Problems and Challenges in Patient Information Retrieval: A Descriptive Study

Sandra Kogan BSc, Qing Zeng PhD, Nachman Ash MD, Robert A. Greenes MD PhD
Decision Systems Group, Brigham and Women's Hospital, Harvard Medical School
Boston, MA

Many patients now turn to the Web for health care information. However, a lack of domain knowledge and unfamiliarity with medical vocabulary and concepts restrict their ability to successfully obtain information they seek. The purpose of this descriptive study was to identify and classify the problems a patient encounters while performing information retrieval tasks on the Web, and the challenges it poses to informatics research. In this study, we observed patients performing various retrieval tasks, and measured the effectiveness of, satisfaction with, and usefulness of the results. Our study showed that patient information retrieval often failed to produce successful results due to a variety of problems. We propose a classification of patient IR problems based on our observations.

INTRODUCTION

With over 52 million Americans seeking health information on the Web, important questions about access and use of this content need to be addressed [1]. In addition to all the medical and health-related content available on the Web for consumers and patients, new initiatives are underway to deliver even more information to patients including their medical records and test results [2].

This is most certainly a positive step forward in an effort to democratize health care on the Web, and improve patient access to information and services [3]. Lack of familiarity with medical vocabulary and concepts, however, presents a serious impediment to access to and interpretation of the information available. Overcoming this problem and successfully supporting patient information retrieval (IR) activities pose challenges for informatics research.

In a previous study, we collected and analyzed patient and clinician terms entered into the Brigham and Women's Hospital (BWH) Web site, to confirm and quantitatively assess their differences. We also analyzed the IR performance resulting from these terms. The results showed that patient terminology does differ from clinician terminology in many respects including misspelling rate, mapping rate to the Unified Medical Language System (UMLS), semantic type distribution, and that patient terms led to poorer results in information retrieval [4,5]. That study, like McCray's study, suggested the need for terminology support for patients [6]. However, without the benefit of studying patients performing retrieval tasks on the Web, we could not observe all the problems they were experiencing or measure their satisfaction and perceived usefulness.

The aim of the current study was to identify and classify the problems a patient encounters while performing IR tasks on the Web. This study was observational, and involved three types of IR tasks performed on two Web sites.

BACKGROUND

To study patient IR, we used two testbed Web sites: MEDLINEplus [7,8] for consumer health content, and BWH's Web site to find a physician [9].

MEDLINEplus is a high quality health care information source provided by the National Library of Medicine (NLM) at the National Institutes of Health [8]. MEDLINEplus aims at serving both health professionals and consumers through the provision of authoritative health information. The content from this Web site includes information about specific diseases and conditions, and links to consumer health information from the National Institutes of Health, dictionaries, lists of hospitals and physicians, health information in Spanish and other languages, and information on clinical trials. MEDLINEplus supports both free-text search as well as browsing. For free-text search, "Advanced Search" is an option that provides more IR support such as spelling correction and word morphology. The major sections for browsing are Health Topics, Drug Information, Dictionaries, Directories, and Other Resources.

The Find-A-Doctor page of the BWH Web site allows the user to search for a doctor by entering information in free-text format. People can search by last name, first name, clinical interests, language and board certification. The page links to a centralized hospital database of physician profiles and their respective clinical interests, self-defined by the physicians, and reviewed by department chiefs. Free-text terms entered into the categories on the search page are matched to the database using a simple partial string matching method. The retrieved results are presented to users.

1067—5027/01/$5.00

METHODS

Patient recruitment was conducted in the patient family learning center of a large academic center where people come to access information about health and disease, and included hospital out-patients and visitors. This study was divided into three parts: (I). Free-form search using MEDLINEplus, (II). Free-form search using the Find-A-Doctor page on the BWH Web site, and (III). Scenario-based search using the Find-A-Doctor page on the BWH Web site. In parts I and II patients were asked to search for any health-related information of interest on the two Web sites. In part III both patients were presented with pre-defined scenarios, and asked to find the relevant information. Post-task interviews were conducted at the end of each part. For purposes of this study, effectiveness, satisfaction and usefulness are measured as follows:

Effectiveness is measured in terms of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) ratios for the search results. The interviewer established the gold standard by reproducing and analyzing the queries after each subject completed the tasks.

TP: Returns content that the subject wanted

TN: Returns no content, and content does not exist

FP: Returns some content, but not what the subject wanted

FN: Returns no content even though the content exists

Satisfaction and usefulness were rated by the subjects during the interview. They were asked if they were satisfied with the result (why or why not?) and if the results were useful to them in any way (why or why not?).

Part I: Patients searched for medical information on the MEDLINEplus Web site. They were asked to search for any health-related information that was of interest to them. Some subjects searched for more than one topic. The effectiveness of the data retrieved, the subject's satisfaction with the result, and the perceived usefulness of the information were recorded. Subjects were given a quick orientation to the Web site (approximately 2 minutes) before initiating their search.

Part II: Patients were asked to find a doctor on the BWH's Web site using the Find-A-Doctor page for any problems or concerns. Same kinds of data as in part I were collected.

Part III: Patients were presented with pre-defined scenarios, and asked to find a doctor using the BWH Find-A-Doctor page for each scenario. A short orientation to the Web site was provided. The following is an example of one of the scenarios:

Scenario 1. You are running to catch a bus on a snowy day in November, slip on the ice and hurt your wrist. You go to the emergency room and find out that you've broken it and will require surgery. You need to find a doctor to examine the wrist and perform the operation. Please try to find a doctor for this problem.

In all three parts, subjects were encouraged to keep trying for as long as they wanted, or needed. Based on our observations, a classification scheme was created for patient IR problems.

RESULTS

A total of 11 patients participated in this study. All reported graduating college or graduate school, and had some computer and Web experience. (Table 1) The interviewer was able to locate relevant content for all topics that patients searched for in Part I and II from the respective Web sites. In Part III, relevant content also existed in the Web site for all three scenarios.

Part I

Subjects were able to search for information on MEDLINEplus using various strategies, such as searching a list of health topics by letter of the alphabet, searching health topics by broad group, or using the free-text search. Most subjects, 82% (9/11) used free-text search instead of browsing the list of health topics. Patients who failed to find what they sought using the list, also moved to free-text search. A total of 15 queries were performed by the 11 patients – some patients performed more than one query (Table 2).

When query results were not useful (67%, 10/15), or when patients were not satisfied with the results (73%, 8/11), the interviewer encouraged them to try an alternate search method (health topics alphabet, or broad groups). In 64% cases, this did not improve the effectiveness, satisfaction or usefulness. The usefulness was also low – only 36%, (4/11) reported that the results were useful to them. Participants experienced some difficulties with the interface of the search results, but since this was not the objective of this study, these results are not further elaborated in the paper.

Queries for part I yielded results in all but three cases (false negative-result): *Aneurism*, *blood preasure* and *prostrate*. In the first case, the subject received no matching results to the first two misspelled queries and assumed no information was available on these

topics and was thus actually satisfied with the results, and felt they were useful.

In the third case, the patient's intention was to search for enlarged prostate (benign prostatic hypertrophy or BPH), but since he had entered *prostrate*, he was quite frustrated with the result (0 matches). The subject was prompted by the interviewer to continue searching for the term using "Health topics: P". The subject searched again and found links to *Prostate Cancer*, *Prostatitis: acute or chronic*. Nothing was listed for BPH and the patient became concerned that this was his doctor's way of giving bad news – that his diagnosis was something more serious.

Overall, the effectiveness was quite low with 20% TP matches, 73% FP matches, and 7% FN matches. Satisfaction was also quite low: 33% satisfied, and perceived usefulness was also 33% (Table 2).

Table 1 Demographic variables by number and percentage (%)

Variable	Patients, N=11
Gender:	
Male	7 (64.0)
Female	4 (36.0)
Age Group:	
19-50	7 (64.0)
>50	4 (36.0)
Education:	
College	6 (54.5)
Graduate School	5 (45.5)
Computer use:	
Never used a computer	0 (0)
<1 year	1 (9.0)
1-2 years	1 (9.0)
2-4 years	5 (45.5)
>4 years	4 (36.5)
Years of WWW use:	
Never used the WWW	0 (0)
<1 year	1 (9.0)
1-2 years	3 (27.2)
2-4 years	4 (36.5)
>4 years	3 (27.2)
Visit other medical Web sites:	
Yes	
No	7 (64.0)
	4 (36.0)

Table 2 Results of MEDLINEplus queries, by patients, N=11. Search terms are displayed as entered.

Search term	Results	Eff	Sat	Use
Exercise stress test	9	FP	No	No
Healthy bones	22	FP	No	DK
Steroids	74	TP	Yes	Yes
Acupuncture	7	FP	No	No
Cat scans	8	FP	No	No
Blood sugar	210	FP	No	No
Hepatites	62	TP	Yes	Yes
Stroke	222	TP	Yes	Yes
Prostrate	0	FN	No	No
Heart problems and smoking	45	FP	No	No
Aneurism	0	FN	Yes	Yes
Blood preasure	0	FN	Yes	Yes
Sore ankle	34	FP	No	No.
Diabetes care	260	FP	No	No
Fat in diet	58	FP	No	No

Eff = Effectiveness as defined by TP (true positive); FP (false positive); TN (true negative); FN (False negative). Sat = Satisfied; Use = Useful; DK = don't know

Table 3 Results of the Find-A-Doctor search on the Brigham and Women's (BWH) Hospital Web site, N=11.

Term entered	Results	Eff	Sat	Use
Heart problems	0	FN	No	No
Adult asthma	0	FN	No	No
Arthitis	0	FN	No	No
Internal Med	51	TP	Yes	Yes
Cat Scan	0	FN	No	No
Blood sugar	0	FN	No	No
Sinusitis	2	TP	Yes	Yes
Endochronology	0	FN	Yes	Yes
Prostate	13	TP	Yes	Yes
General doctor	0	FN	No	No
Neraul	0	FN	No	No

Eff = Effectiveness as defined by TP (true positive); FP (false positive); TN (true negative); FN (False negative). Sat = Satisfied; Use = Useful

Part II

The results for this part show that overall effectiveness was low, 27% TP and 73% FN. Satisfaction was reported to be 36% satisfied and 64% unsatisfied, and same for usefulness (Table 3). The subjects were not always aware that they had

made spelling errors and assumed that if the term entered was incorrect, that the correct term would be provided. One subject felt that the search engine was broken since she knew for a fact that there are "general doctors" at BWH. Another subject was sure of the existence of content, but could not identify her error when she typed in "Endochronology" and received no matches to her query.

Part III

Due to space limitations, the terms entered for each scenario are not detailed here. Instead we provide a summary table of the effectiveness (TP, TN, FN, FP), satisfaction, and usefulness of the results. (Table 4).

Table 4 Mean query results of the three pre-defined scenarios: Effectiveness, Satisfaction and Usefulness (%)

Results	Patients (N=11)
TP	11
TN	2
FN	78
FP	9
Satisfed: Yes	11
Useful: Yes	12

In terms of effectiveness, patients yielded a true positive (TP) result 11% of the time. Satisfaction and usefulness of the results were also quite low.

Subjects encountered many difficulties in this part. For example, in the first scenario (as depicted in Methods), several patients typed in 'Bone' to find a doctor to perform surgery on a broken wrist. The search for 'bone' yielded 28 matches, but none related to orthopedics. The results pointed to specialists dealing with bone metabolism or bone marrow transplantation. Another common term was "surgery" which yielded 182 matches – too many for the subjects to go through manually.

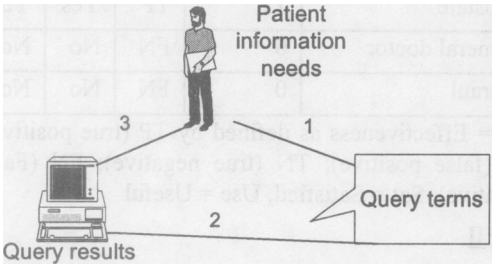

Figure 1 Patient IR cycle: from information needs to query term, to query results and then back to information needs.

Classification of Observed Problems:

The patient IR process can be viewed as a three-part cycle: 1. Formulate query terms based on information needs. 2. Obtain results from search engines by submitting query terms. 3. Use query results to satisfy information needs (Figure 1). Patients encounter problems at each stage (Table 5).

Table 5 Problems patients encounter during IR cycle.

	Problems
1	a. Patients unaware of what their needs are (e.g. do not know what to search for)
	b. Patients aware of their needs but cannot articulate them (e.g. "healthy bone")
	c. Patients can articulate their needs, but make mistakes during the process (e.g. misspelling: *aneurism* or wrong medical term: *CAT scan* instead of *CT scan*)
2	a. Query term does not match the content resource (e.g. mismatch of mental models: *bone* instead of *orthopedics*)
	b. Content poorly organized or indexed (e.g. BPH listed under *B* instead of *P*)
	c. Information source lacking requested content (e.g. no drug info. On Find-A-Doctor)
3	a. Patients can not comprehend query results (e.g. can not understand content written in medical jargon)
	b. Patients misinterpret query results (e.g. assume no results means no content exists)
	c. Inappropriate number of query results (e.g. too much or too little)

DISCUSSION

Patient IR often fails to produce successful results even after iteration. People search the Web for health-related information when they have a need or interest. A Harris poll reported that 70 million adults used the Web to find health information between June 1998 and June 1999, and the numbers grow each month [10]. Based on our research and observations, it is apparent that people have difficulty in getting access to the information they seek. Our study showed overall low rates of effectiveness, satisfaction, and usefulness.

This issue is much larger than that of health literacy alone, which is now moving to the forefront of medical content on the Web for patients [11]. As shown in our classification of problems, not all

problems can be attributed to health literacy. Patient information retrieval affects all interactions using the Web - access to information, services and health care (telemedical programs, monitoring protocols, etc).

Our data suggest that unsuccessful IR is a common problem. All of our patient subjects were highly educated users of technology and all but one have been accessing information on the Web for at least 1-2 years. Yet, the IR outcomes were still quite poor. We postulate that the problem is much more pronounced in the general population.

We also found that these three measures reflect different aspects of IR performance – effectiveness is not correlated with satisfaction and usefulness.

As there are difference kinds of relevance (topic, task and context), the measurements we used in this study do not measure all aspects of IR performance. For instance, we did not attempt to measure the impact of reading the retrieved documents on patients' knowledge or quality of life [12].

The observed IR problems present challenges to informatics research. In order to solve the problems, various layers of support need to be developed specifically for patients. Such support includes lexical tools (e.g., to correct spelling), semantic tools (e.g., to map patient terms to medical concepts), domain knowledge support (e.g., to bridge the gap between mental models), and filtering tools (e.g., to control information overload).

In addition to developing new tools, issues with content and indexing of the content also needs to be addressed. The BWH Web site exemplifies a common problem with health Web sites. This Web site allows one to search for a doctor by matching search terms to the self-described clinical specialty/interest. This is typical of health-related Web sites where content is indexed by medically oriented terms.

Our patient population is likely not typical. Our assumption is that the recruited patient group would perform better than the general population because of their education, knowledge of resources, and access to computers and the Web.

Although this is a study of patient IR, not of particular Web sites, the results are limited by the number of the Web sites studied. We chose one Web site for consumer health content (MEDLINEplus) and one for accessing services (Brigham and Women's Hospital Web site). The content of these Web sites is representative of the health information needs of patients. However, generalizability of the results would require further study.

Acknowledgments

We thank all study subjects for their participation in our research. We would like to thank Dr. Aziz Boxwala for reviewing this paper.

REFERENCES

[1] Horrigan J, Lenhart A, Spooner T et al. The online health care revolution: How the Web helps Americans take better care of themselves. Pew Internet and American Life Project. http://www.pewinternet.org

[2] Eysenbach G. Consumer health informatics. *BMJ.* 2000;320:1713-6.

[3] Milne R, Booth-Clibborn N, Oliver S. Consumer health information needs to be rigorous, complete, and relevant. BMJ. 2000 Jul 22;321(7255):240.

[4] Zeng Q, Kogan S, Ash N, Greenes RA. Patient and Clinician Vocabulary: How Different Are They? Submitted to MedInfo, 2001.

[5] Inniss A. Lack of support for patient on the Web. Technical report 081800, DSG – Brigham and Women's Hospital.

[6] McCray AT, Loane RF, Browne AC, Bangalore AK. Terminology issues in user access to Web-based medical information. Proc AMIA Symp 1999;:107-11

[7] National Library of Medicine. MEDLINEplus. http://www.nlm.nih.gov/medlineplus/. 2000.

[8] Miller N, Lacroix EM, Backus JE. MEDLINEplus: Building and maintaining the National Library of Medicine's consumer health Web service. Bull Med Libr Assoc 2000;88(1):11-7.

[9] Brigham and Women's Hospital. Find-A-Doctor. http://www.brighamandwomens.com/. 2000.

[10] Grandinetti DA. Doctors and the Web: Help your patient surf the net safely. Medical Economics Archive, 03/06/200.

[11] Greenberg B. Literacy needed for good health. Associated Press. 3/3/01 http://www.zeenews.com/links/articles.asp?aid=9448&sid=ENV)

[12] Mizzaro S. Relevance: The whole history. Journal of the American Society for Information Science. 1997;48(9):810-832.