The Cognitive Demands of an Innovative Query User Interface

Di Wang[*], M.D., David R. Kaufman[*†], Ph.D., Eneida A. Mendonça[*], M.D., Yoon-Ho Seol[*], MS, Stephen B. Johnson[*], Ph.D., James J. Cimino[*], M.D.
Department of Medical Informatics[*] and Laboratory of Decision Making and Cognition[†]
Columbia University, New York, NY, USA

Too often, online searches for health information are time consuming and produce results that are not sufficiently precise to answer clinicians' or patients' questions. The PERSIVAL project is designed to circumvent this problem by personalizing and tailoring searches and presentation to the demands of the user and the particular clinical context. This paper focuses on a cognitive evaluation of one component of this project, a Query User Interface (QUI). The study examines the system's ability to allow users to easily and intuitively express their information needs. We performed several analyses including a cognitive walkthrough of the interface and quantitative estimations of cognitive load. The paper also presents a preliminary analysis of usability testing. The analyses suggest that there are features in the QUI that contribute to a greater cognitive load and result in greater effort on the part of the subject. The results of usability testing are consistent with these findings. However, subjects found it to be relatively easy and intuitive to generate well-formed queries using the interface. This study contributed to the iterative design of the interface and to the next generation of the PERSIVAL system.

INTRODUCTION

Clinicians are often confronted with unpredictable healthcare information needs that give rise to questions that need to be addressed in a timely fashion. Patients need specific information to understand and better manage their health problems. Although the World Wide Web has greatly increased the accessibility of keyword searches of online databases such as PubMed, the search results are often less than satisfactory.[1]

Salton[2] characterize 4 stages of the information retrieval (IR) process, including indexing, query formulation, retrieval, and evaluation and refinement. Studies[3] have documented problems related to access to information at all stages of the IR process. Research on evidence-based medicine indicates that the query formulation stage is especially problematic, in that it is often rather difficult to articulate a clear clinical question.[4,5] Even when there is no problem in generating a well formed question, there are significant challenges pertaining to the evaluation and refinement of searches. Confronted with the enormous quantity of biomedical literature, it is difficult for clinicians to filter the immense amounts of information to incorporate relevant evidence to make safe and accurate diagnostic, therapeutic and management decisions.

One solution to these problems is to develop evidence-based decision support applications to provide relevant and up-to-date evidence to clinicians. PERSIVAL (**PE**rsonalized **R**etrieval and **S**ummarization of **I**mage, **V**ideo **A**nd **L**anguage Resource), a system that provides personalized access (by including patients' individual characteristics) to a distributed digital library is designed to facilitate this process.[6] Components in this system include Health Information Needs Tailoring (HINT), Multimedia search, Summarization and Presentations Interface.[7]

In this paper, we describe one of the three components of HINT, the Query User Interface (QUI). Links to QUI are intended to be placed in the clinical information system such that a user can evoke it in the context of an information need related to specific patient data. On the basis of a particular clinical context, the QUI displays a set of automatically generated structured clinical questions such as "what is the best drug for this condition". The physician would then select from a list of questions, the one that best matches his/her information needs.

Systems such as PERSIVAL, which can readily address the information needs of the clinician in real-time, offers significant promise in advancing evidence-based medicine practices. However no matter how impressive the technology, its effective implementation is predicated on its acceptance by the user population. Ease of use and learnability are important attributes in gaining such acceptance. It is therefore important that a query interface presents users with context-sensitive questions that correspond to their information needs and sufficient flexibility and choice to formulate precise queries. If the system lacks this flexibility or imposes significant cognitive demands, then it is unlikely to gain acceptance. This paper presents a cognitive evaluation of the QUI component of the system. The objective is to assess whether the structured interface is sufficiently intuitive for clinicians to construct well-formed questions with a minimum of effort.

THE QUERY USER INTERFACE

Architecture:

The HINT architecture is shown in Figure 1. The QUI is the front-end component of two other components, namely, the Context Extraction and the Query Refinement.

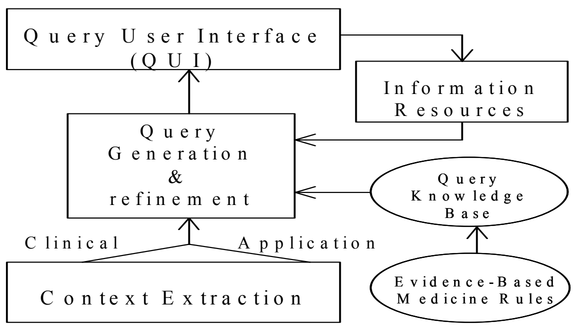


Figure 1. HINT architecture

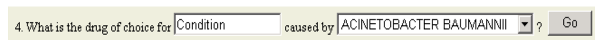
The workflow is as follows: When users view patient data and need information to make clinical decisions, they can invoke HINT to initiate a search. The Context Extraction component produces relevant concepts from *Application Context* (the specific patient data the user is viewing) and *Clinical Context* (all other data in patients' record) then output them to the *Query Refinement* component, which produces query graphs and calls the QUI component.

QUI uses Perl to process the query graphs and dynamically displays structured queries to the user. JavaScript is used to better display certain types of questions. HTML features such as pull down menus, checkboxes, text fields and pop-up windows are used to meet different query display needs. The queries are displayed in natural language style with 1-4 concepts to be selected or entered by users.

HINT produces data with relevant patient's information, user's profile and user's full question and sends the data to the search component after a user completes a question and initiates a search request. The search result is presented to the user and also to the *Query Refinement* component to improve query. Data transfer in the PERSIVAL system is achieved via the use of XML.

Display and Usability Issues

The QUI faces significant challenges in providing the users with sufficient flexibility to generate a pertinent query with a minimum of effort. QUI is designed to adhere to a set of usability standards.[8] The interface uses a range of HTML resources (widgets) for question completion. In the following example, a physician can query the system about a drug of choice by a) typing in a condition (e.g., disease) and b) selecting an organism from the pull-down menu. In this case, the organism choices are drawn from data the user was viewing (a culture result) and the question was drawn from the *Query Knowledge Base* because it was deemed relevant to the culture result.



In the cases where there are multiple concepts in the query graphs, if only one concept element fits into the question context, the pull down menu is used. When multiple concepts fit into the context, a JavaScript popup window is displayed and the user can select multiple choices. To let users enter their own concept value to fit in a query context, a text-field (i.e., "fill in the blank") is the simplest and most intuitive feature to use. A word closely relevant to a semantic type for a text-field will be displayed in the text-field as a hint for users to enter a word of same type. A screen shot illustrating some features of the query interface is presented in Figure 2. The pop up window (Message window) is a JavaScript feature that enables multiple concept selection.

Although the interface adheres to usability guidelines, there are many empirical questions to address concerning the ease of use and intuitiveness. For example, what is a reasonable number of

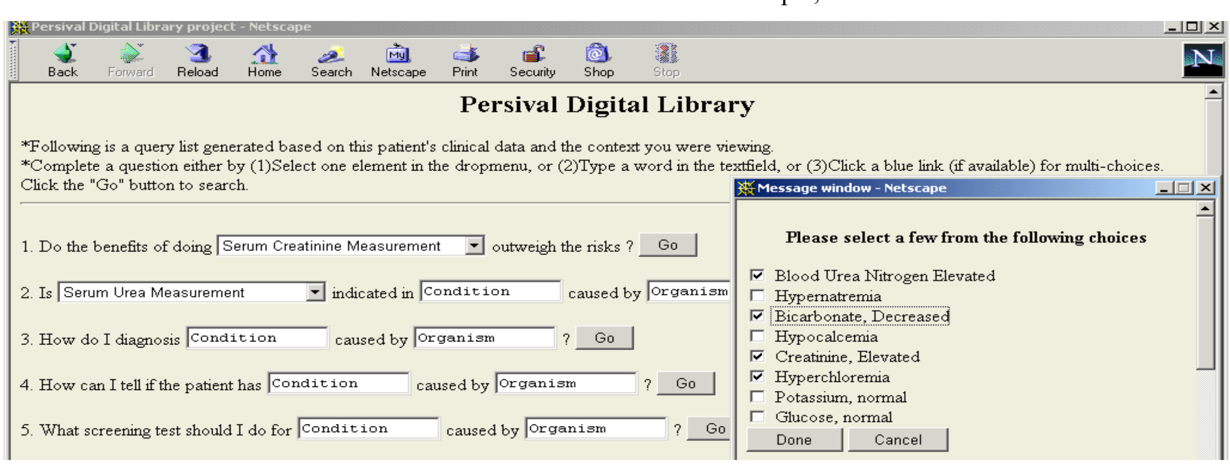


Figure 2. Example of the query user interface

questions to present to the user? Currently the system presents as many as 20 questions. What are some of the challenges involved in constructing well-formed questions using different widgets? For example, can users readily combine the actions of pull-down menus and text-fields? Each of these elements makes distinctively different cognitive demands on the user.

METHODS

We performed a series of a usability inspection analyses drawing on a cognitive task analytic methodology that is described below. In addition, we conducted a usability test experiment to evaluate the challenges of developing well-formed questions.

Cognitive Task Analyses:

Cognitive task analyses allow us to define the demand characteristics of a particular task and to focus attention on the pertinent problem dimensions.[9] In this circumstance, the assumption is that the user has limited attentional and cognitive resources for constructing well-formed queries. For example, menu items of different lengths and text-fields instantiating particular UMLS semantic types (e.g., pharmaceutical agent) are likely to make different demands on the user. This is referred to as cognitive load or mental workload, which is commonly used to characterize the difficulty of a system or interface in human factors research.[10] To begin to assess the cognitive demands of the system, we quantified the particular concept-slots of the interface for specific clinical contexts. We selected four contexts that were representative of the variety of possible interfaces that the system could generate. These include two laboratory contexts and two microbiology contexts.

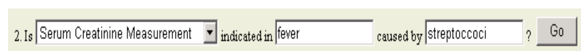
We also employed a usability inspection analysis methodology known as the cognitive walkthrough (CW). The purpose of a walkthrough is to evaluate the process by which users perform a task and the ease with which they can do this. The CW methodology involves identifying sequences of actions and goals needed to accomplish a specific task and provides a reasonably good sense of the difficulties involved. In this paper, we illustrate the results of the CW for a single question.

Usability Testing Experiments:

We are in the process of conducting a series of usability testing experiments. We present preliminary results of the first experiment. In the first study, 6 physicians of varying levels of expertise participated in the study. Subjects were asked to complete 2 tasks in 2 distinct contexts. In addition, subjects were asked to complete a survey that evaluated the interface on several dimensions of interest.

Procedure:

The subjects were asked to complete both tasks in 1) a laboratory test context (metabolic panel) and 2) a microbiology test context (smear culture). In the first task, subjects were initially presented with a clinical context display indicating several test results, including some of which were abnormal. After clicking on the "digital library" link the QUI displays a set of system-generated questions. Subjects are then presented with 5 questions on a sheet of paper and required to match these questions with the appropriate ones on the interface. For example, one of the questions asked, "should I measure serum creatinine in fever caused by streptococci?" The questions were phrased in such a way that they necessitated a certain transformation to instantiate them in the QUI. The following item (question 2) represents the closest match:



In the second task, subjects were asked to spontaneously generate 5 specific well-formed and clinically sensible questions. Throughout the task, subjects were asked to think-aloud; when they failed to do so, they were prompted by the experimenter to voice their thoughts.

Subjects were videotaped and the computer screens were captured by a digital camera using a digital analog NTSC converter.

Questionnaire:

The questionnaire consisted of 18 Likert items in which subjects were asked to express their agreement (or disagreement), on a 5 point scale, with statements about their overall impression of the QUI, the screen layout, learnability, and ease in which they could generate questions. In addition, subjects were presented with 5 different question types and asked to identify which was easiest to use to formulate a question and which was most difficult.

RESULTS

Cognitive Task Analysis:

The results of the cognitive task analysis are presented in Table 1. The average number of concept slots per page (first 10 questions) was 20.25 (ranging from 17 to 27). Each distinct widget makes unique demands on the user. For example, a free text field necessitates only 2 user actions: supplying the concept and clicking on go. However, it requires additional cognitive effort to generate the concept from one's knowledge. A pull-down menu requires the user to click the arrow, scan the list of items (of varying length), and select an item. This involves more actions, but relies on recognitional memory, which substantially reduces the cognitive burden.

However, these menus are of varying length (from 3 to 20 items) and it's reasonable to assume that the longer lists place additional demands on working memory. Pop-up menus, activated by a link on a page, necessitate a minimum of 6 actions, but once again rely on recognitional memory. Compound questions combine the various widgets and are likely to result in a greater exertion of effort. On the basis of our analysis, both microbiology contexts (especially Mic 1) are likely to result in additional effort by subjects since it requires more actions to initiate a search and there are more concept slots on the page.

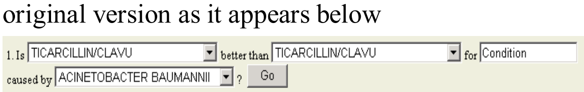
Table 1: Analysis of Interface Complexity

Clinical context	Lab 1	Lab 2	Mic 1	Mic 2
Concept Slots	19	17	27	18
Mean actions to start search	5.0	3.8	8.1	5.4
Free text fields	15	15	10	10
Pulldown menus	2	2	17	8
Average Depth	7.0	5.5	11.5	10.0
Popup windows	2	0	0	0
Average Depth	8	0	0	0
Compound Questions	7	8	10	7

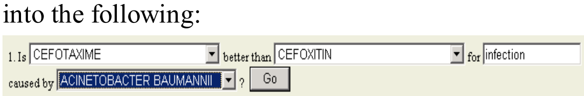
The analysis is additionally complicated by the fact that different fields employ different semantic types. For example, it may be easier to generate a hypothesized condition slot (i.e., disease/syndrome) than a microorganism slot. There are many issues to resolve and we are beginning to address some of these in user testing, which are described below.

Cognitive Walkthrough of A Complex Question:

In this section, we present a cognitive walkthrough of the most complex question presented to users. They are provided with the following question "For Cefotaxime and Cefoxitin, which one is better in treating the Acinetobacter Baumanii infection." There are 2 subprocess involved in this task: 1) find the matching question and 2) instantiate the concept slots with the appropriate values. Q1 is the correct question and users must transform the original version as it appears below



into the following:



The following CW analysis illustrates the complex sequence of Goals and Actions necessary to formulate a question.

Subprocess 1: Find a Matching Question
Goal: Select the question from the list that is likely to best match the stimulus question.
Action: Scan the list
Subgoal: Discriminate among questions that have very similar variable or concept slots
Goal: Find questions that address therapeutic concerns
Action: Evaluate questions with therapeutic content
Problem: 9 of 11 questions are therapeutic queries.
Subgoal: Select alternative strategy to select question
Action/Inference: The question compares the efficacy of 2 drugs
Subgoal: Identify comparative questions
Action: Recognize that Q1 and Q2 are comparatives
Subgoal: Discriminate between Q1 and Q2
Action: Q1 contains a matching literal comparative term "better", so select 1

Subprocess 2: Instantiating the Concept Slots
Subgoal: Instantiate first concept slot
Action: Click on pull-down menu
Subgoal: Find "Cefotaxime"
Action: Scroll to the 14th item on a list of 20
System Response: Cefotaxime appears in the first concept slot
Sub-goal: Instantiate second concept slot
Action: Click on pull-down menu
Subgoal: Find "Cefoxitin"
Action: Scroll to the 13 item on a list of 20
Subgoal: Instantiate the fourth concept slot
Action: Note the desired organism "acinetobacter baumanii" is in the default "organism" slot.
Subgoal: Complete Query
Action: Click on "Go"

The analysis reveals a question of immense complexity necessitating 12 goals and subgoals and 11 corresponding actions. Although it is difficult to ascertain at this point how much effort will be tolerated by clinicians, there is no doubt that formulating a well-formed query in this context involves a substantial cognitive workload.

Results of the Usability Testing:

There was substantial variability in how the six subjects approached the tasks. The sessions ranged from 20 to 60 minutes. All subjects were able to complete most questions without too much difficulty, although some questions were significantly more difficult. Subjects were able to correctly instantiate 54 out of 60 questions (90%). Question 4 in the microbiology context (illustrated in the CW analysis) proved to be the most difficult.

Table 2: Average Time to Complete Questions

	Lab Context	Micro Context
Average Total	34.77 (19.05)*	34.78 (16.33)
Matching	21.57 (14.89)	19.71 (11.70)
Instantiation	12.67 (7.34)	14.59 (8.61)

*Standard Deviations are in parentheses

As indicated in Table 2, the average time to complete a question was 34.8 seconds for the laboratory context (SD=19) and 34.7 for the microbiology context (SD=16.3). As the standard deviations suggest, there was considerable individual differences as well as differences across question types. In general, the questions with a lower cognitive load required less time to complete. The average time needed to identify a matching question was greater across both contexts than the time needed to instantiate the slots. Not surprisingly, the questions with more slots needed more time to instantiate. In addition, certain semantic types also needed more processing time.

Responses to Survey:

The responses to the survey items are summarized in Figure 3. All subjects rated the QUI as easy to learn (learnability) and suggested that questions were easy to generate. Four subjects rated the overall QUI very highly and 2 others rated it rather harshly. A similar split was observed in the ratings of the screen layouts.

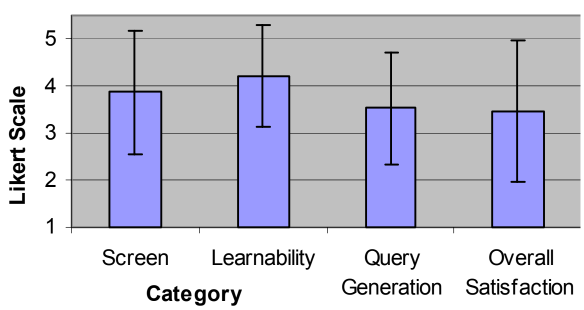


Figure 3. QUI interaction satisfaction

DISCUSSION

It is increasingly recognized that clinicians need access to quality information resources at the point of care and that such access may sharply reduce medical errors and improve health outcomes. This philosophy is embodied in the HINT system. Yet there are numerous unresolved issues about how to facilitate such a process. This paper presents a cognitive evaluation of one component of the HINT system, the query user interface. The evaluation included usability inspection methods and user testing.

The results of these analyses suggest that the QUI may be a promising vehicle for the unproblematic formulation of information retrieval queries. However, the current version of the system imposes a significant cognitive load on the user, rendering the task more difficult than necessary. The study suggests several ways in which the QUI can be improved. These include a better layout with fewer questions, less compound questions, the elimination of redundancy and a reduction in the number of actions to initiate a search. From a methodological vantage point, we need to further refine the notion of cognitive load as it refers to the demands of the search process and correlate quantitative parameters with subjective judgments and performance parameters (e.g., accuracy and efficiency). The research provisionally suggests the considerable promise of the QUI approach in providing timely information to address complex clinical decisions.

Acknowledgments

The authors thank Dr. Vimla Patel and Dr. Alla Keselman for their contributions to this project. This work is supported by a Digital Library Initiative grant from the NSF DLI Phase 2 (IIS-98-17434).

References

1. Price SL, Hersh WR. Filtering Web pages for quality indicators: an empirical approach to finding high quality consumer health information on the World Wide Web. Proc AMIA Symp 1999: 911-5.
2. Salton G. Introduction to Modern Information Retrieval, New York: McGraw-Hill; 1983
3. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? Annals of Internal Medicine 1985; 103(4): 596-9.
4. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 2nd ed. London: Churchill Livingstone, 2000.
5. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995: 123: A12-3.
6. McKeown K, Chang S-F, Cimino JJ, Feiner S, Friedman C, Gravano L, Hatzivassiloglou V, Johnson SB, Jordan D, Klavans J, Kushniruk A, Patel V, Teufel S. PERSIVAL, a System for Personalized Search and Summarization over Multimedia Healthcare Information. Proceedings of the 1st JCDL. 2001. Roanoke, USA.
7. Mendonça EA, Cimino JJ. Building a Knowledge Base to Support a Digital Library, Proceedings of the Tenth World Conference on Medical Informatics, MEDINFO, 2001.
8. Shneiderman B. Designing the user interface. 3rd ed. Addison-Wesley, 1998:
9. Patel VL Arocha, JF, Kaufman DR. A primer on aspects of cognition for medical informatics. Journal of American Medical Informatics Association. 2001. 8: 324-343.
10. Rasmussen J. Information processing and human-machine interaction. An Approach to cognitive engineering. 1986. New York. North-Holland.