Research Paper ■

Providing Concept-oriented Views for Clinical Data Using a Knowledge-based System:

An Evaluation

QING ZENG, PHD, JAMES J. CIMINO, MD, KELLY H. ZOU, PHD

Abstract **Objective:** Clinical information systems typically present patient data in chronologic order, organized by the source of the information (e.g., laboratory, radiology). This study evaluates the functionality and utility of a knowledge-based system that generates concept-oriented views (organized around clinical concepts such as disease or organ system) of clinical data.

Design: The authors have developed a system that uses a knowledge base of interrelationships between medical concepts to infer relationships between data in electronic medical records. They use these inferences to produce summaries, or views, of the data that are relevant to a specific concept of interest. They evaluated the ability of the system to select relevant information, reduce information overload, and support physician information retrieval.

Measurements: The sensitivity and specificity of the system for identifying relevant patient information were calculated. Effect on information overload was assessed by comparing the amount of information in each view with the amount of information in the entire record. Information retrieval accuracy and cost (time) were used to measure the effect of using concept-oriented views on the efficiency and effectiveness of retrievals.

Results: The sensitivity and specificity of the system for identifying relevant clinical information were generally in the range of 70 to 80 percent. Concept-oriented views are effective in reducing the amount of information retrieved (over 80 percent reduction) and, compared with source-oriented views, are able to improve physician retrieval accuracy ($p = 0.04$).

Conclusion: Computer-generated, concept-oriented views can be used to reduce clinician information overload and improve the accuracy of clinical data retrieval.

■ **J Am Med Inform Assoc.** 2002;9:294–305.

In medical practice, information overload is a common problem for clinicians who are confronted with large amounts of patient data in paper charts or electronic medical records.[1] Since clinicians have limited time to review and process the data, information overload may result in errors during the information retrieval and decision-making processes. Furthermore, depending on the clinical tasks, often only certain subsets of the data (which we refer to as views) are of interest to clinicians. Providing appropriate views can be a way to address the problem of information overload.

All views of medical data can be categorized into three classes—source-oriented views (which organize data on the basis of where they were collected); time-oriented views (which primarily use time to organize data); and concept-oriented views (which center on clinical concepts, such as diseases or organ systems).[2] Our work has focused on concept-oriented views.[3] One of the most famous examples of a concept-oriented view is the problem-oriented medical record.[4]

Affiliations of the authors: Harvard Medical School, Boston, Massachusetts (QZ, KHZ); Columbia University, New York, New York (JJC).

Correspondence and reprints: Qing Zeng, PhD, Decision System Group, Thorn Building #309, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115; e-mail: <Qzeng@dsg.harvard.edu>.

Received for publication: 8/24/01; accepted for publication: 12/26/01.

The place and time of data observations are generally recorded and can be used to generate source- and time-oriented views. The relationships between a piece of data and various concepts are typically not explicitly recorded and need to be created, obtained, or inferred to generate the desired concept-oriented views. It can, therefore, be challenging to generate and maintain concept-oriented views of a large amount of patient data on a routine basis. For instance, physicians do not always indicate which medication prescriptions are for which specific chief complaints. Requiring physicians to manually link data with specific concepts for the purpose of creating views is impractical, as it will definitely slow down the workflow.

With manually maintained links, providing multiple concept-oriented views for the same data is extremely difficult. Although physicians are the best source for case-specific relationships, such as the reason why a medication is given to a patient, it is inefficient for non-case-specific relationships to be specified repeatedly. Furthermore, the consistency of the views is subject to human errors and interpersonal differences.

Previous research has shown that it is possible for computer systems to identify relevant data and generate views.[5,6] Although the Physician Workstation System was a pilot system, it demonstrated the ability to generate patient-specific and disease-specific views using domain knowledge modeled in a causal network.[6]

Simulation studies have also been conducted to assess the effects of different clinical data formats on the speed and accuracy of information retrieval.[7–9] These studies have shown that concept-oriented views may improve information retrieval and even medical decision making.

Especially worth noting is the evaluation by Tange et al. of the effect that different views of medical narratives have on the speed and completeness of information retrieval.[9] This quantitative study demonstrated the value of problem-oriented and organ system–oriented views.

However, the quality of computerized identification of relevant data and the effects of providing concept-oriented views generated through such identification have rarely been evaluated. Because computerized identification of relevant data can never be perfect, the accuracy, reliability, and effects of the technique need to be measured to provide a basis for further improvement and to determine the value of the method in production systems.

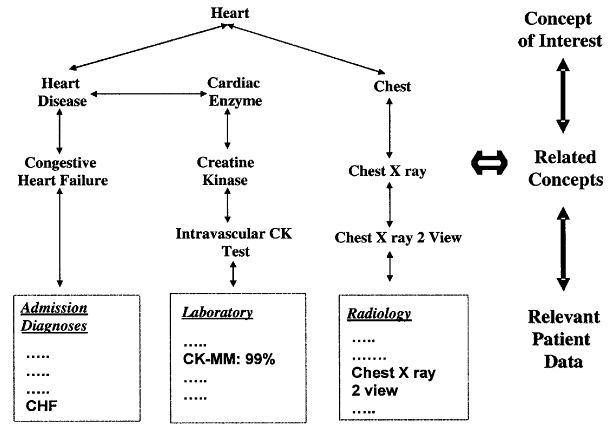

Figure 1 Chart showing how a concept (heart) could be linked to the related clinical data and the general model of linking a medical concept to relevant patient data.

To test several hypotheses about computer-generated concept-oriented views, a general-purpose, concept-oriented view-generation system was developed. We refer to the system as the Query Clinical Information System (QCIS). The system is also capable of handling source- and time-oriented views.

A knowledge-based approach was employed in the system design and implementation.[3,10,11] We discuss the evaluation of the QCIS ability to identify relevant information for concept-oriented view generation, to reduce information overload, and to benefit clinical practice. Although some raw data of the evaluation have been reported before,[12] here we present new data, different analytic methods, and an overall view of the evaluation.

Background

We have developed the QCIS—a multiple-view (source-oriented, time-oriented, and concept-oriented) generation system. On selecting a specific patient, users have the freedom to review the clinical data using one of the three views. The design for providing multiple views was based on the hypothesis that different views complement each other and could be compared. The emphasis of the system is to provide concept-oriented views that are generated by computer identification of data relevant to a user-specified concept of interest. Figure 1 illustrates the key process of linking a concept of interest to relevant patient data.

In the QCIS, the concept-oriented views are generated in four steps—concept selection (selecting the concept of interest), concept expansion (finding concepts related to the concept of interest), data retrieval (obtaining

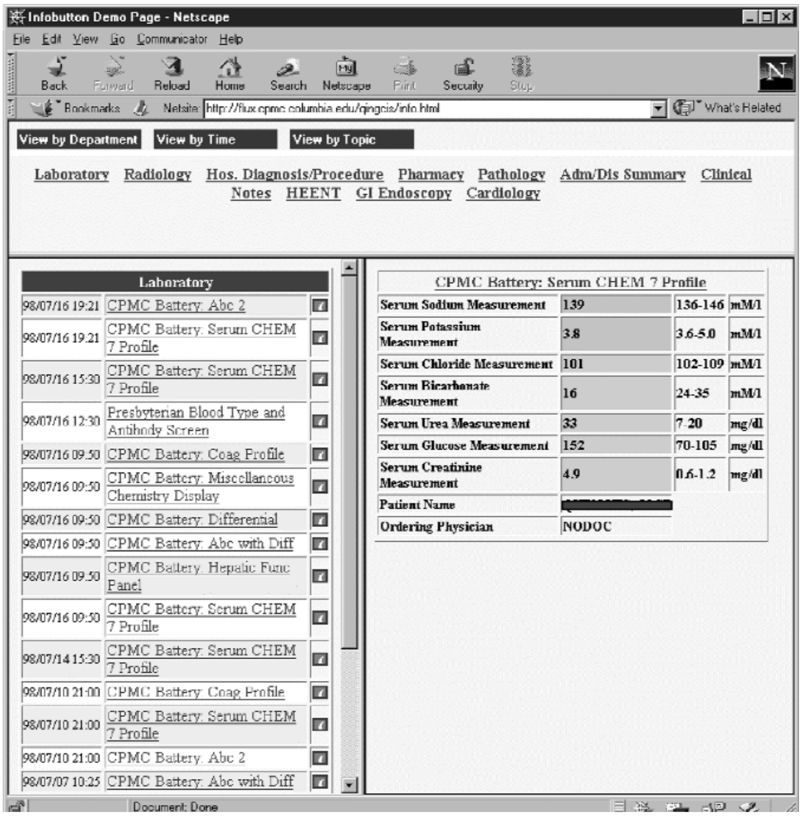

Figure 2 Three types of views were available for users. When View by Department was chosen, a list of clinical departments was shown. After the laboratory department was selected, an index of laboratory reports was displayed. The details of a laboratory report were displayed after a click on the name of the report.

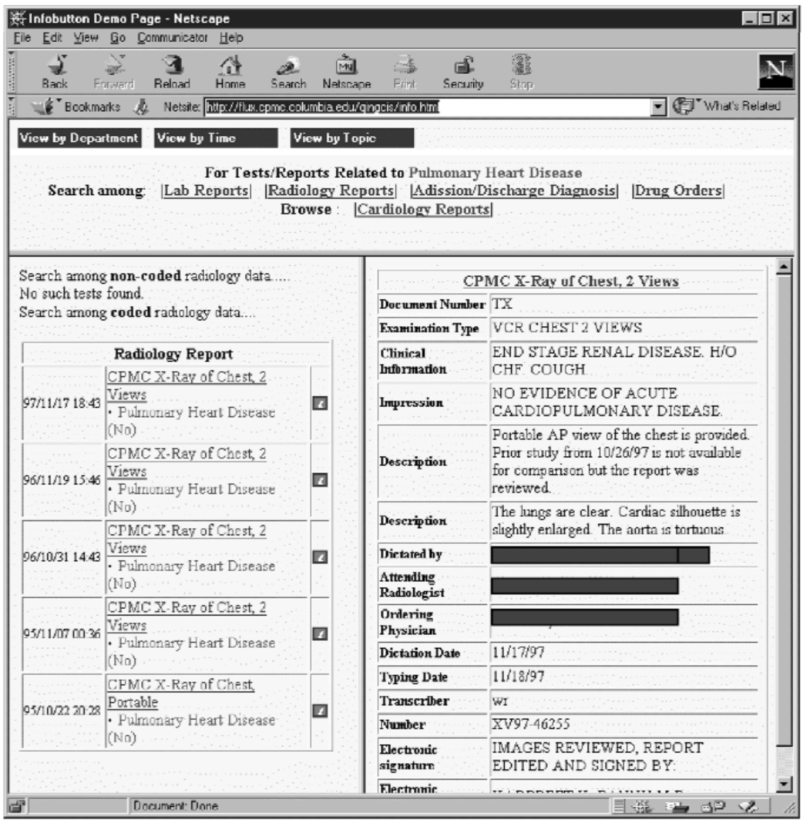

Figure 3 When using concept-oriented views (view by topic), a user must enter a term of interest and selected from a list of matched concepts. This figure shows that after Pulmonary Heart Disease was typed in and chosen from the list of matching concepts as the concept of interest, a list of departments was shown. After selecting Radiology Reports, the system returned a list of radiology reports related to Pulmonary Heart Disease, and the content of a report was displayed after clicking the report name.

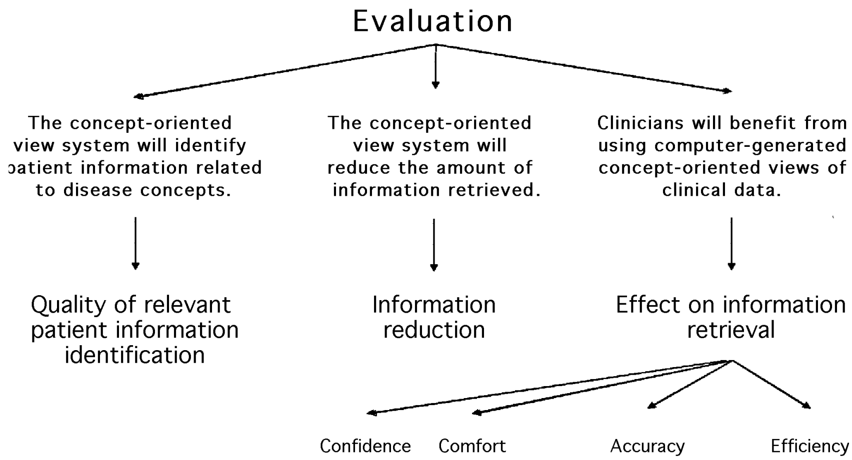

Figure 4 The structure of the evaluation of the QCIS.

information from the patient record that is coded with the expanded set of concepts), and display generation (producing a human-readable set of results).

The implemented system is capable of providing concept-oriented views for eight classes of concepts in the Medical Entities Dictionary (MED)[13]—anatomic entity, measurable entity, specimen, patient problem, sampleable substance, display information, event information, and orderable entity. This accounts for more than 40,000 concepts, which are about 76 percent of all concepts in the MED. The source of the clinical data for the QCIS is the central clinical data repository at New York Presbyterian Hospital (NYPH). Figures 2 and 3 show source (department)-oriented and concept (disease)-oriented views.

Hypotheses

With the evaluation, we tested the following three hypotheses:

- The concept-oriented view system identifies patient information related to disease concepts.
- The concept-oriented view system reduces the amount of information retrieved.
- Clinicians may benefit from using computer-generated concept-oriented views of clinical data.

Evaluation

This study evaluated the ability of the system to identify relevant patient information and the effects of the resulting views on clinical information retrieval. The evaluation was divided into three parts corresponding to the three hypotheses (Figure 4):

- *Quality of relevant information identification*—to determine the sensitivity and specificity of relevant information identification in certain clinical contexts such as validation of diagnostic hypothesis or disease management
- *Reduction of information overload*—to determine the degree of reduction of the amount of information in the concept-oriented views compared with views of the entire patient record
- *Effect on information retrieval*—to determine whether there is any advantage in using the concept-oriented views to retrieve information for patient care purposes, compared with using more traditional ancillary department-oriented views

Methods

Quality of Relevant Information Identification

This part of the evaluation was designed to measure the sensitivity of the system (true-positive fraction) and specificity (true-negative fraction) in identifying patient information related to concepts of interest.[14] Here diseases were used as concepts of interest, and laboratory tests and medication orders were used as patient information.

A set of patients ($n = 693$) was randomly selected from the NYPH database of all patients who had hospital visits between September 1996 and September 1998. We first randomly selected a set of patient visits and then identified the set of unique patients by their medical record numbers. The patients' laboratory test results, medications, and diagnoses during this time were also retrieved.

To derive the gold standard for relevant medications, we used the MICROMEDEX drug index, a widely used knowledge base of drugs and clinical review articles.[15] From the diagnoses in the patient data set, seven diseases (pulmonary embolism, tuberculosis, hypertrophic cardiomyopathy, acute renal failure, heart failure, pancreatitis, and multiple sclerosis) with review articles indexed in MICROMEDEX were randomly selected. For each of the seven diseases, all drugs mentioned in the disease reviews were extracted to form the gold standard, regardless of comments on their efficacy.

The following are a few lines from the clinical review for thromboembolic stroke treatment:

> H. THROMBOLYTICS:
>
> 1. ALTEPLASE (rtPA):
>
> a. INDICATIONS: Recommended treatment within 3 h of onset of acute ischemic stroke, IF emergent ancillary care and facilities to manage bleeding complications are readily available. Intracranial hemorrhage and changes indicating a large stroke must be excluded by head CT prior to administration. Do not treat strokes older than 3 h or massive in size.

Gold standards for relevant laboratory tests were established by a survey of six physicians. Because the relevance between diseases and laboratory tests can vary between the task of disease diagnosis and the task of disease management, we established a gold standard for each task. From the diagnoses in the patient data set, five diseases (sickle-cell anemia, congestive heart failure, coagulation defects, diabetes mellitus, and acute pancreatitis) discussed in *Harrison's Principles of Internal Medicine*[16] were randomly selected. A physician expert compiled a list of candidate laboratory tests by taking the union of all potential laboratory tests identified, using *Harrison's Principles of Internal Medicine* for each of the five diseases.

The surveyed physicians were given two questionnaires, each containing a matrix of the candidate tests and five diseases. On one questionnaire they were asked to identify the tests they might order to confirm the diagnosis of each disease. On the second questionnaire they were asked to identify the tests useful for the management of each disease. They were also asked to suggest any tests that were not included in the candidate list. We calculated a set of value for the gold standard: When four of six physicians agreed that a test was relevant, the test was considered to be relevant.

Because most terms used in laboratory and drug gold standards could not be mapped directly to the coding concepts of our patient data set, a translation process was done semi-automatically to make sure that the gold standards could be applied to the data set. For the drug gold standard, terms used in the MICROMEDEX clinical review articles were first mapped manually to MED concepts. Then a program was used to translate these concepts to concepts used to code the medication orders. For example, rifampin, which was mentioned in the clinical review article for tuberculosis treatment, was first mapped to the MED concept with the same name and then translated to a group of medication concepts such as CPMC DRUG: RIFAMPIN 150 MG CAP.

Although it was possible to use controlled vocabulary in composing the questionnaires for a laboratory gold standard, general terms such as "urinalysis" were used for physician convenience. The terms used in the questionnaires were mapped to classes of concepts before we translated the concepts to laboratory test concepts using a computer program.

For each disease, both the gold standard and the QCIS were used to identify relevant medications in the patient data set. Sensitivities and specificities were then computed empirically. Similarly, one set of diagnostic and management gold standards (four of six agreement) and the system were used to identify relevant laboratory tests in the patient data set. Sensitivities and specificities (diagnostic and management) of the system were calculated on the basis of the gold standards.

Relevance is a fuzzy concept, so statistics were collected on how often a particular number of physicians (0 to 6) considered a test relevant, to provide a sense of the "fuzziness."

Information Overload Reduction

This part of the evaluation was designed to measure the degree of reduction of the amount of information in the concept-oriented views, compared with views of the entire patient record. Using the same data set created for the relevant information identification, 21 diseases were randomly selected from 1,094 diagnoses in the test data set. The system was then used to identify relevant laboratory tests and medications for each disease. We calculated the average number of laboratory tests or medications included in each disease-specific view and the total number of laboratory tests or medications for each patient.

Certain types of hospital visits involved procedures, such as routine physical exam or exercise test, and produced no laboratory tests or medication orders. Patients with no laboratory tests or medication orders were excluded. Among the 653 patients, 444 had laboratory tests and 71 had medication orders.

Effect on Information Retrieval

Evaluation on the effect of using concept-oriented views involved several aspects of information retrieval—accuracy, user comfort, user confidence, and cost (time used). Both subjective and objective measures were taken.

We selected three patient cases with moderate amounts of information, and formulated questions necessitating patient information retrieval. To protect the privacy of the patients, all confidential information was manually altered. On the basis of an information needs study we conducted,[11] questions for the three patient cases were designed by a physician who was not involved in the system development and was not a physician subject later in the study. The questions were phrased in such a way that they had definite correct answers that could be expressed in simple words. For example, for a patient discharged on digoxin, the question "What was the latest digoxin level?" was asked. The questions did not reflect all the identified information needs, because certain needs could not be satisfied by the available patient data in the QCIS. For instance, information about health care provided by providers outside of the NYPH system was generally unavailable.

Thirteen physicians volunteered for the study. Physician subjects were given a brief description for each case and then answered questions by using the system to gather necessary patient information (Appendix). As mentioned before, the system could offer two views—the traditional department-oriented view and the newer concept-oriented view. Using the latter view, users could specify a concept of interest and review computer-identified relevant information. For each case in our study, a physician was instructed to use only department-oriented views or only concept-oriented views.

We adopted a design to ensure that every case was solved an equal number of times, that the department-oriented view and concept-oriented view approaches were used overall an equal number of times for each case, and that a physician never encountered the same case twice.

Prior to working on the cases, each physician was given a 5-min introduction and was informed of the purpose and methods of this study. Basic functionality of the system was demonstrated by the investigators using a specially prepared sample case.

During the evaluation, physicians provided written answers, which were later graded according to the gold standard. The answers were scored as correct or incorrect, and no partial credits were given. A Fisher exact test was performed to determine whether using one particular type of view lead to better accuracy, i.e., higher proportion of correct answers.

In some cases a physician could not answer questions because of unexpected technical problems. For example, the network connection between the outpatient clinic, where part of the evaluation took place, and the computer where the system is located was occasionally unstable. Fortunately, such incidents were rare, and the unanswered questions were regarded as missing data points and excluded from analyses.

After answering the questions, physicians were also asked to fill out a brief one-page questionnaire. The questions were designed to collect the background information on the physicians and their subjective opinions on how comfortable and confident they were using different types of views. The questionnaire allows physicians to provide additional comments and suggestions.

Efficiency is another measurement for information retrieval. To measure efficiency, we employed the portable usability laboratory at NYPH.[17,18] Throughout the evaluation, interactions between the physician subjects and the system were videotaped. Using these tapes, we measured the length of time it took for each person to answer a question. As a general principle, the counting for each question started after a person finished reading a question and ended when a person finished writing an answer. For the information retrievals to be performed more naturally, and because user interface was not the focus of this study, the physicians were not asked to think aloud. However, the physicians did sometimes make verbal comments and ask questions.

Results

Quality of Relevant Information Identification

The sensitivities (true-positive fraction) and specificities (true-negative fraction) of the view generated by the system are shown in Table 1. Most of the sensitivities and specificities are in the 70 to 80 percent range. The sensitivity of the system was exceptionally low in identifying laboratory tests for disease management (55 percent), which may be due to the QCIS' lack of knowledge in that area.

Taking a closer look at how often physicians agreed with each other on whether a test is relevant to a disease, we see that about 69.7 percent of the tests were

Table 1 ■

Sensitivity and Specificity of the Relevant Laboratory and Drug Information Identification

Data Type	Context	Gold Standard	Sensitivity (average)	Specificity (average)
Laboratory	Diagnosis	4/6 agreement	86%	75%
Laboratory	Management	4/6 agreement	55%	77%
Drug	Drug	Clinical review article	81%	80%

considered irrelevant and 1.6 percent of tests were considered relevant by all physicians. This indicates that the participating physicians agreed with each other completely more than 70 percent of the time. On the other hand, one or more physicians disagreed on whether a test was relevant about 30 percent of the time, which indicates the existence of interpersonal differences.

Information Overload Reduction

The concept-oriented views showed promising potential for information overload reduction. On average, each concept-oriented view (in this case, a disease-oriented view) contained only a fraction of all information about patients. The amount of information in the views, however, varied significantly from concept to concept as well as from patient to patient. The average number of laboratory tests and drugs per patient from 1996 to 1998 are 170.5 and 20.6, respectively, whereas the average number of tests and drugs for each patient in a view are 24.0 and 3.2, respectively. In other words, the average number of tests and drugs for each patient in a view are 14 to 16 percent of the total number of tests and drugs for each patient.

Physician Subjects

A total of 13 physicians (7 attending physicians, 2 residents, 3 interns, and 1 medical student) participated in this part of the evaluation and filled out the questionnaires. DOS and the Web are the two major platforms for clinical systems at NYPH. Six physicians had used only DOS-based clinical systems, one physician had only used the Web-based clinical systems, and six physicians had used both types of systems. Some of the recruited physicians were associated with the Department of Medical Informatics at Columbia University, so seven of the recruited physician subjects had experience in developing clinical systems. However, except for three subjects, no participants had ever seen the QCIS, and only one subject had briefly used the system before the evaluation.

User Comfort and Confidence

The comfort and confidence ranking were assigned on a 1 to 5 scale (on which 1 indicated very uncomfortable/not confident; 2, uncomfortable/not confident; 3, neutral; 4, comfortable/confident; and 5, very comfortable/confident). The physicians' opinions regarding the overall level of comfort and confidence in using the QCIS ranged from 2 to 5 (Table 2). Most users (11 of 13) found the system comfortable to use and were confident using it.

As shown in Tables 3 and 4, most physicians were comfortable and confident using either view, although some did rank one view higher than the

Table 2 ■

Users' Overall Ranking of Confidence and Comfort in Using the System

Comfortable	Confident				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	1	0
3	0	1	0	0	0
4	0	0	0	7	1
5	0	0	0	1	2

NOTE: Confidence and comfort were ranked on a scale from 1 to 5, as follows: 1, very uncomfortable/not confident; 2, uncomfortable/not confident; 3, neutral; 4, comfortable/confident; 5, very comfortable/confident.

Table 3 ■

Users' Ranking of Comfort Using Department-oriented View and Concept-oriented View

Concept-oriented	Department-oriented				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	0	1	0
3	0	0	0	0	0
4	1	0	0	3	3
5	0	1	0	2	1

NOTE: Comfort was ranked on a scale from 1 to 5, as follows: 1, very uncomfortable; 2, uncomfortable; 3, neutral; 4, comfortable; 5, very comfortable.

other. Because many cells in the contingency tables contained zeros (Tables 2 to 4), kappa statistics for assessing the agreement between the level of comfort and the level of confidence were not performed.

Information Retrieval Accuracy

Physician subjects answered more questions about patient cases correctly when using concept-oriented views (Table 5). We evaluated the difference in physicians' performance between the two different views using a one-sided Fisher's exact test, because of small counts in the contingency table. We found a statistically significant difference in the overall proportions of the correct answers by using the two views ($p = 0.04$). However, we did not detect any additional statistically significant difference in the stratified analyses by case or question, which may be because of the small sample size in each stratum (Table 5).

Six physicians made mistakes using department-oriented views, and two physicians made mistakes using the concept-oriented views. None of the physicians made mistakes using both views. The errors were not evenly distributed over all questions. For half (six) of the questions, perfect accuracy was achieved regardless of what view was used. For the remaining half (six) of the questions, mistakes were made; furthermore, subjects using concept-oriented views made fewer mistakes in five and tied with those using department-oriented views in one.

Information Retrieval Efficiency

Using a *t* test of mean of time to answer a question (a log transformation was applied to the data to reduce the skewness and variability), we did not find a statistically significant difference in time between the two views. Time spent on each question was highly variable, as indicated by the large standard deviations (Table 6). Such high variability could be related to the view, the case, or the subject. Because of the small sample size, we were not able to study the underlying reason, which will be subject to future investigation.

Table 4 ■

Users' Ranking of Confidence Using Department-oriented View and Concept-oriented View

Concept-oriented	Department-oriented				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	2	0
3	0	0	0	0	0
4	1	0	0	5	2
5	0	0	0	1	2

NOTE: Confidence was ranked on a scale from 1 to 5, as follows: 1, very not confident; 2, not confident; 3, neutral; 4, confident; 5, very confident.

Table 5 ■

Sample Sizes, Error Rates, and the *p* Value, Comparing the Proportions of Correct Answers (Pooled and Stratified by Case and by View)

Case Size	Sample	Error Rate (%)		*p* Value
		Department View	Concept View	
1	49	12	0	0.12
2	50	24	12	0.23
3	19	10	0	0.52
Pooled	118	17	3	0.04

Table 6 ■

Time Spent by Physicians to Answer a Question Using Concept-oriented vs. Department-oriented Views

	Time (sec)	
	Mean Time	SD
Group using department-oriented views	100.49	97.87
Group using concept-oriented views	112.81	112.58

Discussion

A multiple-view generation system with a focus on concept-oriented views was implemented and evaluated. This section discusses the significance, limitations, implications, and future research directions of this work.

Significance

Quality of Relevant Information Identification

The evaluation of quality of views generated by this system was the first study of the ability of the system to identify relevant clinical information and filter out irrelevant information for concept-oriented views. The evaluation showed that when identifying laboratory or medication information, the sensitivities and specificities of the system were generally moderately high (70 to 80 percent). Such performance obviously left large room for improvement.

Three major factors contributed to the difference between the QCIS and the gold standards in identifying relevant clinical information. First, the system does not have a complete or evenly distributed knowledge of medicine. Besides the medical knowledge already existing in the MED, the UMLS and DXplain were the major knowledge sources for the QCIS.[19,20] The knowledge we extracted from UMLS was based on the subjects of biomedical scientific papers published in the past decade, and the knowledge from DXplain focused mainly on the diagnosis of diseases.[10]

Second, relevance is a fuzzy concept. About 30 percent of the time, the six physicians recruited for creating the gold standard did not completely agree on the relevance of a laboratory test to a disease. Consequently, it was not surprising that the QCIS sometimes did not agree with the gold standard.

Third, the gold standards were not really "gold." For practical reasons, the physicians identified laboratory tests related to a disease using higher-level concepts such as "urine analysis." Often, only some of the specific tests grouped under such higher-level concepts are directly related to a disease. To avoid subjective judgments, the medication gold standard included all drugs mentioned in MICROMEDEX clinical review articles regardless of comments on their efficacy. The gold standards, therefore, are not perfect regarding the relevance of laboratory tests or medications.

Given the resulting sensitivity and specificity values, the accuracy of clinician information retrieval still improved using the concept-oriented views. This showed that automated selection of relevant clinical information is a very promising technique. The study also validated the knowledge-based approach of our system, including the knowledge sources and acquisition methods.

Information Overload

Information overload was the original problem that inspired the system development. Although previous statistics were unavailable, many papers recognize this problem and propose different types of views as a solution.[21–26] Our quantitative evaluation confirmed that concept-oriented views, which contain much less information than whole records, could be used to reduce the amount of information presented to users interested in particular topics. Especially noteworthy is the finding that reducing the amount of information presented did not compromise the accuracy of information retrieval.

Effects on Information Retrieval

Several simulation studies have evaluated the effects of different clinical data formats on the speed and accuracy of information retrieval.[27,28] They all showed that a fixed, specially designed, flow sheet or spreadsheet type of clinical data display (such as a microbiology display) improved the speed and accuracy of information retrieval.

More recently, Tange et al.[9,29] evaluated different views of medical narratives generated by a pilot system, particularly the effects of the views on the speed and completeness of information retrieval. The authors showed that viewing information by disease and organ system could speed up information retrieval. The views, however, had no effect on the completeness of the information retrieval.

In our study, we measured accuracy, cost (in time), comfort, and confidence of information retrieval. We also compared concept-oriented views with traditional, source-oriented views.

Accuracy

The accuracy of physician information retrieval improved with use of concept-oriented views, compared with use of traditional department-oriented views. This confirmed the general conclusion of previous studies that different data formats might improve information retrieval accuracy. Because the evaluation questions given to participating physicians were questions they would be asking during routine practice, such findings imply benefits in real clinical settings.

Cost

We did not find statistical difference in retrieval times between the two views. For different questions and various cases, one type of view was not always better than the other. For instance, for cases with very little information, the organization of information by problems might not confer much benefit.

This result should not be viewed as contradictory to findings of previous studies that showed improved speed in information retrieval with use of concept-oriented views. No previous study or system asked users to type in a concept of interest, and cases and questions used for testing in each study were different as well.

Comfort and Confidence

Physicians' opinions on how comfortable and confident they felt when using different types of views differed. This confirmed Newell and Simon's general propositions[30,31] that information processing is dependent on characteristics of the problem solvers and the tasks and that there are individual differences in problem solving. Such differences should be taken into consideration in clinical information system design. As much as standard data and knowledge representation are needed in computer systems, customized representations for users are also needed.

Components of the Evaluation Model

We have made an effort to evaluate both the functionality and the utility of a concept-oriented view- generation system. In other words, we studied whether the system functions as designed and whether using the system will be beneficial. Our results verified the importance to answer both questions.

Regarding functionality, the QCIS was able to identify relevant clinical information with moderately high sensitivities and specificities and reduce the amount information presented to users. Such results would have been much harder to interpret without the evaluation of the utility of the views. It would have been natural to assume that the moderate sensitivities and specificities of the concept-oriented views imply errors in clinician information retrieval or large reduction of information in the concept-oriented views translates into time saving.

Interestingly, when physicians used the QCIS to perform information retrieval tasks, their accuracy improved with the concept-oriented views while the time of information retrieval didn't change significantly, compared with the more standard department-oriented views. This demonstrates the importance of combining the functionality and utility studies.

On the other hand, we could have studied the utility of concept-oriented views with manually constructed views. Since our goal was to automate the view-generation process, and since computer-generated views could be very different from those created by medical experts, it is not ideal to study only manually constructed views. As we studied the utility (information retrieval accuracy and time) of computer-generated views, the results needed to be interpreted in the context of the system functionality (sensitivity, specificity, and amount of information reduction). Had the functionalities of the system been different, the utilities would probably have been different as well.

Limitations

Concept-oriented views can center on all kinds of concepts, including diagnostic strategies and therapeutic goals. Some concept-oriented views require the identification of case-specific relationships, such as those between a laboratory test and a symptom in the context of a patient's medical history. As discussed in a previous paper,[3] those relationships are best established by clinicians, and the QCIS was not designed to establish those case-specific relationships.

The QCIS is capable of generating concept-oriented views for nine classes of concepts by discarding those concepts that are not likely to be related to them. The evaluation reported in this paper focused on views of only one class of concepts—the disease concepts.

Recruiting physician subjects for evaluation can be difficult because of the already high demands on physicians' time. Nevertheless, because physicians are the potential users, it was necessary to involve them in the study. For the evaluation, the number of subjects was limited. The main incentive for the physicians who did participate in the evaluation of this system was to support clinical information system research. So it was not surprising that the majority of participants (7 of 13) had some connection to the medical informatics department.

Since most participants were unfamiliar with the system being evaluated, training was needed. However, five minutes of training could be inadequate. For example, the subjects asked some questions about what was clickable, and a number of users commented that they finally started to get the idea of concept-oriented views only after they had finished the tasks.

Comfort and confidence are subjective measures that can be influenced by a number of factors other than the system. For example, participants may report feeling more comfortable and confident than they really are because of their sympathy toward the system developers or the general research goals. On the other hand, some subjects may feel that they are being evaluated along with the system and thus may become more critical of the system.

Accuracy and cost (in time) are objective measurements that provide valuable data for analysis. In certain cases, conclusions could be drawn—for example, that concept-oriented views led to more accurate retrievals. In other cases, however, our analysis identified the need for further research, showing, for example, that more research is needed to determine which type of view could save time in the answering of which kind of question.

Implications and Future Research

The evaluation indicated that it is possible to generate concept-oriented views using automated relevant information-identification methods. Regarding the quality of relevant information identification, the performance of the system is not comparable with the performance of physicians in all areas. The use of such computer-generated views, however, still proved to be beneficial regarding certain aspects of information retrieval. This shows that automated concept-oriented view generation is a direction worth further exploration.

Although we focused on the concept-oriented views, we did implement a system that provides multiple views. It is our belief that different types of views meet the needs of various clinical tasks and users. Considering the complex nature of medical practice and the differences among individual practitioners, an ideal system would offer multiple types of views and hybrid views.

The system evaluation would be improved with more subjects and more cases. Subjects could be given more complicated clinical tasks, such as preparing discharge summaries and making diagnoses, instead of straightforward information retrieval questions. In such studies, information retrieval may be extended to areas other than laboratory tests and drugs in the system.

Conclusion

A formal evaluation has been conducted on a knowledge-based, concept-oriented view-generation system for clinical data. The automated methods employed by the system to determine the relevance of patient information were shown to be promising. The evaluation also verified the effective use of concept-oriented views to reduce information overload and improve the accuracy of information retrieval.

The authors thank the physician subjects who participated in this evaluation study and Dr. Gai Elhanan, Dr. Osama Alswailem, and Dr. Shabina M. Ahmad for serving as physician experts. They also thank Dr. George Hripcsak, Dr. Soumitra Sengupta, Dr. Gilad Kuperman, and Dr. Suzanne Bakken for reviewing the manuscript.

References ■

1. Rose JS. Medicine and the Information Age. Tampa, Fla.: American College of Physician Executives, 1998.
2. Dore L, Lavril M, Jean FC, Degoulet P. An object-oriented computer-based patient record reference model. Proc Annu Symp Comput Appl Med Care. 1995:377–81.
3. Zeng Q, Cimino JJ. A knowledge-based, concept-oriented view generation system for clinical data. J Biomed Inform. 2001; 34(2):112–28.
4. Weed LL. Medical records that guide and teach. N Engl J Med. 1968;278(11):593–600.
5. Suermondt HJ, Tang PC, Strong PC, Young CY, Annevelink J. Automated identification of relevant patient information in a physician's workstation. Proc Annu Symp Comput Appl Med Care. 1993:229–32.
6. Tang PC, Annevelink J, Fafchamps D, Stanton WM, Young CY. Physicians' workstations: integrated information management for clinicians. Proc Annu Symp Comput Appl Med Care. 1991:569–73.
7. Bodenreider O, Nelson SJ, Hole WT, Chang HF. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. Proc AMIA Annu Symp. 1998:815–9.
8. Balas EA, Pryor TA, Hebertson RM, Haug PJ, Twede M. A computer method for visual presentation and programmed evaluation of labor. Obstet Gynecol. 1991;78(3 pt 1):419–23.
9. Tange HJ, Schouten HC, Kester AD, Hasman A. The granularity of medical narratives and its effect on the speed and completeness of information retrieval. J Am Med Inform Assoc. 1998;5(6):571–82.
10. Zeng Q, Cimino JJ. Automated knowledge extraction from the UMLS. Proc AMIA Annu Symp 1998:568–72.
11. Zeng Q, Cimino JJ. Providing multiple views to meet physician information needs. Proceedings of the 33rd Hawaii International Conference on System Sciences (Jan 4–7, 2000; Maui, Hawaii) [CD-ROM]. Washington, DC: IEEE Computer Society, 2000.
12. Zeng Q, Cimino JJ. Evaluation of a system to identify relevant patient information and its impact on clinical information retrieval. Proc AMIA Annu Symp. 1999:642–6.
13. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Inform Assoc. 1994;1(1):35–50.
14. Zou KH, Zhou XH. Receiver operating characteristic (ROC) analysis. Statistics in Epidemiology Report. Newsletter of the Statistics in Epidemiology Section of the American Statistical Association. 2001:1–9.
15. Drug information. MICROMEDEX Web site. Available at: http://www.mdx.com/products/healthcare/druginfo/. Accessed 1999.
16. Fauci AS, Braunwald E, Isselbacher KJ, Martin JB (eds). Harrison's Principles of Internal Medicine. 15th ed. New York: McGraw-Hill, 2001.
17. Kushniruk AW, Patel VL, Cimino JJ. Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces. Proc AMIA Annu Fall Symp. 1997:218–22.
18. Cimino JJ, Patel VL, Kushniruk AW. Studying the human-computer–terminology interface. J Am Med Inform Assoc. 2001; 8(2):163–73.
19. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc. 1998;5(1):1–11.
20. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. Dxplain: an evolving diagnostic decision-support system. JAMA. 1987;258(1): 67–74.
21. Tang PC, Annevelink J, Suermondt HJ, Young CY. Semantic integration of information in a physician's workstation. Int J Biomed Comput. 1994;35(1):47–60.
22. Powsner SM, Tufte ER. Graphical summary of patient status. Lancet. 1994;344(8919):386–9.

23. Powsner SM, Tufte ER. Summarizing clinical psychiatric data. Psychiatr Serv. 1997;48(11):1458–61.
24. Hayes GM. Medical records: past, present, and future. Proc AMIA Annu Fall Symp. 1996:454–8.
25. Wyatt JC. Clinical data systems, part 1: data and medical records. Lancet. 1994;344(8936):1543–7.
26. Miller PL, Blumenfrucht SJ, Rose JR, et al. Hydra: a knowledge acquisition tool for expert systems that critique medical workup. Med Decis Making. 1987;7(1):12–21.
27. Elson RB, Connelly DP. The impact of anticipatory patient data displays on physician decision making: a pilot study. Proc AMIA Annu Fall Symp. 1997:233–7.
28. Willard KE, Connelly DP, Johnson JR. Radical improvements in the display of clinical microbiology results: a Web-based clinical information system. Am J Med. 1996;101(5):541–9. Erratum: Am J Med. 1997;103(1):86.
29. Tange HJ, Dreessen VA, Hasman A, Donkers HH. An experimental electronic medical-record system with multiple views on medical narratives. Comput Methods Programs Biomed. 1997;54(3):157–72.
30. Newell A, Shaw JC, Simon HA. Elements of a theory of human problem solving. Psychol Rev. 1958;65:151–66.
31. Newell A, Simon HA. Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall, 1972.

Appendix

Evaluation Questionnaire for Patient Cases

Case 1

Mr. Smith is a 74-year-old male patient with a history of congestive heart failure.

1. The patient was discharged on digoxin. What is his latest digoxin level test result?
2. What is his latest serum potassium value?
3. When was it the first time the patient's heart was shown enlarged on a chest x-ray?
4. What anti–heart failure drugs were given to this patient during his last hospitalization? (Jul 5–9, 1998)
5. Did this patient have pulmonary edema during his last hospitalization (Jul 5–9, 1998)?

Case 2

Mrs. Jones is a 78-year-old female patient with a history of diabetes.

1. How many times has she been hospitalized for diabetes?
2. What was the first glycosylated Hb level for the last hospitalization (Jun 3–12, 1998)?
3. What was the highest random blood sugar she had during her last hospitalization? (6/03/98 - 6/12/98)
4. And when did it occur?
5. What are the patient's latest lipid levels?

    LDL ________

    HDL ________

    Triglyceride ________

Case 3

Mrs. Green is a 67-year-old woman with a history of pancreatitis.

There are some drugs for which there is a definite association with acute pancreatitis; like azathioprine, sulfonamides, thiazide diuretics, furosemide, tetracycline, valproic acid, pentamidine.

1. Was this patient ever on any of these medications ?

It is important to identify patients with acute pancreatitis who have an increased risk of dying. Ranson and Imrie have used multiple prognostic criteria and have shown that mortality rate increases when three or more risk factors are identifiable either at the time of admission to the hospital or during the initial 48 hr of hospitalization

The RANSON/IMRIE CRITERIA are:

At admission or diagnosis:
Leukocytosis >16,000/mL.
Hyperglycemia >11 mmol/L (>200 mg/dL).
Serum LDH >400 IU/L.
Serum AST >250 IU/L.

During initial 48 hr:
Fall in hematocrit by >10 percent.
Hypoxemia (PO2 <60 mmHg).
Increase in BUN by >1.8 mmol/L (>5 mg/dL).
Hypoalbuminemia [albumin level <32 g/L (<3.2 g/dL)].

2. Please check whether the criteria were present at this patient's last admission (Apr 7, 1998) and initial 48 hr (Apr 7–8, 1998).