Scenario-based Assessment of Physicians' Information Needs

Yoon-Ho Seol, David R. Kaufman, Eneida A. Mendonça, James J. Cimino, Stephen B. Johnson

Department of Biomedical Informatics, Columbia University College of Physicians and Surgeons, New York, New York

Abstract

Physicians' information needs routinely arise during their practice. Several studies have demonstrated that a useful way to understand the nature of the needs is to examine questions posed by physicians during the course of medical care. This paper presents an analysis of clinical questions collected when physicians were engaged in reviewing clinical cases. The main objective of the study was to investigate characteristics of the physicians' information needs to provide insight into the development of a conceptual guidance approach in information retrieval. The analysis focused on categories, search contexts, and patterns of the questions with respect to the scenarios. The results of the study showed that physicians' information needs exhibited distinct characteristics according to the scenarios, and that their needs could be expressed with a relatively small number of question patterns. These observations supported our approach of using patterns of information needs in facilitating digital information access.

Keywords:

Information Storage and Retrieval; User-Computer Interface; Medical Informatics Computing; Information Science

Introduction

The explosive growth of medical knowledge necessitates that physicians seek knowledge from diverse information sources. To understand and improve the services that provide access to relevant professional information, multiple studies have investigated health care providers' information needs and information seeking behavior [1-8]. These studies show that there are identifiable classes of information needs and patterns of information seeking behaviors among health care practitioners.

Studies typically focus on the ability of physicians to generate questions regarding the practice of medicine. Various methodologies have been employed to elicit and capture physicians' information needs in the form of clinical questions. For example, Covell and colleagues[1] used interviews and self-reporting questionnaires to collect questions from internists and other specialists. The study showed that the most frequently asked questions in outpatient practice were about treatment (31%) and diagnosis (25%). Timpka, et al. [2] analyzed a collection of the most recent incidence of clinical questions that physicians had during their practice. Their study reported similar patterns of information needs from general practitioners: most questions were

concerned with choice of therapy (44%) and diagnosis (36%). Chambliss and colleagues [3] asked family physicians to call after their half-day practice to report unanswered questions. Among the 103 questions collected, 37% concerned treatment and 29% concerned diagnosis. Questions about drug information occurred 15% of the time. A more recent study by Ely, et al. [4] also used self-reported questionnaires and presented a taxonomy consisting of 69 types of questions based on 1101 questions collected from 103 family doctors. A subsequent study [5] revised the taxonomy, consisting of 64 question types, and reported that the most common types of questions were about the diagnosis and treatment in the form of **What is the cause of symptom X?**, **How should I manage disease or finding X?**, and **Is drug X indicated in situation Y?**

With the proliferation of online medical information sources, the questions posed to these sources are becoming a useful means to characterize the information needs of healthcare professionals. Haynes, et al. [6] studied records of online access and showed that 41% of searches were on therapy and 6% of requests were on side effects. Strasberg and his co-workers [7] analyzed queries asked of their information retrieval system and reported that the most frequent requests were on drug information.

Another approach to collecting physicians' information needs is the use of specific patient data from which clinical questions are generated. For their analysis of needs, Cimino and Barnett [8] used clinical summaries of real patients to collect questions from clinicians. They suggested that questions might have more validity if they were collected from physicians based on specific patient data. The study documented the value of specific patient situations in investigating clinicians' information needs.

The primary aim of this study is to gain an understanding of physicians' information needs expressed while they are reviewing clinical cases. The assessment provides an opportunity to examine the patterns of clinical questions that may represent a substantial percentage of the variance of the needs and identify the contexts in which physicians formed questions.

Background

To support access to digital information sources, we have developed a conceptual guide for information retrieval, based on a knowledge base that contains classes of information needs [9,10]. These classes, called *generic queries*, are used in conjunction with electronic medical records to construct patient-specific questions. To build such a knowledge base, we devel-

oped a formal representation of clinical questions using the UMLS knowledge sources, called the Generic Query model.
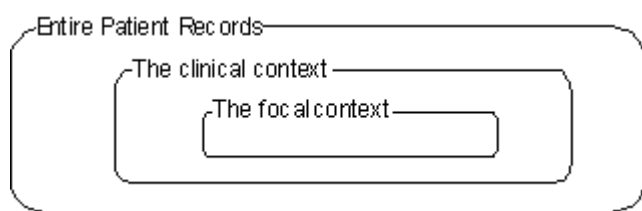

Figure 1 - relationships of the contexts in patient records

The Generic Query model makes use of the abstraction of concepts and interconcept relationships that UMLS semantic types and semantic relationships provide. A generic query describes semantic relationships between concepts explicitly by using the UMLS knowledge sources. For instance, a clinical question "Is drug X indicated for condition Y" can be represented with two concept nodes and a relation that connects the concept nodes to each other:

[Pharmacologic Substance: X] →(treats)→

[Pathologic Function: Y]

Our model produces a generic query that has the following characteristics:

- Captures generalizations about user queries in a single template.
- Separates instances of clinical concepts from questions that refer to them.
- Provides slots for appropriate types of the concepts.
- Identifies explicitly the relationship that links the concepts.

To augment the quality of the knowledge base, we have investigated a method to extract plausible clinical questions from the medical literature and demonstrated the potential of the method for inferring physicians' information needs [11]. The study presented in this paper is part of a series of efforts to investigate physicians' information needs in the context of information retrieval by using diverse sources of data.

Method

Thirteen physicians participated in the experiment. All subjects were also pursuing degrees in Biomedical Informatics at Columbia University.

Clinical Cases

Three scenarios, representing three patients with cardiovascular problems, were prepared for the experiment. The first case was a patient who is a 46-year old female and was admitted for gastrointestinal workup for bleeding. Her lab values revealed decreased hematocrit from 27 to 19 in a two-week period. The second case presented a 65-year-old woman with a history of hypotension. She underwent an atrial septal defect (ASD) closure and a radiofrequency (RF) maze procedure. The third case was a 47-year-old male with a history of hypertension, hypercholes-

terolemia, and Hodgkin's disease. He was admitted to the hospital and underwent coronary artery bypass grafting (CABG).

Each clinical scenario consisted of two segments: (1) the clinical history that would serve as background information for the patient in question, called *the clinical context*; and (2) the current focus of patient data related to the specific task subjects were performing at the moment that an information need arises, called *the focal context*. For example, if a subject's task was to review lab results, the lab data become the focal context, and the patient records that were potentially relevant to the current focus were regarded as the clinical context. Figure 1 depicts the contexts of patient records. As the physician moves through medical records, the focal context changes according to her immediate focus or concerns, and shifts to different parts of the records. The clinical context represents a subset of the patient records that is relevant to the focal context. The clinical context also changes according to the focal context as different kinds of background information become more significant.

Figure 2 illustrates how the three cases were constructed, focusing on the different contexts of patient data. Case 1 presented lab results containing hemoglobin and hematocrit values as the focal context. Case 2 provided microbiology data as its focal context. Case 3 gave a previous discharge summary as the focal context.

Procedure

Each subject was asked to review all three cases. The order in which the cases were presented to the subject was randomized to diminish order effects. Each case was presented in two segments to remind the subjects of the difference between the two contexts of patient data. The subjects were given the second segment (the focal context) after notifying the experimenter when she/he had finished the first segment (the clinical context). The subjects were allowed to refer to the first segment whenever they needed to while reviewing the second segment.

The participants were expected to generate questions related to the patient information in the second segment, the focus of the case, but they were not restricted in their ability to ask any questions from the first segment. However, the procedure was concerned with the subject's understanding of the focus of the patient data for each case when generating clinical questions.

The subject's verbal expressions were audiotaped and transcribed for analysis. The coding scheme was iteratively developed based on discussions in lab meetings in which the authors participated.

Analysis

The questions were categorized according to four types; treatment, diagnosis, etiology, and prognosis. For the purpose of consistency in categorization, a simple rule was applied to distinguish between diagnosis and etiology types of questions. If a question was articulated explicitly as a (differential) diagnosis or contained clinical findings from test results and asked the causes of those findings, the question was categorized as a diagnosis type. Otherwise questions were classified as an etiology type. For instance, when test results showed abnormal findings, indicating *anemia*, a question such as **What is the cause of *anemia*?** was regarded as a diagnosis type. But, a similar form of the

question, **What is the cause of *gastritis*?** was categorized into an etiology type because the concept *gastritis* was not indicated in the test results.

The questions were also examined based on the use of the contexts in the questions. The aim of this analysis was to investigate how patient data, located in different contexts, were used to express the physicians' information needs. The differentiation between the contexts provides us with a hypothesis for delimiting the space of information needs. According to the origins of concepts in a question with respect to the contexts, the question was classified into one of the following: (1) a question from the focal context only; (2) a question from the clinical context only; or (3) a question involving both contexts.

C a s e   1

Discharge Summary		Lab. Results	Admission Note
			Lab. Results

```
03/12/2000      04/01/2000
                        04/08/2000
```

C a s e   2

Admission Note	
Lab. Results	Lab. Results

```
11/30/1999      12/02/1999
```

C a s e   3

	Admission Note
Discharge Summary	Lab. Results

```
11/11/1999      12/12/1999
```
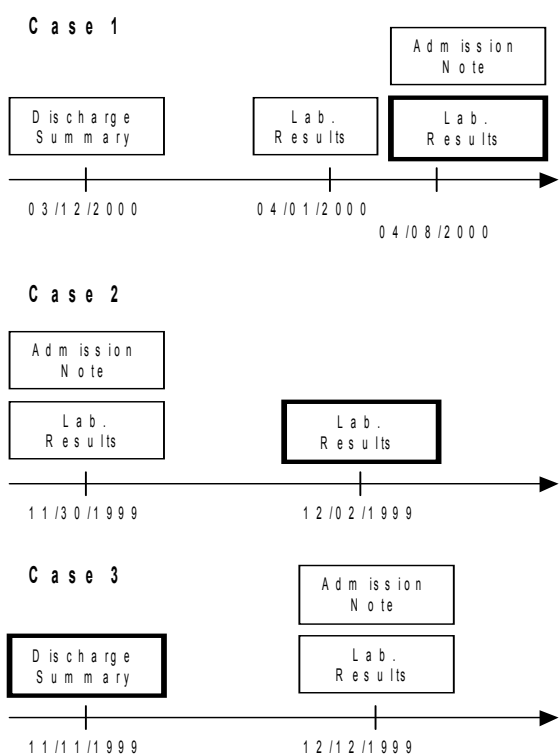
Figure 2 - Clinical cases with different contexts. Patient data in the solid box indicate the focal context and the other boxes represent the clinical context. Arrows denote timelines of the patient data.

The questions were further analyzed by using the generic query model to delineate semantic structures of the questions. The analysis examined whether similar patterns of the questions occurred in the information needs of the subjects. This estimation provided quantitative measures that indicated how many questions shared similar patterns.

Results

A total of 197 questions were collected from the thirteen subjects. Sixty-eight questions were obtained from Case 1 (mean = 5.23, SD = 1.48). From Case 2, thirty-nine questions were generated (mean = 3, SD = 1.08). Ninety questions were gathered from Case 3 (mean = 6.92, SD = 8.74). Questions that asked overview or review information were excluded from the analysis, leaving a set of 182 questions. The time spent by the subjects

to review the three scenarios varied from 40 to 75 minutes (mean = 54.23, SD = 10.96).
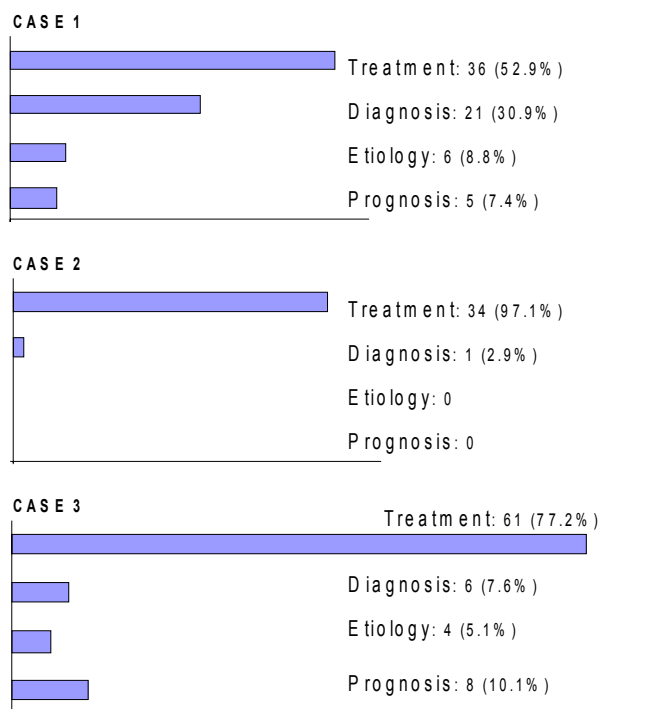
CASE 1

Treatment: 36 (52.9%)
Diagnosis: 21 (30.9%)
Etiology: 6 (8.8%)
Prognosis: 5 (7.4%)

CASE 2

Treatment: 34 (97.1%)
Diagnosis: 1 (2.9%)
Etiology: 0
Prognosis: 0

CASE 3

Treatment: 61 (77.2%)
Diagnosis: 6 (7.6%)
Etiology: 4 (5.1%)
Prognosis: 8 (10.1%)

Figure 3 - Question types based on the requested

The 182 questions were categorized based on the types of information they requested. Figure 3 shows the distribution of the questions by type. Within the treatment questions from Case 1, the subjects were more likely to ask questions about complications of therapeutic procedures related to diseases. These kinds of questions accounted for 22 (61.1%) of 36 treatment questions, whereas, in Case 2, 5 (14.7%) out of 34 treatment questions were associated with complications of the therapeutic procedures.

In the context analysis of the questions, the results showed that many questions were generated by combining the concepts from the focal and clinical contexts. Of 68 questions that were generated from Case 1, 37 questions contained concepts from the focal context. Of those, 23 questions involved concepts from both contexts. Case 2 and Case 3 also demonstrated similar observations with smaller frequencies (i.e., 6 out of 33 questions in Case 2 and 12 out of 76 questions in Case 3). Table 1 shows the distribution of the questions based on the context classification.

Table 1: Distribution of the questions based on the contexts

Case	Context	Frequency
Case 1	F	14 (20.6%)
	C	31 (45.6%)
	FC	23 (33.8%)
Case 2	F	27 (77.1%)
	C	2 (5.7%)
	FC	6 (17.2%)
Case 3	F	64 (81%)
	C	3 (3.8%)
	FC	12 (15.2%)

F = focal context only; C = clinical context only;
FC = both focal and clinical contexts.

Table 2: Question pattern frequency observed in the cases

Question Pattern	Case 1		Case 2		Case 3	
	Q	S	Q	S	Q	S
Does drug X (or therapeutic procedure X) treat condition Y?	3	3	22	11	23	10
Is drug X is more effective than drug X1…X2 for condition Y caused by organism Z?			5	4		
Does drug X (or therapeutic procedure X) cause condition Y?	20	10	5	4	4	2
What are the side effects of drug X?			2	2	2	2
What is the dose of drug X?					3	3
Are there any interactions between drug X1 and drug X2?	1	1			3	3
Is low dose of drug X as effective as high dose of drug X?					1	1
When should drug X be stopped/started?					1	1
What is the treatment for condition X associated to condition Y?	5	3			5	3
What is the treatment for condition Y caused by drug X (or therapeutic procedure X)?	2	2			18	10
When should therapeutic procedure X be stopped/started?	1	1				
How many units should be given to treat condition Y?	2	2				
Does therapeutic procedure X affect physiologic function Y?	1	1			1	1
Is treatment X1 is more effective than treatment X2…Xn?	1	1				
What are the causes of condition Y?	11	8			1	1
What is the normal range of test result X?					6	5
Is condition Y1 associated with condition Y2?	15	10			2	2
Is behavior W associated with condition Y?					1	1
Does organism Z cause condition Y?			1	1		
Does condition Y affect body substance X?	1	1				
What is the prognosis of condition Y?	1	1			3	3
What is the prognosis of condition Y1 associated with Y2?	1	1				
What is the prognosis of condition Y complicated by therapeutic procedure X?	3	3			5	5

Q = the number of questions expressed by a generic question; S = the number of subjects who asked the question.

Using the generic query model, we identified a number of common question patterns. For 68 questions from Case 1, 15 distinct semantic patterns were recognized. The number of clinical questions associated with each generic question differed but the results showed that a small number of semantic patterns accounted for a large number of questions. Table 2 presents the patterns of questions derived from the cases and their corresponding frequency. Note that three generic questions represented 46 questions (67.6%) in Case 1. A frequency analysis of question patterns in the other two cases reported similar findings. For Case 2, a total of 35 questions were associated with 5 generic questions. Of those, two generic questions expressed 27 questions (77.1%). Sixteen generic questions were observed from a set of 79 questions in Case 3 and two of them represented 41 questions (51.9%). We observed considerable overlap among question patterns across the cases. However, their frequency differed markedly. For example, the question pattern, **Does drug X (or therapeutic procedure X) treat condition Y?** accounted only for 3 (4.4%) of 68 questions in Case 1, but the pattern was associated with the highest frequency in Case 2 and Case 3. Similarly, the pattern **Does drug X (or therapeutic procedure X) cause condition Y?** represented 20 (29.4%) of 68 questions, but it was associated only with 5 (14.3%) of 35 and 4 (5.1%) of 79 questions in Case 2 and Case 3, respectively.

Discussion

The purpose of the study was to examine physicians' information needs by means of clinical scenarios. The experiment provided several characteristics of the clinical questions that would be useful to explore the generic query approach in assisting clinicians' information retrieval.

The results show that the most frequently asked questions were treatment-related. The comparison of the treatment questions between Case 1 and Case 2 showed differences in the nature of information being requested. With Case 1, the subjects' needs were centered on the causes of problems and their associations with other conditions of the patient, whereas, with Case 2, the subjects focused on how to treat the problems indicated in the test results. This observation suggested that physicians' information needs were affected by types of lab data.

The analysis of the questions in terms of the search contexts revealed that the subjects were likely to ask questions by joining concepts from both contexts. This indicated that the physicians' information needs occurred not only from the data they were reviewing for the specific task at the moment, but also from the broad context that incorporated background information into the current focus of patient data. Compared to Case 1, Case 2 showed higher rate of the questions involving concepts only from the focal context. With Case 1, there was considerable interest in understanding problems indicated in the test results by using the background information, whereas, with Case 2, the subjects were focused on treatment options for problems identified in the test results and used less background information in formulating questions. Case 3 showed the highest rate of the questions that were constructed by only using the focal context. The reason for this is possibly related to the types of patient data. As discharge summaries provided a concise description of a pa-

tient status, many questions could be asked without referring to background information.

The results of the experiment have provided some evidence that supports the generic query approach. The frequency statistics of question patterns observed in each case suggest that many clinical questions could be expressed by a smaller number of generic questions, which was concordant with the main ideas of the work of Ely and his colleagues [4,5].

This study is part of a research program to validate and extend the generic query approach by employing convergent research methodologies. We have conducted several other studies including one in which the generic query model was applied to a large set of documents from the medical literature. A comparison of the semantic patterns from the experiment against the ones from the literature analysis [11] demonstrates that similar patterns were identified from the literature with fairly high frequency of occurrence in the documents.

The experiment reported here used clinical questions that were collected from a small number of subjects in one setting and were based on just 3 clinical scenarios. Given these limitations, the information needs identified in this study may not reflect the spectrum of the broader population of physicians' needs. However, this study demonstrates the feasibility of our method for detecting and characterizing information needs that can be applied in a larger study.

The observations from the study revealed significant characteristics of physicians' information needs. The results provide support for the generic query model and lend credence to the instantiation of this model in the context of a system that supports online information access.

Acknowledgement

References

[1] Covell DG, Uman GC, Mannning PR. Information Needs in Office Practice: Are They Being Met? *Annals of Internal Medicine*. 1985;103(4):596-599.

[2] Timpka T, Ekstrom M, Bjurulf P. Information Needs and Information Seeking behavior in Primary Health Care. *Scand J Prim Health Care*. 1989; (7):105-109.

[3] Chambliss ML, Conley J. Answering Clinical Questions. *Journal of Family Practice*. 1996; 43:140-144.

[4] Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, Evans ER. Analysis of questions asked by family doctors regarding patient care. *BMJ* 1999; 319: 358-361.

[5] Ely JW, Osheroff JA, Gorman PN, Ebell MH, Chambliss ML, Pifer EA, Stavri PZ. A taxonomy of generic clinical questions: classification study. *BMJ* 2000; 321: 429-432.

[6] Haynes RB, McKibbon KA, Walker CJ, Ryan N, Fitzgerald D, Ramsden MF. Online access to MEDLINE in Clinical Setting. *Annals of Internal Medicine*. 1990;112:78-84

[7] Strasberg HR, Hubbs PR, Rindfleisch TC, Melmon KL. Analysis of Information Needs of Users of the Standard Health Information Network for Education. *Proc. AMIA Annual Fall Symposium*. 1999:965-969

[8] Cimino C, Barnett GO. Analysis of Physician Questions in an Ambulatory Care Setting. *Comp and Biomed Research*. 1992; 25:366-373.

[9] Cimino JJ, Aguirre A, Johnson SB, Peng P. Generic Queries for Meeting Clinical Information Needs. *Bulletin of the Medical Library Association*. 1993; 81(2): 195-206.

[10] Seol YH, Johnson SB, Cimino JJ. Conceptual Guidance in Information Retrieval. *Proc. AMIA Annual Fall Symposium*. 2001;1026.

[11] Seol YH, Johnson SB, Cimino JJ. Knowledge Acqusition of Generic Queries for Information Retrieval. *Proc. AMIA Annual Fall Symposium*. 2002;1160

Address for correspondence

Yoon-Ho Seol, PhD.
Dept. of Biomedical Informatics
Columbia University
622 West 168th Street, Vanderbilt Clinic, 5th Floor
New York, New York 10032
Email: seol@dbmi.columbia.edu