Answering Information Needs in Workflow

Eneida A. Mendonça, David Kaufman, Stephen B. Johnson
Department of Biomedical Informatics, Columbia University, New York, NY, USA

Abstract
During the process of patient care, clinicians frequently experience the need for information about treatment, diagnostic workup, disease progression and other aspects of patient management. In most of these situations, it is difficult or impossible for the clinician to immediately access appropriate information resources. Most information needs are never adequately articulated or recorded, and consequently are forgotten by the end of the day. Moreover, when clinicians do recall information needs, they often don't act on them, due to the significant limitations of current retrieval systems and the exigencies of clinical practice. This paper describes and discusses the architecture of an information system called CIQR ("seeker"): Context-Initiated Query and Response. CIQR enables clinicians working in the field to pose queries and receive responses, without interrupting their workflow. CIQR is the outgrowth of several years of research in information retrieval with the goal of finding out what information the user really wants to know and delivering it when and where it is needed. A unique aspect of the described approach is that the user submits open-ended, multi-sentence questions, not just key words. This enables the user to provide contextual background related to the question, such as pertinent characteristics of the patient, the purpose of the query, and the kind of materials the user is seeking. These items provide vital clues for constructing search strategies that are better tuned to the user's environment and emergent goals.

INTRODUCTION

Needs for information arise continuously during the course of clinical practice, especially for physicians in training, for example when examining a patient, participating in rounds or attending conferences. Information needs include questions about the best current scientific evidence available for treatment and diagnosis, as well as factual information available in medical textbooks, databases and other reference materials. Gaps in knowledge can have significant detrimental effects in patient care, and can contribute to medical errors. Filling these gaps is an important part of clinical training. However, while clinicians are engaged in work it is usually difficult or impossible to access appropriate information resources. Later, when workflow presents opportunities for seeking information, clinicians face numerous barriers to obtain what they need. Most frequently, information needs are forgotten because they were never adequately articulated or recorded. Moreover, when clinicians do recall information needs, they don't act on them, due to the significant limitations of current retrieval systems.

In this paper, we describe a model of an information system called CIQR ("seeker"): Context-Initiated Query and Response. CIQR enables clinicians working in the field to pose queries and receive responses, without interrupting their workflow. CIQR is the outgrowth of several years of research in information retrieval with the goal of finding out what information the user really wants to know and delivering it when and where it is needed.
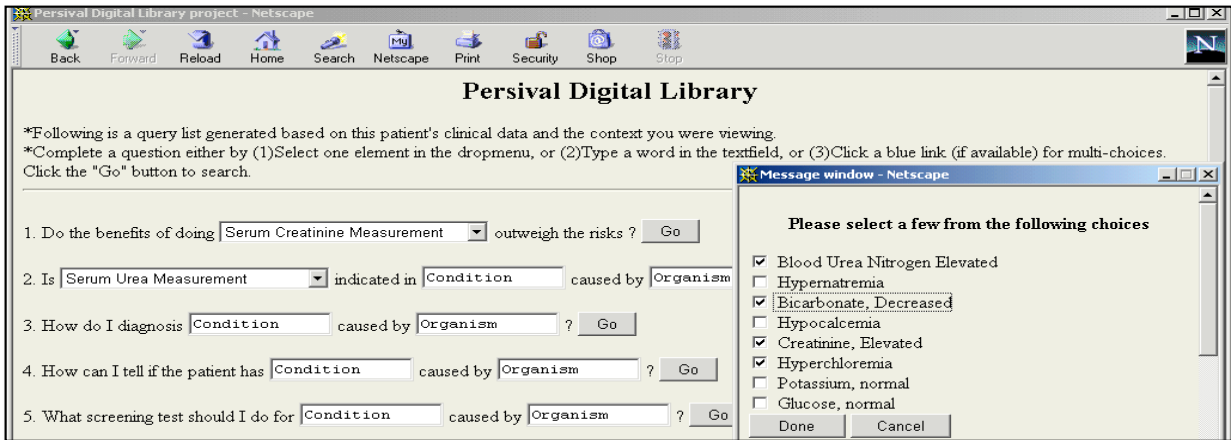
Figure 1. Query user interface component of HINT

BACKGROUND

Information seeking is a conscious effort to acquire information in response to a gap in knowledge. While simple information needs that arise in clinical settings can be addressed by asking a colleague, many other information needs require more comprehensive or authoritative sources. However, these may have associated costs, especially in time-sensitive situations, which may preclude their use by health care practitioners. Our previous efforts concentrated on the situations in which an information need arises while the user is viewing information in an electronic patient record. The objective was to provide access to knowledge at the point of care. The patient record was used to establish a context that could be used to guide searches for information according to what is known about the specific patient, and what is know about patterns of queries. We have developed a multidisciplinary framework that incorporates theories and methods of evidence-based medicine, natural language processing, and cognitive methods of empirical inquiry.

The HINT system (Health Information Needs Translator) was developed over the past five years by Drs. Johnson, Mendonça, and Kaufman[1] as part of a larger system that provides personalized access to distributed information resources, by incorporating patients' individual characteristics.[2] HINT explores new techniques for searching on-line information resources and presenting results to the user. The project uses information about the patient to customize the

presentation of results, and makes use of evidence-based medicine questions to help elicit user information needs. As illustrated in Figure 1, HINT provides users with context sensitive questions which the user completes and then sends the data to the search component. The search result is presented to the user and can subsequently be further refined by either an automated process or with the user's input. On the basis of analysis of context, the user is presented with a series of priority-ordered question stems with concepts selected from the relevant parts of the patient case. He or she can select a concept from either a pulldown menu, pop-up window or by typing it in a textbox.

This project focused on a central problem in information retrieval: finding out what it is that the user really wants to know. In particular, our research examined how to assist clinicians and patients to express their information needs at the point of care, while viewing information in an electronic patient record. Development of this system involved several innovative technologies:

- *XML Patient Database* – We created an electronic medical record based on documents annotated in XML (Extensible Markup Language), and accessed through the Tamino native XML database. Medical documents use the Clinical Document Architecture (CDA), an ANSI standard (under the Health Level 7 initiative) for the format of clinical documents. We have also expanded the use of the database to be a repository for eXstensible Stylesheet Language for Transformations (XSLT) documents.[3]
- *Clinical Context Identification* –This includes what the user is looking at (what the user clicks on), as well as other information in the medical record deemed to be related (based on statistical or knowledge-based models). This research allowed us to form a semantic representation of clinical context.[4,5]
- *Clinical Question Formulation* – This determines a likely set of clinical questions based on a semantic representation of the clinical context (what the user was looking at) and a knowledge base of known structures of generic clinical questions. The knowledge base was developed using the theory of evidence-based medicine and numerous empirical studies of information needs. A graph-matching algorithm compares the clinical context against the knowledge base of potential information needs. The system returns a ranked list of potential questions, filled in with concepts from the clinical context.[6]
- *Question Analysis* — A query analyzer (AQUA) extracts semantic content from text selected from the patient's medical record, representing each sentence as a semantic graph. Semantic types and relations are drawn from the Unified Medical Language System (UMLS). The AQUA parser was developed using a machine learning approach, rather than handcrafting grammar rules. The AQUA parser has been evaluated in previous work on its ability to generate syntactic[7] and semantic parses.[8]
- *Speech Input* – The user can provide input to the system through speech though WeaVe (Web voice enabling). WeaVe anticipates patterns of spoken input using a "speech grammar", a series of rules indicating what can be said at each Web page. The speech grammar is generated automatically by transforming XML content into speakable units.[9]

THE MODEL

In our current work, we are supporting clinicians who are actively engaged in work, which is when most information needs arise. To accomplish this, we must capture the user's question in the field in the easiest possible way, process the question centrally, and deliver a response at a later time when the user is free to review it. Our model leverages the expertise of human information providers (reference librarians), arriving at an automated system by incremental steps. The model involves three components (Figure 2):

1. **Question Analyzer**: identification of high-level information needs in natural language questions. Free-text requests are mapped into a high level semantic representation based on an ontology of clinical information needs. The mapping rules are acquired using machine-learning techniques.
2. **Search Strategy Generator**: High-level information needs are translated into complex search strategies that are adapted to user needs and capabilities of information resources. Models of complex, multi-resource search strategies are derived from observations of human search expertise. Translation rules are developed using machine learning.
3. **Response Delivery**: The information located by search strategies is extracted from one or more information resources. The pieces are integrated into a coherent response for delivery to the user, typically via email. Response can be viewed on mobile devise or workstations, depending on workflow.

Clinician information needs are captured through text and voice entry on handheld devices. These are transmitted to the Query Analyzer, which converts speech input to text, and text to a structured, semantic analysis of the information need. The structured query is then processed by the Strategy Generator, which retrieves information from various sources according to the specific needs specified by the user. Finally, the Response Delivery module packages and transmits the retrieved information to the user for viewing on handheld or desktop devices. In the early stages of the system development, the Strategy Generator module is simulated by a human reference librarian, who attempts to obtain answers to questions with the aid of automated search strategies. The architecture also enables the librarian to request clarification of user queries and obtain follow up information from the user via the same pathways. As the system becomes more automated, an increasing number of the questions will be answered without mediation by the human expert.
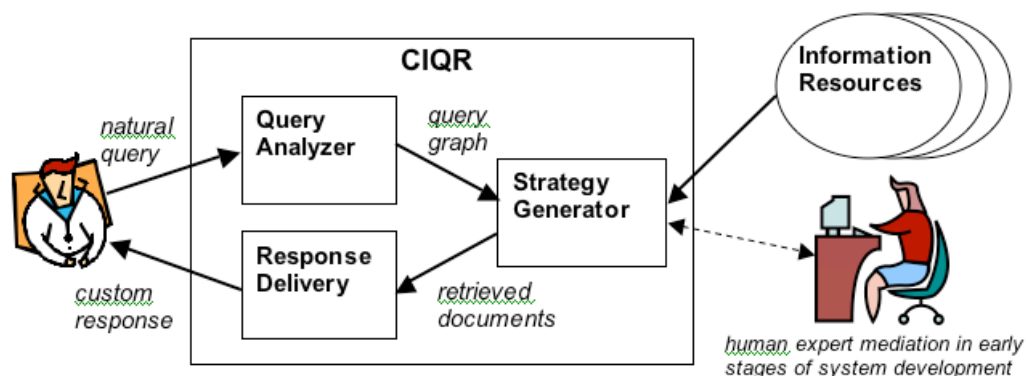

Figure 2. System Architecture of CIQR

CURRENT STATUS

The current project requires a deeper understanding of clinical workflow, how information needs arise during this process and what steps clinicians take in pursuing these needs. We have begun a series of studies that examine workflow in a variety of clinical settings: paramedics in the field, doctors and nurses in emergency departments, and physicians on internal medicine service. We are using observational techniques to detect emergence of information needs and to track the information seeking process. At present we have found that it is most effective if the observer can also serve as a participant in the clinical process, e.g. paramedic or medical student. This method makes clinicians being observed more comfortable, eases the process of obtaining study approval and provides additional insight into the subtleties of the clinical process.

In an ongoing study, we are using cognitive methods to investigate librarians' strategies for conducting clinically-related searches. This experiment is intended to simulate a client consultation in which a clinician needs to find specific information in order to treat or diagnose a patient. The librarians are presented with 3 brief clinical scenario followed by 3 questions for each scenario. They search for information using available search engines, such as in PubMed or OVID. The primary sources of data collected include: 1) the think-aloud protocol which is transcribed and time stamped, 2) video capture of subject and display, 3) output from the search, and 4) the stored search strategies, which can be downloaded as a text file. Our objective is not only to describe a search strategy, but to understand the thought processes that inform the various actions and decisions taken by the librarian. The data provides rich insight into the process of search and should prove to be useful in our endeavors to automate expert strategies.

In the current implementation, several components of the HINT system are being enhanced and adapted to the new architecture. In particular, we are analyzing free-text queries instead of medical records. In this model, the contextual information is determined in part from the question itself. Our analysis of actual queries occurring in the field show that they tend to be open-ended, multi-sentence questions that contain crucial information about the user's environment and goals. This module builds upon the AQUA parser[7]. AQUA does not have explicit grammar rules. Instead, it acquires its parsing knowledge using transformation-based learning (TBA).[10,11-13] This method requires a training set that pairs each user question with its target semantic analysis. AQUA learns a set of rules to transform free-text into a structured form, which can then be used to parse new questions.

The voice component to be developed will be constructed from commercial tools (ViaVoice, Dragon) and employ open-source solutions and standard protocols (e.g., XML, voiceXML) for the implementation of system's interfaces, following a model developed by Tachinardi and collaborators.[14] In our preliminary testing of the speech engines, we obtained 90-94% accuracy using a general English vocabulary. Even though many of the errors will be addressed by using an appropriate medical vocabulary, we believe the system will need to send the

transcribed text back to the user for verification. The user can edit and resubmit the request, replacing the original search. Our preliminary work with speech indicate that the number of errors per request are relatively minor (usually 1-2 words), suggesting that the request may be edited on a mobile device. More complex editing will probably require sitting down at a workstation, which will require integration with workflow.

DISCUSSION

Patient care is a complex, knowledge-intensive discipline. Even experienced clinicians have gaps in knowledge, which can be compounded by the ever-growing body of scientific evidence. Physicians in training have larger gaps because they are in the process of acquiring new knowledge and because they cannot always access knowledge when learning opportunities best present themselves. The focus of this project is the gaps that arise spontaneously during the course of clinical practice. These are particularly important, because they concern knowledge that is necessary to make decisions effectively, e.g., diagnosing a patient and ordering appropriate therapy. However, addressing these gaps can be extremely challenging because of the lack of timely access to knowledge. Frequently, when the need is recognized the clinician is too busy to pursue it, and when time becomes available, the need is already a distant memory.

CIQR provides an environment in which we can study how information needs arise during clinical practice and how they can be addressed through communication with a semi-automated reference library service. The advantage of the proposed model is that we can immediately begin to capture needs and deliver information for a small number of users. In this way, pieces of the communication process (query analysis, search strategy, and result display) are automated in an incremental manner as they are understood. In addition, CIQR is complementary to other methods of seeking information (e.g., electronic medical record or a drug reference), and therefore should not interfere with established patterns of workflow.

Our model raises several questions on the culture of information seeking. Intercepting information needs as they occur during clinical practice and addressing them in a timely manner that fits with clinical workflow may provide timely provision of information, optimizing patient management. The new model has also the potential to change the culture of clinical training, by encouraging trainees to identify gaps in their knowledge and to articulate their information needs. Additional questions can be raised on the innovative prototyping methodology for developing an information system that draws on the active participation of human information experts (reference librarians). While our ultimate goal is to develop a fully automated system, our approach leverages human abilities and progressively gives way to automation as the sophistication of the system evolves.

CONCLUSION

Information needs arise at critical junctures in the course of clinical work. Many of these needs go unmet and can potentially compromise patient care. Enabling clinicians to express their information needs in natural language while practicing can enhance their access to information and thus improve medical decision making. Current information retrieval systems do not afford ready access to relevant knowledge resources. In this paper, we present a novel approach to addressing information needs in the context of workflow. The model is based on prior research related to providing proximal access to knowledge at the point of care. We have a developed a multi-faceted framework which incorporates natural language processing, speech recognition, machine learning and information retrieval methods. The development work is richly informed by ethnographic studies of clinical practice, cognitive analysis of librarian search strategies, and real time capture of information needs using a range of input devices. The design of our research program has the potential to change the culture of clinical training, by encouraging residents to identify gaps in their knowledge and to articulate their information needs. If we are successful, this could represent a significant advance in evidence-based practice.

References

1. Mendonça EA, Cimino JJ, Johnson SB, Seol YH. Accessing Heterogeneous Sources of Evidence to Answer Clinical Questions. J Biomed Inform. 2001 Apr;34(2):85-98.

2. McKeown KR, Chang SF, Cimino JJ *et al*. PERSIVAL, a System for Personalized Search and Summarization over Multimedia Healtcare Information. Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries. 2001: 331-40.

3. Johnson SB, Campbell DA, Krauthammer M *et al*. A Native XML database design for clinical document research. Proc AMIA Symp. 2003: 883.

4. Mendonça EA, Cimino JJ, Johnson SB. Using narrative reports to support a digital library. Proc AMIA Symp. 2001: 458-62.

5. Mendonca EA. Using Automated Extraction from the Medical Record to Access Biomedical Literature. Doctoral Dissertation. New York: Columbia University, 2002.

6. Seol YH, Johnson SB, Cimino JJ. Conceptual guidance in information retrieval. Proc AMIA Symp. 2001: 1026.

7. Campbell DA, Johnson SB. A transformation-based learner for dependency grammars in discharge summaries. Proceedings of the ACL-02 Workshop for Computational Linguistics. Association of Computational Linguistics,

2002.

8. Campbell DA. A Natural Language Processing System to Assess User Needs in Information Retrieval. Doctoral Dissertation. New York: Columbia University, 2004.

9. Starren JB, Charney ML, inventors. A method and system for voice activating web pages. 60/250,809.12-1-2000.

10. Brill E. A report of recent progress in Transformation-based Error-driven Learning. Proceedings of the ARPA Workshop on Human Language Technology. 1994.

11. Ken S, Carberry S, Vijay-Shanker K. An Investigation of Transformation-Based Learning in Discourse. Machine Learning: Proceedings of the 15th International Conference. 1998.

12. Higgins D. A transformation-based approach to argument labeling. Proceedings of CoNLL-2004. 114-7.

13. Ramshaw LA, Marcus MP. Text Chunking Using Transformation-Based Learning. Proceedings of the Third ACL Workshop on Very Large Corpora. 1995.

14. Tachinardi U, Gutierrez MA, Pires FA, Kobayashi LO, Furuie SS. Abstracting complexities to support development and deployment of healthcare applications on multiple mobile devices. Medinfo 2004. 2004: 1876.