval effectiveness of medical information on the Web. International Journal of Medical Informatics 62, 155–163.

Bowman, C., Danzig, P., Manber, U., Schwartz, F., 1994. Scalable Internet resource discovery: research problems and approaches. Communications of the ACM 37 (8), 98–107.

Brill, E., 1995. Transformation-based error-driven learning and natural language processing. Computational Linguistics 21, 543–565.

Chau, M., Zeng, D., Chen, H., 2001. Personalized spiders for Web search and analysis. In: Proceedings of the First ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'01). ACM Press, New York, pp. 79–87.

Chen, H., Schufels, C., Orwig, R., 1996. Internet categorization and search: a self-organizing approach. Journal of Visual Communication and Image Representation 7 (1), 88–102.

Chen, H., Houston, A.L., Sewell, R.R., Schatz, B.R., 1998. Internet browsing and searching: user evaluations of category map and concept space techniques. Journal of the American Society for Information Science 49 (7), 582–603.

Chen, H., Fan, H., Chau, M., Zeng, D., 2001. MetaSpider: meta-searching and categorization on the Web. Journal of the American Society for Information Science & Technology 52 (13), 1134–1147.

Gauch, S., Wang, G., Gomez, M., 1996. Profusion: intelligent fusion from multiple different search engines. Journal of Universal Computer Science 2 (9), 637–649.

Harman, D., 1991. How effective is suffixing? Journal of the American Society for Information Science 42 (1), 7–15.

Hearst, M., 1995. TileBars: visualization of term distribution information in full text information access. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'95). ACM Press, New York, pp. 59–66.

Hersh, W.R., 1996. Information Retrieval: A Health Care Perspective. Springer, Berlin, Germany.

Howe, A.E., Dreilinger, D., 1997. SavvySearch: a meta-search engine that learns which search engines to query. AI Magazine 18 (2), 19–25.

Hull, D.A., 1996. Stemming algorithms—a case study for detailed evaluation. Journal of the American Society for Information Science 47 (1), 70–84.

Kiley, R., 1999. Medical Information on the Internet: A Guide for Health Professionals. Churchill Livingstone, London.

Keonemann, J., Belkin, N., 1996. A case for interaction: a study of interactive information retrieval behavior and effectiveness. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'96). ACM Press, New York, pp. 205–212.

Kohonen, T., 1995. Self-Organizing Maps. Springer, Berlin, Germany.

Krovetz, R., 1993. Viewing morphology as an inference process. In: Proceedings of 16th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '93). ACM Press, New York, pp. 191–202.

Lawrence, S., Giles, C.L., 1998a. Inquirus, the NECI meta search engine. In: Proceedings of the 7th International World Wide Web Conference, available at: http://www7.scu.edu.au/programme/fullpapers/1906/com1906.htm.

Lawrence, S., Giles, C.L., 1998b. Context and page analysis for improved Web search. IEEE Internet Computing 2 (4), 38–46.

Lawrence, S., Giles, C.L., 1999. Accessibility of information on the Web. Nature 400, 107–109.

Lin, X., 1997. Map displays for information retrieval. Journal of the American Society for Information Science 48 (1), 40–54.

Lin, X., Soergel, D., Marchionini, G., 1991. A self-organizing semantic map for information retrieval. In: Proceedings of the 14th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'91). ACM Press, New York, pp. 262–269.

Lovins, J.B., 1968. Development of a stemming algorithm. Mechanical Translation and Computational Linguistics 11 (1–2), 22–31.

Orwig, R., Chen, H., Nunamaker, J.F., 1997. A graphical self-organizing approach to classifying electronic meeting output. Journal of the American Society for Information Science 48 (2), 57–170.

Porter, M.F., 1980. An algorithm for suffix stripping. Program 14 (3), 130–137.

Salton, G., 1986. Another look at automatic text-retrieval systems. Communications of the ACM 29 (7), 648–656.

Salton, G., 1989. Automatic Text Processing. Addison-Wesley, Reading, MA.

Salton, G., Wong, A., Yang, C.S., 1975. A vector space model for automatic indexing. Communications of the ACM 18, 613–620.

Shneiderman, B., 1997. Designing information-abundant Web sites: issues and recommendations. International Journal of Human–Computer Studies 47, 5–29.

Selberg, E., Etzioni, O., 1995. Multi-service search and comparison using the MetaCrawler. In: Proceedings of the Fourth World Wide Web Conference, available at: http://www.w3.org/Conferences/WWW4/Papers/169/.

Selberg, E., Etzioni, O., 1997. The MetaCrawler architecture for resource aggregation on the Web. IEEE Expert 12 (1), 1997.

Sutcliffe, A.G., Ennis, M., Hu, J., 2000. Evaluating the effectiveness of visual user interfaces for information retrieval. International Journal of Human–Computer Studies 53, 741–763.

Tolle, K., Chen, H., 2000. Comparing noun phrasing techniques for use with medical digital library tools. Journal of the American Society for Information Science 51 (4), 352–370.

Veerasamy, A., Belkin, N.J., 1996. Evaluation of a tool for visualization of information retrieval results. In: Proceedings of the 19th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'96). ACM Press, New York, pp. 85–92.

Voorhees, E., Harman, D., 1998. Overview of the sixth Text REtrieval Conference (TREC-6). In: Voorhees, E., Harman, D. (Eds.), NIST Special Publication 500-240: The Sixth Text REtrieval Conference (TREC-6). National Institute of Standards and Technology, Gaithersburg, MD, pp. 1–24.

Westberg, E., Miller, R., 1999. The basis for using the Internet to support the information needs of primary care. Journal of the American Medical Informatics Association 6, 6–25.

Zamir, O., Etzioni, O., 1999. Grouper: a dynamic clustering interface to Web search results. In: Proceedings of the Eighth World Wide Web Conference, available at: http://www8.org/w8-papers/3a-search-query/dynamic/dynamic.html.