Journal of Biomedical Informatics **34,** 112–128 (2001)
doi:10.1006/jbin.2001.1013, available online at http://www.idealibrary.com on IDEAL®

A Knowledge-Based, Concept-Oriented View Generation System for Clinical Data

Qing Zeng*,[1] and James J. Cimino†

**Decision Systems Group, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts 02115; and †Department of Medical Informatics, Columbia University, New York, New York 10032*

Received February 15, 2001; published online July 16, 2001

Information overload is a well-known problem for clinicians who must review large amounts of data in patient records. Concept-oriented views, which organize patient data around clinical concepts such as diagnostic strategies and therapeutic goals, may offer a solution to the problem of information overload. However, although concept-oriented views are desirable, they are difficult to create and maintain. We have developed a general-purpose, knowledge-based approach to the generation of concept-oriented views and have developed a system to test our approach. The system creates concept-oriented views through automated identification of relevant patient data. The knowledge in the system is represented by both a semantic network and rules. The key relevant data identification function is accomplished by a rule-based traversal of the semantic network. This paper focuses on the design and implementation of the system; an evaluation of the system is reported separately. © 2001 Academic Press

Key Words: concept-oriented view; information overload; knowledge-based system; relevant patient data identification.

INTRODUCTION

Information overload is a well-known problem for clinicians, who often must read bulky paper charts. As electronic medical records (EMRs) gradually replace paper charts, many problems are being solved. The amount of data available for clinicians to review, however, has not been reduced. In fact, as technology improves, it helps capture even more patient data. Because clinicians may not have enough time to review and process all of the data, their ability to perform clinical tasks such as information retrieval and decision making may be hampered.

Depending on their goals, clinicians may only be interested in certain subsets of the data, which we refer to as *views*. It would therefore be useful to reduce the quantity of information for clinicians to process by generating such views automatically.

The most common view of the EMR is the *source-oriented view*, which organizes the data based on where they were collected. The other common view is the *time-oriented view*, which primarily uses time to organize data. Other views can be envisioned that center around clinical concepts such as diagnostic strategies, therapeutic goals, which are referred to as *concept-oriented views*. The most famous example of a concept-oriented view is the *problem-oriented medical record* [1].

Generating and maintaining concept-oriented views of patient data are not trivial. In source and temporal views, place and time of data observations are used to organize data. In concept-oriented views, data are organized according

[1]To whom correspondence should be addressed at Decision System Group, Thorn Building, No. 309, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115. E-mail: Qzeng@dsg.harvard.edu.

1532-0464/01 $35.00

to the underlying medical concepts and relationships between concepts. Although place and time of data observations are generally recorded, relationships between a datum and concepts are not. Such data-concept links need to be created, obtained, or inferred to generate the concept-oriented views.

When view generation relies on manually maintained links, it is extremely difficult to have various concept-oriented views. Limited research has been done to establish relationships between patient data and medical concepts using automated methods [2]. Since automated methods demand extensive medical knowledge in a computer-understandable form, how to acquire, model, store, maintain, and utilize such knowledge are still issues for investigation.

There have been several simulation studies on the impact of different clinical data formats on the speed and accuracy of information retrieval [3–5]. Although the value of concept-oriented views has been shown through evaluations [5–7], no evaluation has ever been carried out on the automated view generation methods in conjunction with medical knowledge. Also, few evaluation studies of computer-generated, concept-oriented views of any sort have appeared in the published literature [6]. The quality and impact of concept-oriented views generated by automated methods still needs to be evaluated.

To study computer-generated concept-oriented views, a general-purpose, concept-oriented view generation system was developed and evaluated. We refer to the system as the querying clinical information system (QCIS). Given a user-selected concept of interest, QCIS can use medical knowledge to generate a view containing information relevant to the concept from a patient record. In addition to concept-oriented views, the system is also capable of handling source- and time-oriented views. This paper will describe the knowledge-based approach that was employed in the system design and implementation.

BACKGROUND

Information Overload

Modern medical records are comprehensive documents that contain information such as patient medical history, hospital discharge summaries, and electrocardiograms. The information is collected for clinical, legal, financial, and administrative purposes from many sources such as primary care physicians, pathologists, and social workers [8]. Although it is necessary to keep extensive medical records, it is not always easy for physicians to retrieve information from them [9, 10].

The amount of available data can overwhelm clinicians, making it hard for them to identify the desired information [9–11]. In a survey conducted by Tange, retrieval of specific tests was considered by physicians to be a weak aspect of paper-based medical records, as significant as poor legibility [12].

EMRs do not have many of the problems associated with paper-based records. For example, handwriting is not a problem with most EMRs. (Electronic records composed of images of scanned paper documents are the exception.) EMRs, however, still contain a great amount of information and are not automatically more comprehensible because they are electronic [10]. A study at New York Presbyterian Hospital (NYPH) conducted and reported by Elhanan showed that many patients have over 1000 laboratory tests per hospitalization per person [13]. Although the test results are stored in a database and are easily accessible, browsing through them or searching for a specific test can be tedious.

The problem of information overload can impair a clinician's ability to make medical decisions. When clinicians are overwhelmed by the amount of information in medical records, they actually cannot get the information they need and end up with "informaciation" (lack of crucial information to take care of patients) [9]. Such "informaciation" makes decision making difficult because medical decisions are best made when based on past and present patient information [14, 15].

Information Processing Theory

Our intention in studying patient data views that can organize clinical data based on their relation with a subject (problem or task) to reduce information overload is consistent with the propositions of information processing (or human problem solving) theory. The origin of information processing theory was the work by Miller and Newell [15, 16]. Newell and Simon provided five general propositions of a theory of human problem solving [17]:

- Human problem solving is information processing.
- Information processing is dependent on the characteristics of the problem solver and the task.
- There are individual differences in problem solving.
- Different tasks require different in0formation processing.
- The nature of the task and intelligence of the problem solver determine problem-solving behavior.

Since information processing is dependent on the characteristics of the problem solver and the task, it is reasonable to theorize that different information organization and presentation schemes (views) are required to best meet the needs of different users and tasks.

Views

Although clinicians do need access to a patient's entire record, they often seek answers to specific questions and only wish to browse certain subsets of data (views). The problem of information overload can be eased by providing views specific to the tasks that motivated the clinicians to obtain patient information [5]. For example, specially designed summary views of patient data have been shown to reduce time spent searching and to improve the quality of medical decisions [5 7, 7 18–20].

Views can be classified by how they are organized. Dore pointed out four major types of view: the time-oriented view, which is organized by occurrence time; the encounter-oriented view, which is organized by observation time; the source-oriented view, which is organized by data origin (radiology, laboratory, cardiology, etc.); and the concept-oriented view which is organized by medical concepts (disease, findings, etc.) [21]. Since occurrence- and observation-oriented views are organized by different (but similar) times, we will consider three major types of views: source-oriented, time-oriented, and concept-oriented. Figure 1 shows how those views are projected over an EMR.

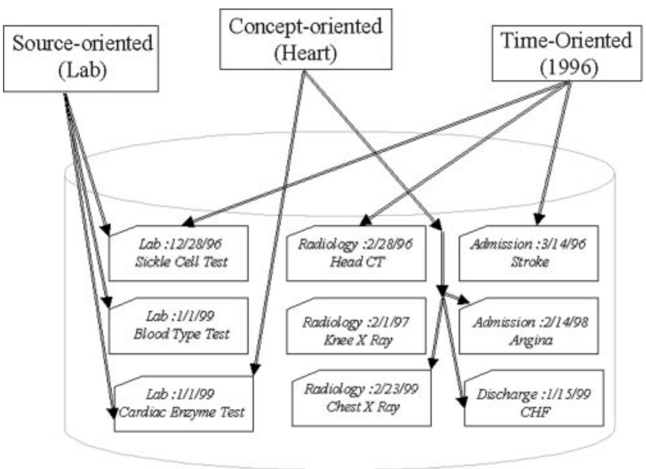

FIG. 1. Source-oriented, time-oriented, and concept-oriented views are organized respectively by the source, time, and meaning and potential use of clinical data.

Source-oriented views are probably the most commonly used views in current EMRs. A typical example is the ancillary department view, which organizes information by its department of origin, such as laboratory, radiology, pharmacy, and cardiology. Because there is normally a very limited number of data origins and they are routinely recorded along with the data, generating source-oriented views is relatively straightforward.

Time-oriented views are common for paper records. Various longitudinal summary EMR displays are also good examples of this type of view. Although data collection or event occurrence times are also routinely entered into data repositories, there can be more variance in time-oriented than source-oriented views; depending on the frequency of a type of event, the appropriate unit of time for data organization may differ. For instance, the exact hour that a mammography test was administrated may not be crucial, but the hour and minute of a blood sugar test could be clinically relevant. Research has been conducted in designing and generating time-oriented views [10, 22, 23]

Very few medical record systems provide concept-oriented views to users. Although concept-oriented views are still rarely available, they have been desired by clinicians. Past pilot studies have validated the potential value of concept-oriented views [6]. Because there may be tens of thousands of concepts by which a view could be oriented, concept-oriented views have the most variety. Concept-oriented views are also hardest to generate because the relationship between patient data and medical concepts is not always explicit. The most famous generic example of a concept-oriented view is the problem-oriented medical record (POMR) proposed by Weed [1]. Despite some criticisms [24], problem-oriented views have generally been accepted as a major enhancement to the format of the medical record [5].

Generating Concept-Oriented Views

In the past, most concept-oriented views were generated manually. For instance, PROMIS, a POMR system developed by Weed, required physicians to manually link problems with patient data and did not gain as much acceptance as the idea of POMR [25–27].

Manually creating and maintaining concept-oriented views can be prohibitively tedious. Take the problem-oriented view again as an example: having a physician link each piece of patient data to one or more problems would require a significant amount of work. Moreover, different health providers of a patient may disagree on what the problems are, and on the problem-data links. Some physicians view symptoms, such as headache, as patient problems, whereas others classify problems as diagnoses or hypotheses. Because diagnoses may evolve over time based on new evidence, maintaining consistent problem-oriented views can be complicated.

Just as physicians may never be able to anticipate and link a test result to all possible problems, especially a later diagnosis, the computer has its own limits: it may never be able to know and explain why each test was ordered for a patient. Some connections between patient data and patient problems are case-specific, so only the physicians who are taking care of the patient can establish them. There are, however, non-case-specific connections that can be inferred from medical knowledge, and such connections can be generated by automated systems.

Automating the generation of concept-oriented views is challenging because of the knowledge required to link patient data to medical concepts. For instance, to know that a chest X-ray may reveal information about the heart, the system first needs to know that the heart is located in the chest and that the chest is the target body part of a chest X-ray. As trivial as the example may seem, there are few formalized sources for such knowledge.

In addition to the lack of sources, the quantity and variety of the required knowledge are also obstacles for knowledge representation, acquisition, and utilization. There are hundreds of thousands of diseases, findings, drugs, tests, and procedures. The number of possible relationships among them is many times the number of concepts. Furthermore, not all relationships bear the same semantics, and such differences could be clinically relevant.

Few systems have attempted automated concept-oriented view generation; the physician workstation system developed by Tang's group for outpatient care is the most significant example [28, 29]. A prototype of the system was implemented using a methodology that combined a patient-specific physiological model with functions to determine relevance of patient information for clinical display generation and critical event monitoring. For example, when hypertension was the problem of focus, the display contained relevant disease parameters such as blood pressure and heart rate.

The multiple-view system for medical narratives developed by Tange was another pilot system that intended to provide multiple views of the EMR to users [30]. This system contained a search structure that supported several views of the medical data. The information in the database, however, was prelabeled and the relevant information identification issue was not directly addressed.

Some laboratory displays are also concept-oriented; for example, the cardiac laboratory display focuses on laboratory tests related to the heart. Though normally predefined (as defined by human experts), the contents of a laboratory display can be partially or fully defined in a dynamic manner (as defined by systems drawing on medical knowledge bases) [31–33]. This can be regarded as a form of automated concept-oriented view generation as well.

There have been very few evaluations of automated view generation systems. Tange performed an evaluation of the views generated by his system [6]. However, the most challenging issue—automated identification of relevant information—was not fully addressed by the evaluation. It therefore remains to be proven that views can be generated through automated relevant information identification and that the views generated using such methods can benefit clinicians.

System Design and Implementation

We developed QCIS to be capable of handling all three major types of views: source-oriented, time-oriented, and concept-oriented. Upon selecting a specific patient, users have the freedom to view the clinical data using one of the three types. This design was based on our hypothesis that different views complement each other. The system's ability to generate various views also made it possible to study the differences among them. The emphasis of QCIS is to provide concept-oriented views that are generated by computer identification of data that are relevant to a user-specified concept of interest (Fig. 2).

The content of the concept-oriented views was limited to coded clinical data (including those extracted from text reports through natural language processing [34]) because the

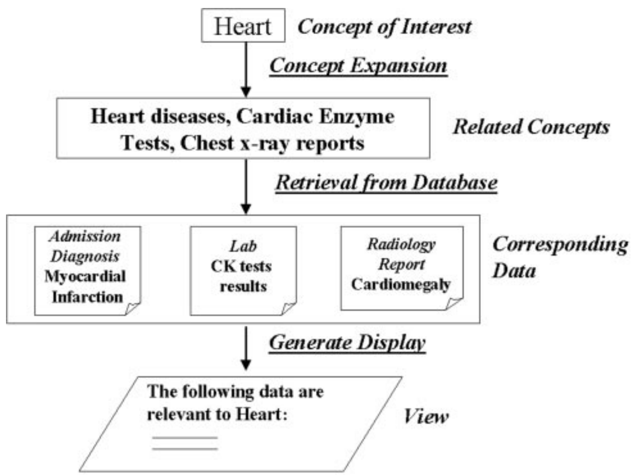

FIG. 2. Concept-oriented view generation process.

concept-oriented views were organized by the content and meaning of data, and computer systems can only "understand" coded data. The source of the clinical data is the central clinical data repository at NYPH [35]. The available coded data from the NYPH include laboratory tests, inpatient medication orders, admission/discharge diagnoses, and radiology procedures.

QCIS is designed to be generalizable: it generates concept-oriented views regardless of the unique characteristics of certain classes of concepts. The system has knowledge (in the form of relationships and rules) about the unique attributes of different concepts and how to use these attributes to generate views. We will describe the system's knowledge representation, acquisition, architecture and implementation.

Knowledge Representation

Since we intend for our system to generate concept-oriented views through an automated process instead of manual linking patient data and medical concepts, it needs domain knowledge to identify relevant concepts and retrieve corresponding data. We chose an existing semantic network (the Medical Entities Dictionary (MED) [36]) and rules to represent the knowledge. For the semantic network, we used several existing semantic relationships and created additional relationships help in identifying relevant concepts.

The knowledge needed for identification of relevant concept consists of the relationships between concepts and the logic about how the relationships should be used to make valid connections. For instance, to connect chest with heart disease requires knowledge of the relationships between heart diseases and heart, and heart and chest. However, it is also possible to link chest with renal disease since kidney and chest are both anatomical entities. We use logic to establish which connections are reasonable for our purpose and which are not.

For this project, examples of typical concepts were used to examine what relationships are involved in identifying relevant concepts. For each example concept, relevant concepts were manually identified and connected by a domain expert (Fig. 3). To use the existing knowledge in the MED, efforts were made to use existing MED relationships. When appropriate relationships could not be found in the MED, the author proposed new relationships that were reviewed and approved by a domain expert.

The kinds of relationships needed to identify related concepts are diverse. Some relationships are definitional and thus permanent; for example, heart is a part of the cardiovascular system. Other relationships contain a certain degree of uncertainty and are subject to change, for example, the medications used to treat a disease. Some relationships can be described qualitatively or quantitatively; for instance, the relationship between diabetes and elevated blood sugar value can be described qualitatively. Some relationships, such as the relationship between cardiovascular disease and observations for suspected cardiovascular disease, may not easily fit into a qualitative or quantitative model. Certain relationships are local; for example, latex fixation test is one of the NYPH "gold top" chemistry tests. Other relationships, such as "chest is the o bservation site of chest X-ray" are universal.

A semantic network is an ideal representation scheme for these relationships. Although some of the relationships may be best represented by other schemes, a semantic network is one scheme that has the capability to accommodate different types of relationships. Another reason for choosing a semantic network is that the MED is a semantic network and its framework and knowledge can be reused.

Fourteen kinds of new semantic relationships along with 14 kinds of extant relationships were identified for relevant concept identification. (Table 1). These relationships are in pairs because in the MED, every relationship has a reciprocal relationship.

As mentioned previously, logic about how the relationships should be used to make valid connections is another kind of knowledge that needed to be represented. We used rules to represent this logic. Rules were also used to represent the administrative knowledge for patient data retrieval such as which query module to call for a class of concepts. All the rules were grouped into modules according to the purpose and application domain of the rules. They are described in detail in the section on system architecture.

Knowledge Acquisition

Knowledge acquisition is important for our approach because a large amount of knowledge is needed to generate concept-oriented views. For example, knowledge about relationships between laboratory tests and diseases is necessary to generate disease-oriented views. A detailed description of the knowledge acquisition work for this system was published in an earlier paper [37].

For this project, three sources were employed: domain experts, published literature, and computer-based resources. Because of the amount of knowledge required, computer-based resources where knowledge may be acquired using automated methods were the preferred sources. For instance, there are thousands of diseases and laboratory tests and

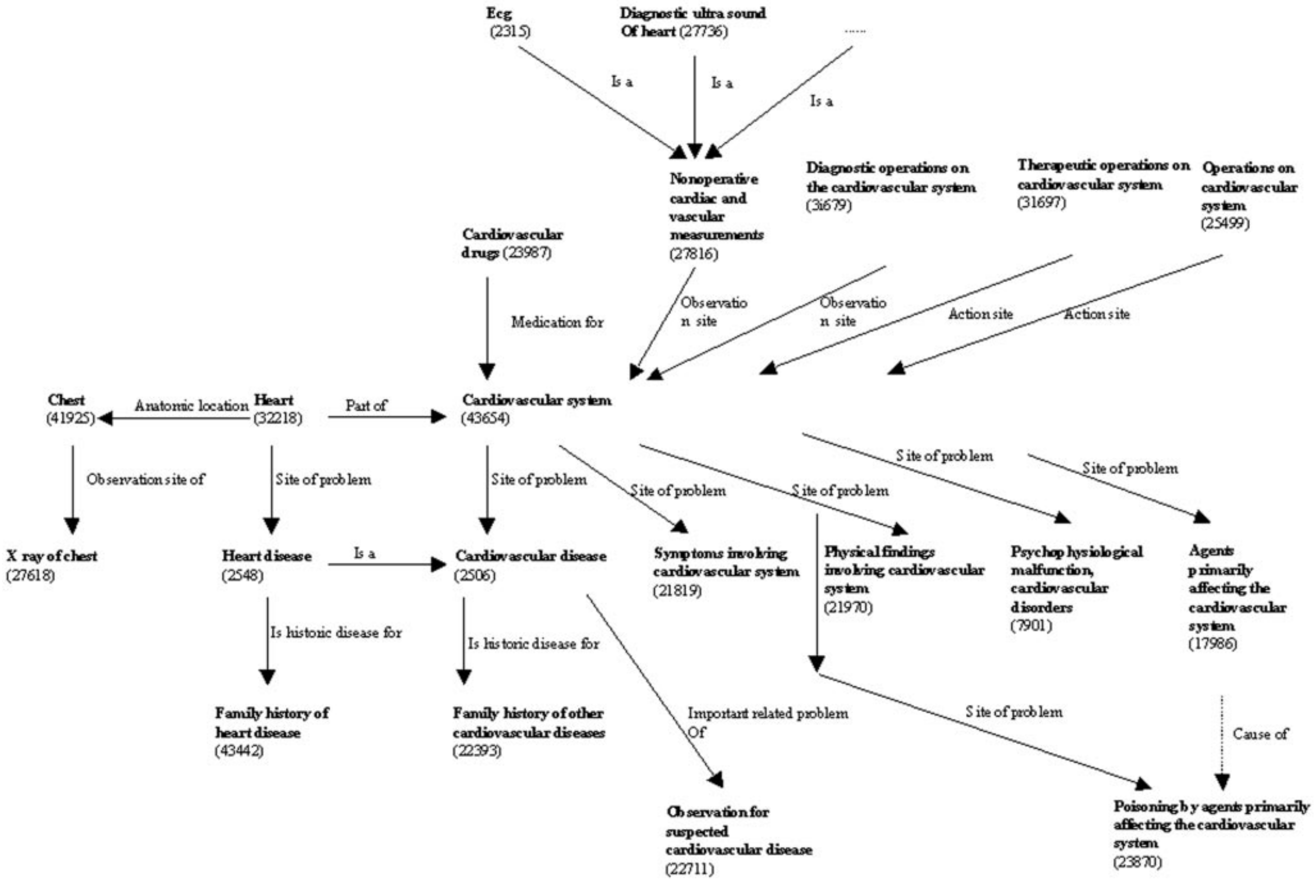

FIG. 3. This chart demonstrates how the MED concept "heart" could be linked to some other MED concepts via semantic relationships.

there could be tens to hundreds of thousands of relationships between diseases and laboratory tests. Some knowledge, however, was not available in computer understandable format; such knowledge was available only from the literature or domain experts. For example, knowledge about physiological relationships between organs was obtained from textbooks because no appropriate computer-based source was found.

Knowledge extraction from computer-based sources. UMLS [38] and DXplain [39] are the two computer-based sources that were explored because they could offer knowledge about disease–chemical relationships (e.g., drug treatment or contradiction). From UMLS, we extracted the disease–chemical relationships from the co-occurrence (MRCOC) data on disease and chemical concepts. From DXplain, the disease–chemical relationships were inferred following the links from laboratory findings in the disease descriptions to laboratory tests and then to its analytes. The detail of the methods was reported in a prior paper [37].

From UMLS, 389,655 relationships were extracted. Through these relationships, 3388 diseases and 5683 chemicals were linked. From the DXplain, 6992 pairs of disease–chemical relationships were extracted. We added all these relationships to the MED.

There is no gold standard for the disease–chemical relationship knowledge acquired from UMLS and DXplain because relevance can only be defined in context. Knowledge acquired from the UMLS MRCOC was tested against DXplain knowledge and literature knowledge to get a general estimation of sensitivity. MRCOC's sensitivity regarding disease–drug chemical relationships was very good (93%), but sensitivity regarding disease–laboratory chemical relationships was less satisfactory (68%). The ultimate evaluation, however, is of the performance of the system when the knowledge is put to use.

TABLE 1
Extant and New Semantic Relationships Identified for the Relevant Concept Identification

Extant semantic relationships	
Descendant of	Descendants
Has parts	Part of
Specimen	Specimen of
Substance measured	Measured by
Has problem site	Site of problem
Pharmaceutic component	Pharmaceutic component of
Substance sampled	Sampled by
New semantic relationships	
Observation site	Observation site of
Action site	Action site of
Has anatomic location	Anatomic location of
Historic disease	Is historic disease for
Has important related problem	Important related problem for
Has related pharmaceutical chemical	Related pharmaceutical chemical for
Has related measurable entity	Related measurable entity for

Knowledge extraction from literature. Some knowledge regarding anatomical relationships was manually acquired from textbooks because no computer-based source was available. The 37 extracted relationships regarding anatomic systems were added to the MED.

Knowledge extraction from domain experts. Knowledge was acquired from domain experts to create rules for concept expansion and query construction. For administrative knowledge needed to construct queries, the domain experts were the database designers and administrators. For expansion rules, a physician served as the domain expert. The knowledge acquisition generally started with discussion with the domain experts. Then the knowledge was formalized by the system developer and presented to the experts for modification and approval. This process was sometimes repeated several times before we obtained the approval of experts.

System Description

The concept-oriented view generation process in our system consists of four parts: *concept selection* (selecting the concept of interest), *concept expansion* (finding concepts related to the concept of interest), *data retrieval* (obtaining information from the patient record that is coded with the expanded set of concepts), and *display generation* (producing a human-readable set of results). Take the concept **heart** as an example (Fig. 2). First, a user could start the process by inputting the string "heart" and QCIS would map it to concepts such as **heart, heart disease,** and **heart failure**. Then the user would select the *concept* **heart** which would become the concept of interest for a view. In the second step, the system would traverse the MED's semantic network to find the concepts that are related to **heart**. This would result in a list of disease terms (**heart diseases**), laboratory tests (including **cardiac enzyme tests**), and radiography reports (including **chest X-rays**). In the third step, the necessary queries would be performed by the system to obtain all occurrences of these concepts in the EMR, such as an admission diagnosis of myocardial infarction, some creatine kinase tests results, and a radiology report indicating cardiomegaly. In the final step, these data would be organized into a display for presentation to the user. The system's architecture is shown in Fig. 4.

Concept selection. There are tens of thousands of concepts by which a view could be oriented, so QCIS relies on user input to locate the concept that the user might be interested in. After a user inputs a term into QCIS, the system maps the term to concepts in the MED. The lexical lookup functions are provided by the MED. QCIS first tries to find exact matches and when exact matches cannot be found, the system attempts to find partial matches. The concepts that matched the user term are displayed and users can select one matching concept which becomes the concept of interest for a view.

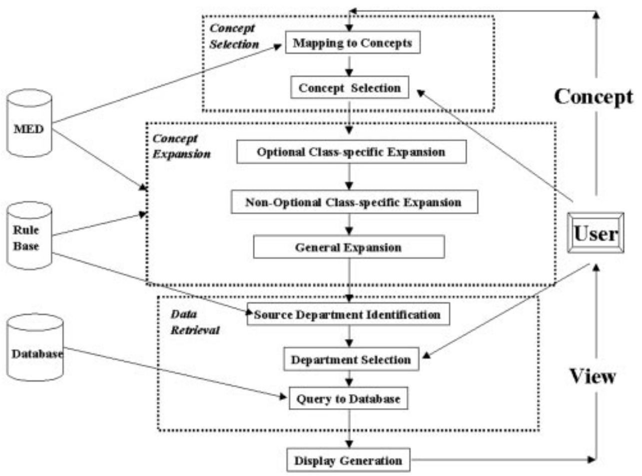

FIG. 4. This chart shows the view generation process for a user-selected concept. The process consists of four parts: concept selection, concept expansion, data retrieval, and display generation. The system also interacts with the MED and a rule base for knowledge and the NYPH database for clinical data.

Concept expansion. Given a concept of interest, QCIS identifies the related concepts. As discussed in the knowledge representation section, two types of knowledge are involved in this concept expansion process: (1) relationships between concepts; and (2) logic, which guides the system to establish meaningful connections between concepts based on the relationships. In QCIS, the relationships are stored in a semantic network, and the logic is represented by expansion rules.

In addition to previously existing relationship knowledge in the MED, new relationships were acquired and added. The following is an example of some relationships:

UNSPECIFIED CIRCULATORY SYSTEM DISORDER
—has related pharmaceutical chemical:
ALPROSTADIL
ASPIRIN
CAPTOPRIL
DIPYRIDAMOLE
. . . .

Expansion rules were constructed based on expert knowledge. The following is an example of a few rules:

If (the concept is an anatomic concept and has part X), then {*add X to the set of related concepts*}

If (the concept is an anatomic concept and the site of problem for Y), then {*add Y to the set of related concepts*}

If (the concept is a problem and has related pharmaceutical chemical Z), then {*add Z to the set of related concepts*}

. . .

As illustrated in the example, some rules only apply to a certain class of concepts. (In QCIS, expansion rules were only defined for high-level concept classes such as anatomic entities and diseases instead of individual concepts, which limited the number of rules needed.) These rules are grouped into modules by class and referred to as *class-specific expansion rules*. Other rules apply to all concepts and are referred to as *general expansion rules*. "If concept A is related to concept B, then all the descendants of concept A are related to concept B" is one such rule. The distinction between general and class-specific rules was made to ensure their proper application of rules.

Not all concepts are directly searchable in the patient database. For example, **CK** is a chemical concept and would not be found in the coded patient data in our clinical database, whereas **serum creatine kinase mm measurement** is a concept that is used to code laboratory results. Some expansion rules were specifically written to look for searchable concepts corresponding to unsearchable concepts. Because these rules must be applied for data retrieval, they are called *non-optional rules*. Consequently, the remaining rules are referred to as *optional rules*.

During concept expansion, the concept of interest first goes through the class-specific *optional* expansion, which results in a set of related concepts, including the concept of interest itself. Then the class-specific non-*optional* expansion is performed on each of the concepts in this set, generating a new set of related concepts. In the last step, each related concept undergoes a general expansion, which produces a final set of related concepts.

The application of general rules is not elective in QCIS; therefore, distinctions were not made between *optional* and non-*optional* general expansion rules. For class-specific expansions, a classification process determines if any class-specific rules apply. If no rules apply to each element in the set of related concepts, no expansion would occur and the concept(s) would be passed on to the next step.

Data retrieval. Concept expansion provides an answer to the question "What to retrieve?" "Where and how to retrieve?" is the next question. Identifying possible source departments of a concept is the first step because it enables QCIS to know where to search in the database. Queries are then constructed and executed for concepts in different departments.

To determine the potential source departments of a searchable concept requires both general classification knowledge in the MED and knowledge of the local EMR system. For instance, based on the knowledge that heart disease is an ICD-9 disease, and admission/discharge diagnoses are coded in ICD-9, it is reasonable to conclude that the department where admission/discharge diagnoses are stored is a place to search for heart disease. In this instance, in addition to the knowledge of classification and of the EMR system, QCIS also employs some simple logic. The knowledge of the EMR system and the logic are represented in the rules. In the implementation of this system, both these rules and the expansion rules are stored in a rule base. After applying the EMR-specific rules to determine the possible source departments, a list of departments is presented to users so that they may select the appropriate one for data retrieval.

With both paper charts and EMRs, browsing has always been better supported than searching. Searching for specific concepts is a low-priority function for most clinical databases, except for databases primarily designed for research purposes. Queries to the NYPH clinical database for specific concepts tend to be somewhat more complex and less efficient than common queries, such as a query for all latest information from a given department. Although QCIS uses common standards such as HL7 and SQL, the actual queries being used can still be fairly system-specific. To encapsulate the system-specific details and reduce complexity, a data structure was designed to export the requested data from the query modules to other parts of the system and interact with the data access modules (DAMs).

Query modules were developed for each department. When using the modules, retrieving specific concepts can be as simple as supplying a list of concepts and receiving results from a data structure that was specially designed for data export. In the modules, queries are constructed for different departments. In some cases, multiple queries are needed to retrieve a concept from a possible source department. Since data returned by the DAM are not all in the same format and may contain extraneous information, a filter process is also included. After filtering, data are packed into the data structure for export.

Display generation. Faced with a variety types of data, the system's displays had to make the data representation reflect their differences while maintaining consistency. For example, most laboratory tests have numerical values and associated normal ranges while admission/discharge diagnoses are coded text terms. To achieve uniformity in the window design (layout, color, font, etc.), efforts were made to achieve consistency within the system and to conform to the existing Web-based clinical information system (CIS) browsers at NYPH as well. For example, abnormal findings are flagged in red, regardless of the source department.

As in any system, a large amount of information to present cannot fit into one screen or window all at once. In choosing design strategies such as "overview + details" or "focus + context," the determining factor was how we anticipate the views would be used for various clinical tasks. For instance, laboratory data are known for their huge volume; even the number of test results related to a particular concept could be quite large. When displaying all laboratory tests related to a concept, the same tests are grouped together and only the latest results of each laboratory test are shown. Such a design provides users with an overview of which laboratory tests have been performed and what the latest status is. The complete history of a test is available by clicking the test name, if a user is interested in it. In the case of chest X-ray and mammography reports, the number of reports is not normally as large as laboratory test results. Each report, however, can be long and complex. When giving users an index of reports that contain the relevant concepts, the relevant concepts are displayed along with each report name. By doing so, QCIS provides users with a focal point before they go into the full-length reports.

RESULTS

The implemented system is capable of handling 8 classes of concepts in the MED: anatomic entity, measurable entity, specimen, patient problem, sampleable substance, display information, event information, and orderable entity. There are more than 40,000 concepts covered by the 8 classes, and they account for 76% of all concepts in the MED.

The screen shots illustrate how QCIS works. Users log in to the system by supplying a medical record number. Users then can choose from three types of views: view by department (source-oriented view), view by time (time-oriented view), and view by topic (concept-oriented view) (Fig. 5a).

When users choose to view by department, a list of clinical departments is shown (Fig. 5b). Not all patients have data in all departments, so only the departments that contain data are shown. When a department is selected, an index of available clinical reports is displayed (Fig. 5c). Since there can be thousands of reports for one patient from one ancillary department, the reports are organized in reverse chronological order, with 20 reports per screen shown. The content of each report can be reached by clicking the report name; it will be displayed in another frame (Fig. 5d).

When users choose to view by time, a list of years is shown (Fig. 6). After a year is chosen, a calendar-like display is generated for that year. In the display, each cell (shown as a dot) represents a day in a clinical department. If there are any data for the patient on a particular day in a department, the corresponding cell will be highlighted in the color unique to that department (Fig. 7). When a highlighted cell is clicked, all data from the department for the patient on that day are shown.

When users choose to view by topic, they first need to type in and submit a term as the focus (Fig. 8). This term is mapped to MED concepts; the matches, if any, are displayed in a pop-up window (Fig. 9). Users may select a matching concept by clicking it, whereupon it becomes the concept of interest for the view. The system then performs concept expansion and identifies the department where related concepts may be found. A short list of departments that may have data relevant to the concept is displayed. When a user clicks on a particular department, relevant clinical data from that department are shown (Figs. 10 and 11).

Evaluation

We evaluated QCIS's ability to identify relevant patient information and the impact of the resulting views on clinical

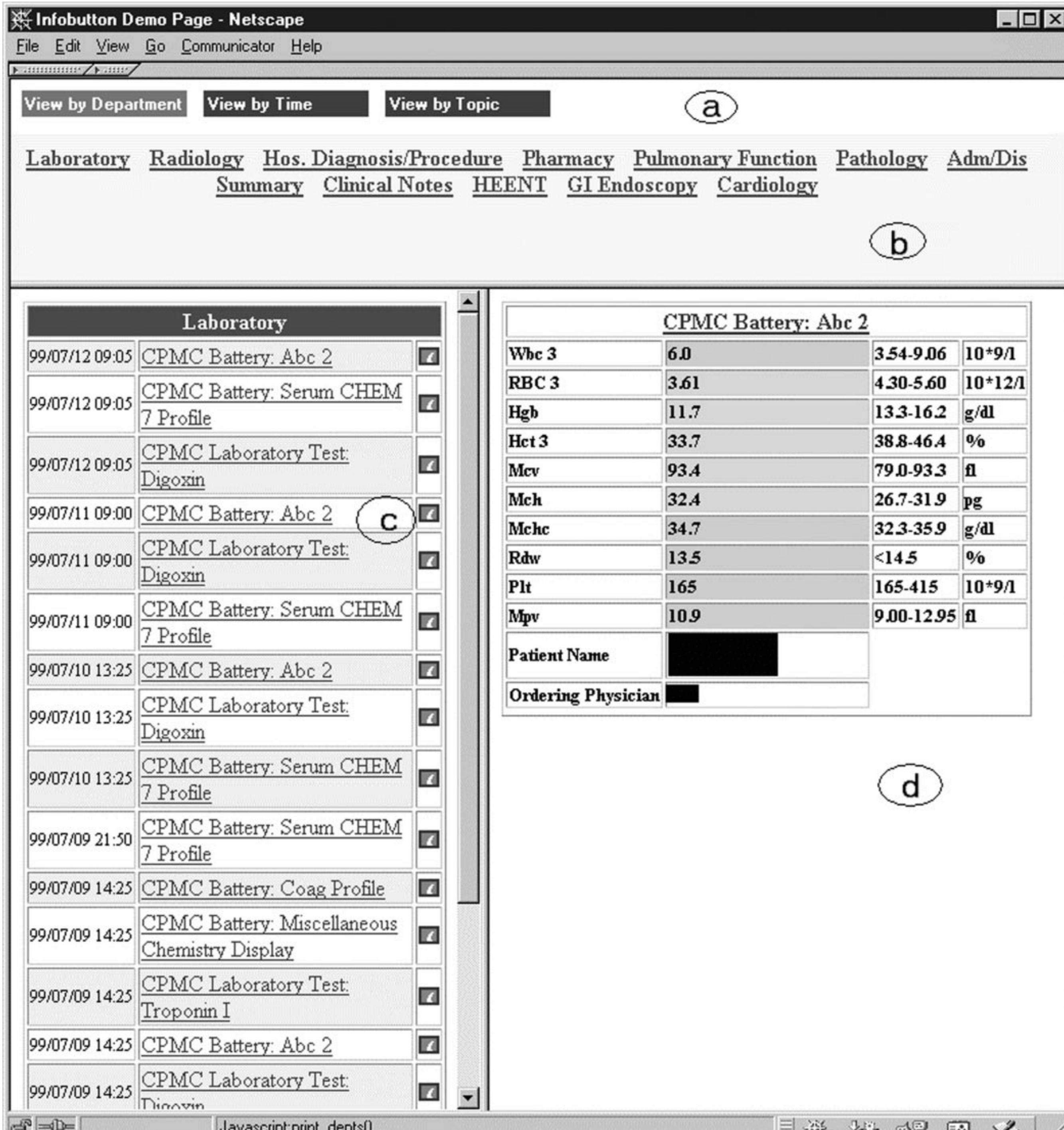

FIG. 5. Three types of views were available for users (a). When *View by Department* was chosen, a list of clinical departments was shown (b). After the laboratory department was selected, an index of laboratory reports was displayed (c). The details of a laboratory report were displayed after a click on the name of the report (d).

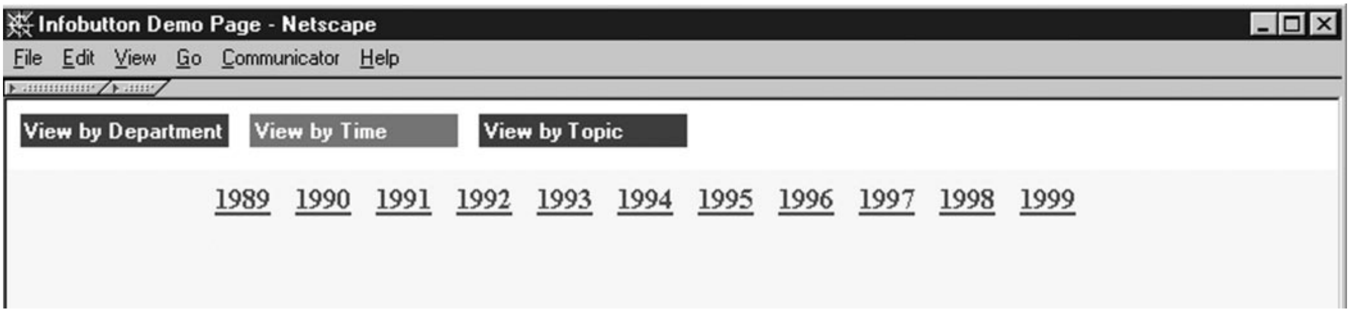

FIG. 6. When *View by Time* was chosen, a list of years was displayed.

information retrieval. The detailed methods and results of the evaluation will be reported in a separate paper. The evaluation was divided into three parts.

1. Quality of Relevant Information Identification: To determine the sensitivity and specificity of relevant information identification in certain clinical contexts such as hypothesis validation or disease management.

2. Information Overload Reduction: To determine the degree of reduction of the amount of information in the concept-oriented views comparing with viewing the whole patient record.

3. Effect on Information Retrieval: To determine if there is any advantage to using the concept-oriented views to retrieve information for patient care purposes, compared

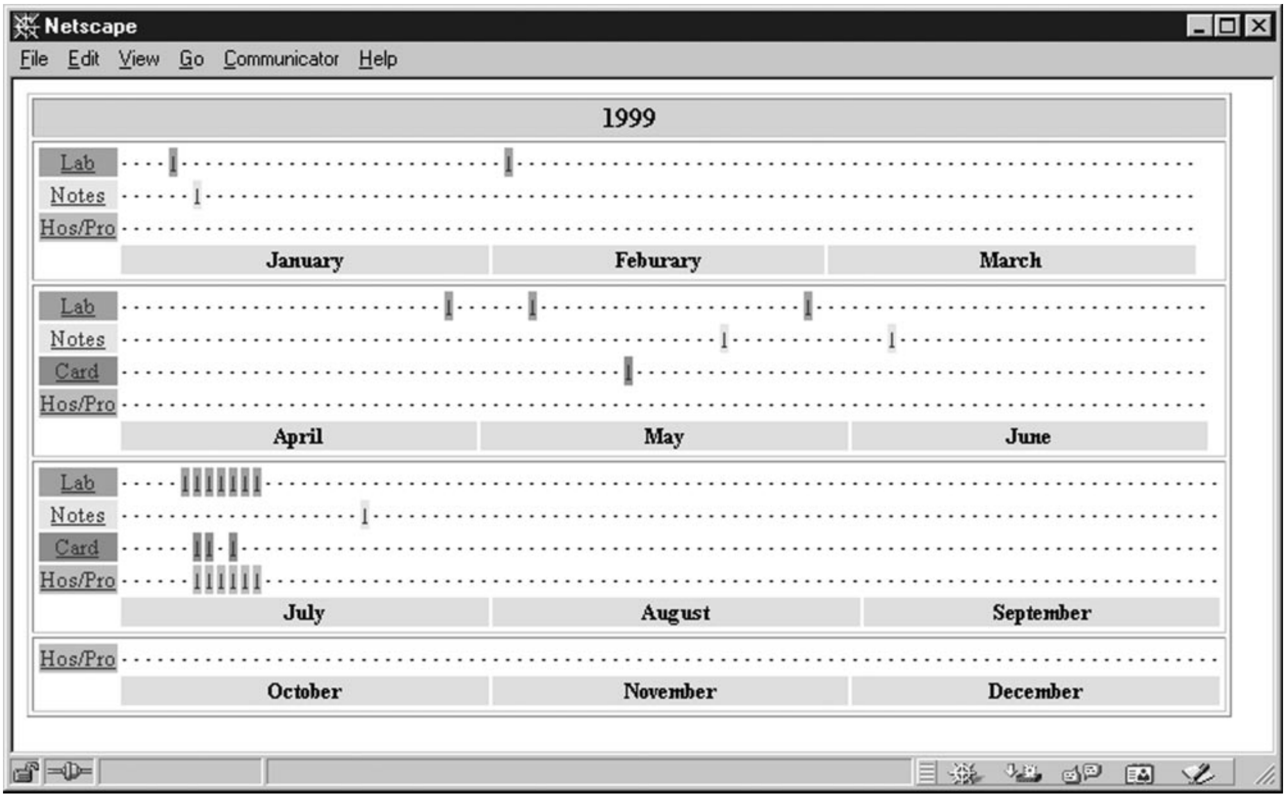

FIG. 7. After selecting the year 1999, available clinical reports in that year was shown. Cells were highlighted with different colors to represent data from different clinical departments on various dates.

FIG. 8. After *View by Topic* was selected, a medical term (*congestive heart failure*) was supplied to the system as the concept of interest.

with using more traditional ancillary department-oriented views.

The evaluation showed that in certain areas QCIS's ability to identify relevant clinical information (sensitivity, area under receiver operating characteristic (ROC) curve) was comparable to physicians' ability. This finding also validated the knowledge sources employed by the system. The evaluation further revealed that the concept-oriented view system reduced the amount of information retrieved and improved physician information retrieval accuracy compared with a source (department)-oriented view system.

DISCUSSION

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

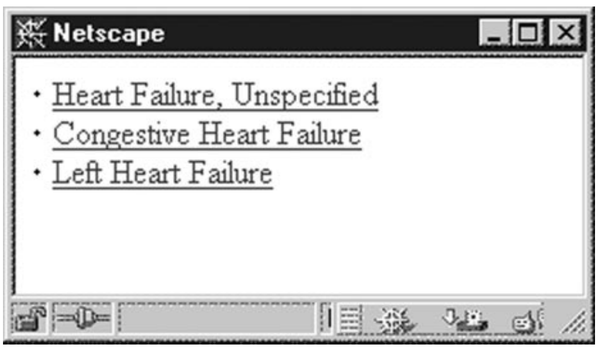

FIG. 9. Three MED concepts matching *congestive heart failure* were displayed in a pop-up window.

Significance

System design and implementation. In the past, various kinds of concept-oriented views have been proposed, and systems capable of generating such views were developed [1, 2, 5, 7, 21, 27, 28, 32, 33, 40, 41]. The features that distinguished our system from previous systems are that it is knowledge-based, capable of handling a large number of concepts, and that it is capable of working with real clinical data.

A key step in concept-oriented view generation is to identify information relevant to the concepts of interest. Some systems require users to manually identify the relevant information for each patient [27, 42–44]. For example, clinicians might need to manually link blood sugar tests to "diabetes" so that the system may include the tests in a diabetes view. In some systems, relevant information is predefined by domain experts [5, 7, 32, 41]. For example, blood sugar tests can be predefined as part of a diabetes view. The system discussed in this paper used a more general, knowledge-based approach [28, 33]. Domain knowledge is employed to identify information related to concepts. Contents in a view were no longer to be directly defined by clinicians. For example, a system might have the knowledge that sugar is an important chemical related to diabetes and therefore include blood sugar tests in a diabetes view.

Most of the previous systems could only handle one or two types of concepts, such as organ systems and patient problems [27, 28, 40]. Our knowledge-based approach, along with the knowledge acquisition and representation efforts, enabled our system to handle a wider range of concepts (over 40,000), including anatomic entity, measurable entity, specimen, patient problem, sampleable substance, display information, event information, and orderable entity.

Some significant concept-oriented view generation systems, for instance Tang's and Tange's systems, were pilot systems that worked with mock clinical databases [21, 28, 40]. Like Weed's and Barrows' systems, this system works

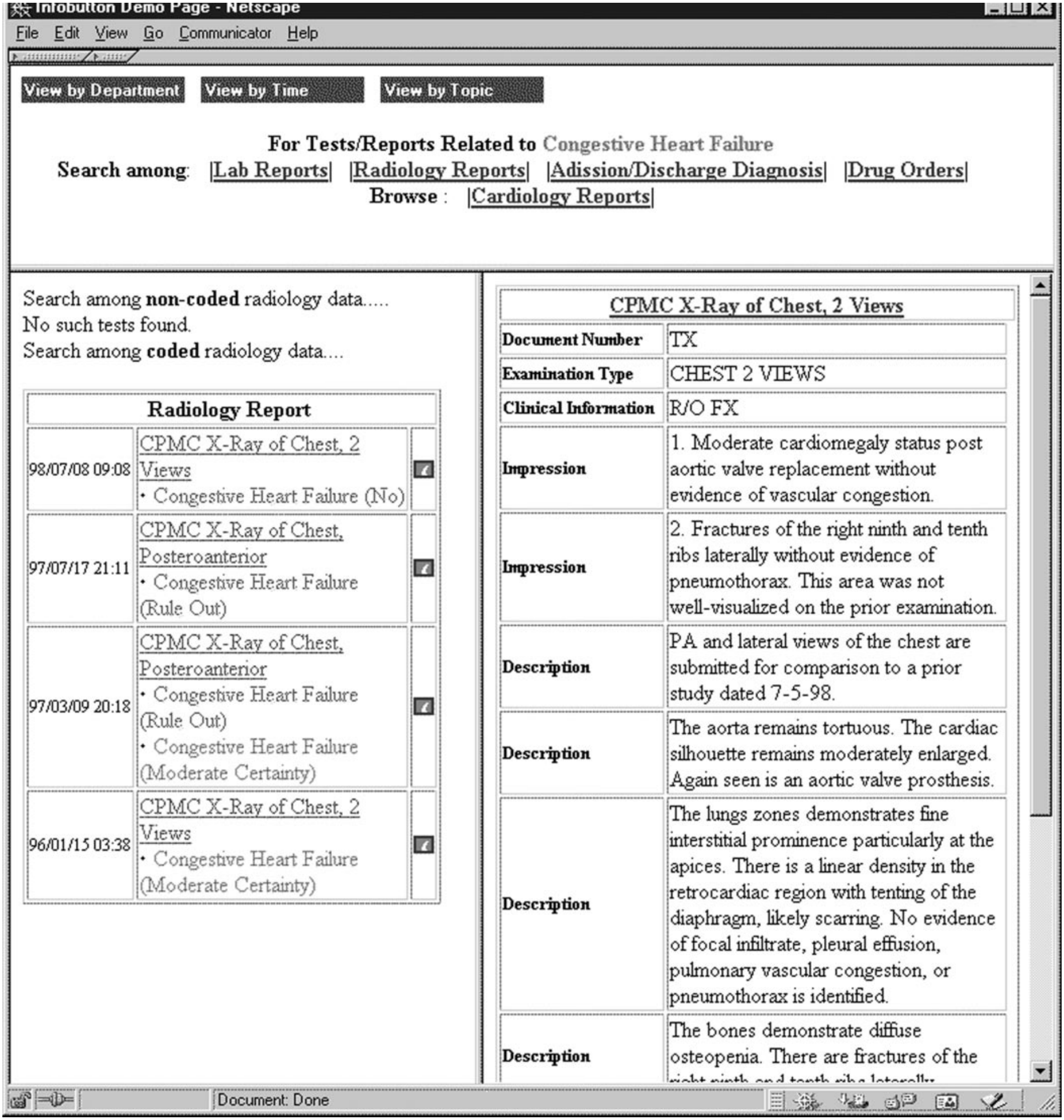

FIG. 10. When *Congestive Heart Failure* was chosen as the concept of interest, a list of departments was shown. After selecting *Radiology Reports*, the system returned a list of radiology reports related to congestive heart failure, and the content of a report was displayed after clicking the report name.

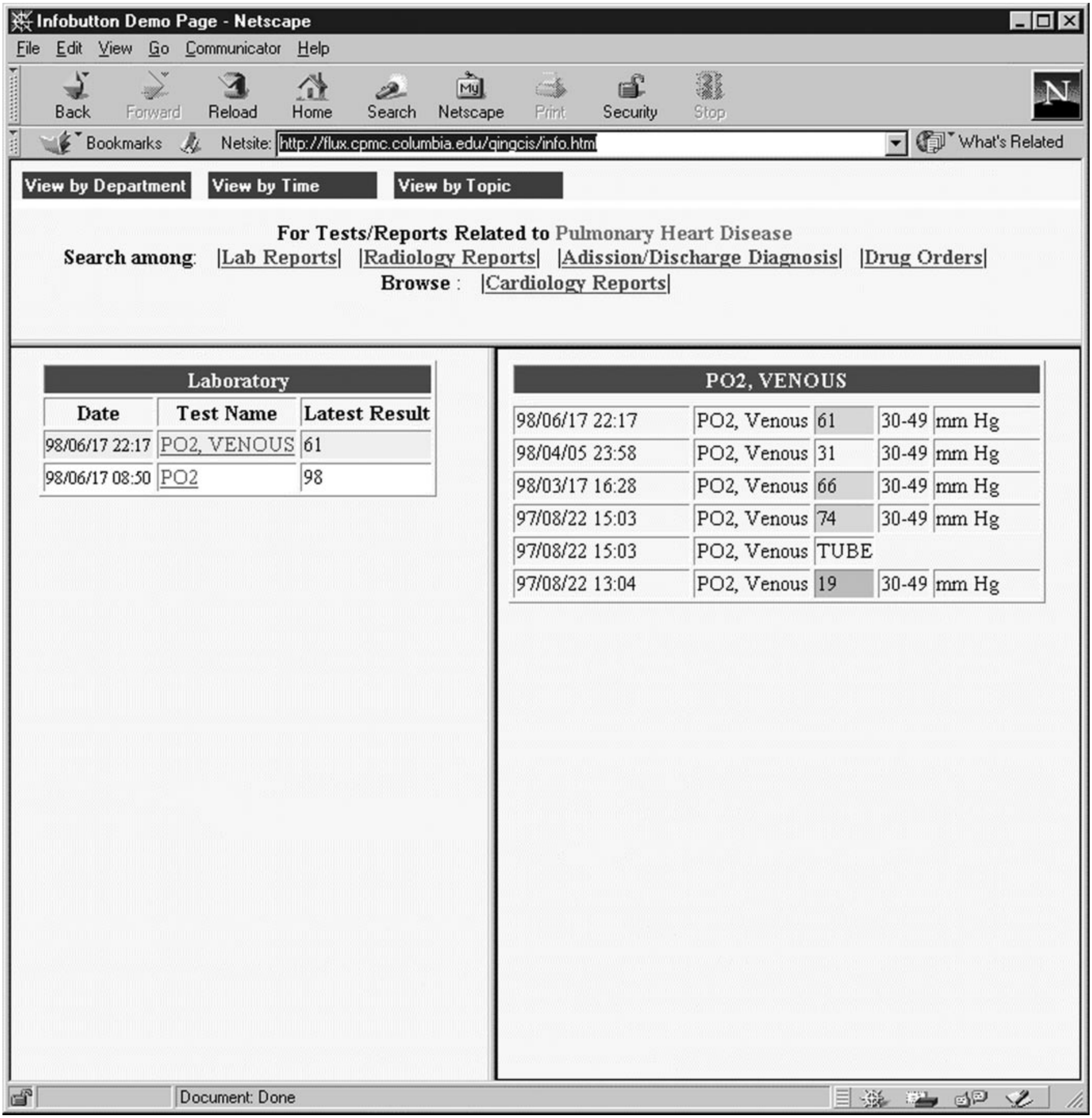

FIG. 11. When *Pulmonary Heart Disease* was chosen as the concept of interest, a list of departments was shown. After selecting *Lab Reports*, the system returned a list of the most recent laboratory reports related to pulmonary heart disease. Past results of a test were displayed after clicking the test name.

with a real clinical database which offers a complexity and variety that is hard to find in artificially-created test patients' data [27, 32].

Through system implementation, the chosen representations (a semantic network and rules) were shown to be sufficient and efficient for the knowledge needed by this view generation system. The same knowledge can often be represented by different schemes. For each system, difficult choices have to be made regarding which scheme to use.

Very few of the earlier concept-oriented view generation systems used a knowledge-based approach, and little discussion could be found about the knowledge sources and knowledge acquisition methods involved. For our system, knowledge was extracted from several sources (experts, published literature, and computer-based resources) for the concept-oriented view generation, although major efforts were made to extract disease-chemical relationship knowledge from the UMLS and DXplain using automated methods.

Previous research has been carried out on using the UMLS as a knowledge source, but not for the purpose of concept-oriented view generation [38, 45–54]. Our work can be viewed as an extension of the ongoing efforts to use the UMLS as a knowledge source. The knowledge in DXplain was collected to make diagnoses. In our system, it was transformed and reused for view generation. Later evaluations of the overall system did validate that UMLS and DXplain were appropriate knowledge sources for concept-oriented view generation.

Limitations

QCIS was utilizing a real-life clinical repository as its foundation. The content of the concept-oriented views was limited to the available kinds of coded data in the repository. For example, notes are not coded, and information contained in the notes could not be included in the generated concept-oriented views.

Although the certainty of some knowledge of relationships is hardly arguable, for example, that of a sickle cell test to sickle cell disease, people can disagree on other knowledge, such as what expansion rules are valid. In our system, the expansion rules were proposed by the author and approved by a domain expert. Physicians may choose the concept of interest for a view, but they cannot define the view by changing expansion rules. This may inconvenience and distress some users.

The knowledge on UMLS and DXplain was not originally collected for use in view generation. It is therefore inherently challenging to reuse their knowledge. They, however, were explored because of their broad coverage, free availability, and simple electronic formats.

Concept-oriented views do not guarantee the coverage of all patient data. Some patient data may not be included in views centered around user-selected concepts. The only way for users of this system to make sure that they have seen all the data about a patient when using concept-oriented views is to select all of the concepts in the MED.

Because people are used to having only one or two views of the EMR, it may take time for some to accept concept-oriented views. Those who are in the habit of browsing may feel that browsing is the only safe way to obtain information from EMRs and that using concept-oriented view may result in missing the whole picture. This concern is legitimate given that automatic view generation research is still at an early stage. As technology improves, the quality of concept-oriented views will improve and may gain more acceptance.

Implications

Although concept-oriented views such as problem-oriented views are desirable, very few systems offer them because of the difficulty of generating and maintaining the views. This work showed that it is possible to generate concept-oriented views through automated relevant patient information identification.

It is not the our intention to claim that concept-oriented views are superior to other views, although the concept-oriented view is the focus of this project. On the contrary, we believe that different types of views complement each other. Considering the complex nature of medical practice, an ideal system would be able to offer multiple types of views and hybrid views.

The concept-oriented view can be suitable for three purposes:

1. To filter out extraneous information. As we have mentioned in the background, information overload is a problem many physicians face. Filtering out extraneous information would allow users to focus on information relevant to specific tasks.

2. To detect relevant information. Many systems have been developed to monitor drug–drug interaction because of the clinically significant relationships between drugs. Such relationships also exist between other concepts, for example, between drugs and problems. Views can be used to identify such relationships and to provide necessary information related to a concept such as a disease or an organ.

3. To locate specific tests, reports, or findings. Most

EMRs are designed for browsing instead of searching because of the assumption that physicians need to review all information and do not need assistance with searching because they already know where to look for a specific piece of information. Physicians, however, do need to search for specific information sometimes. For example, a physician may want to know if cardiac catheterization was ever performed on a patient. Some physicians may not know exactly which department handles catheterization (e.g., cardiology or radiology) and, even if they do, they will need to check through all records in that department for the event.

Future Direction

Knowledge is still the most challenging issue for concept-oriented view generation. For future studies, the use of additional knowledge sources should be explored. Knowledge from new sources may complement sources currently being used and improve the system's sensitivities; such sources may also help to improve specificities if used to confirm knowledge already acquired. For example, some commercial knowledge bases may have more accurate knowledge about disease-drug chemical relationships. Expansion rules are another piece of knowledge that can be improved by gathering more expert opinions.

In addition to improving the knowledge quality, QCIS could also be designed to allow users to define their own versions of certain types of knowledge that are subject to disagreement, such as expansion rules. This could give users more control over the view generation process and could increase their confidence in using the system. It, however, would also require users to formalize their knowledge, which would be difficult for some users.

CONCLUSION

We have developed a general-purpose, knowledge-based approach to the generation of concept-oriented views for coded patient data. This approach was tested through the implementation of a multiple-view generation system with a focus on concept-oriented views. Given a user-selected concept of interest, the system retrieved relevant coded clinical data and presented them as a view. Four major categories of knowledge resources were employed: existing knowledge bases, on-line information sources, domain experts, and medical literature. The system used a semantic network and rules as its knowledge presentation scheme and performed relevant data identification through a rule-based traversal of the semantic network.

REFERENCES

1. Weed LL. Medical records that guide and teach. N Engl J Med 1968; 278(11):593–600.
2. 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.
3. Fries JF. Alternatives in medical record formats. Med Care 1974; 12(10):871–81.
4. Willard KE, Connelly DP, Johnson JR. Radical improvements in the display of clinical microbiology results: a web-based clinical information system [published erratum appears in Am J Med 1997 Jul; 103(1):86]. Am J Med 1996; 101(5):541–9.
5. 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.
6. 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.
7. 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.
8. Institute of Medicine, Steen EBE, Dick RSE. The computer-based patient record: an essential technology for health care. Washington DC: National Academy Press, 1991.
9. Rose JS. Medicine & the information age. Tampa, FL: American College of Physician Executives, 1998.
10. Powsner SM, Tufte ER. Graphical summary of patient status. Lancet 1994; 344(8919):386–9.
11. Kinkhorst OM, Lalleman AW, Hasman A. From medical record to patient record through electronic data interchange (editorial). Int J Biomed Comput 1996; 42(1–2):151–5.
12. Tange HJ. The paper-based patient record: is it really so bad? Comput Methods Programs Biomed 1995; 48(1–2):127–31.
13. Elhanan G, Socratous SA, Cimino JJ. Integrating DXplain into a clinical information system using the world wide web. Proc AMIA Annu Fall Symp 1996; 348–52.
14. Hendrickson RM, Lee J. Physicians' impressions of old medical records. J Fam Pract 1983; 17(6):1081–2.
15. Newell A, Shaw JC, Simon HA. Elements of a theory of human problem solving. Psych Rev 1958; 65:151–66.
16. Miller GA, Galanter E, Pribram KH. Plans and structure of behavior. New York: Holt, 1960.
17. Newell A, Simon HA. Human problem solving. Englewood Cliffs, NJ: Prentice-Hall, 1972.
18. Elson RB, Faughnan JG, Connelly DP. An industrial process view

of information delivery to support clinical decision making: implications for systems design and process measures. J Am Med Inform Assoc 1997; 4(4):266–78.

19. Wyatt JC. Clinical data systems, part 1: data and medical records [see comments]. Lancet 1994; 344(8936):1543–7.
20. Hayes GM. Medical records: past, present, and future. Proc AMIA Annu Fall Symp 1996; 454–8.
21. 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.
22. Powsner SM, Tufte ER. Summarizing clinical psychiatric data. Psychiatr Serv 1997; 48(11):1458–61.
23. Fries JF. Time-oriented patient records and a computer databank.-JAMA 1972; 222(12):1536–42.
24. The problem of the problem-oriented record. N Engl J Med 1973; 288(21):1133–4.
25. Fleming JW. The problem-oriented medical record. Is there more than promise in promis? JFMA 1973; 60(6):34–7.
26. Gilroy G, Ellinoy BJ, Nelson GE, Cantrill SV. Integration of pharmacy into the computerized problem-oriented medical information system (promis)—a demonstration project. Am J Hosp Pharm 1977; 34(2):155–62.
27. Ali MY. Problem-oriented medical information system (promis) for Australian medical education. Med J Aust 1983; 2(6):293–5.
28. 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.
29. 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.
30. 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.
31. Elhanan G, Cimino JJ. Automated integration of a diagnostic decision support system with a large scale clinical information system—aspects of feasibility. Proc AMIA Annu Fall Symp 1998.
32. Barrows RC Jr, Allen B, Fink DJ. An x window system for statlab results reporting. Proc Annu Symp Comput Appl Med Care 1993; 331–5.
33. Hripcsak G, Allen B, Cimino JJ, Lee R. Access to data: comparing accessmed with query by review. J Am Med Inform Assoc 1996; 3(4):288–99.
34. Friedman C. Towards a comprehensive medical language processing system: methods and issues. Proc AMIA Annu Fall Symp 1997; 595–9.
35. Johnson S, Friedman C, Cimino JJ, Clark T, Hripcsak G, Clayton P D. Conceptual data model for a central patient database. Proc Annu Symp Comput Appl Med Care 1991; 381–5.
36. 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.
37. Zeng Q, Cimino JJ. Automated knowledge extraction from the UMLS. Proc AMIA Symp 1998; 568–72.
38. Selden CR, Humphreys BL. Unified Medical Language System (UMLS). U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Library of Medicine, 1997.
39. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. Dxplain. An evolving diagnostic decision-support system. JAMA 1987; 258(1):67–74.
40. Tange HJ, Hasman A, de Vries Robbe PF, Schouten HC. Medical narratives in electronic medical records. Int J Med Inf 1997; 46(1):7–29.
41. Elhanan G, Cimino JJ. Controlled vocabulary and design of laboratory results displays. Proc AMIA Annu Fall Symp 1997; 719–23.
42. Stratmann WC, Goldberg AS, Haugh LD. The utility for audit of manual and computerized problem-oriented medical record systems. Health Serv Res 1982; 17(1):5–26.
43. Cheong PY, Goh LG, Ong R, Wong PK. A computerised outpatient medical records programme based on the summary time-oriented record (stor) system. Singapore Med J 1992; 33(6):581–7.
44. van Ginneken AM, Stam H, Duisterhout JS. A powerful macro-model for the computer patient record. Proc Annu Symp Comput Appl Med Care 1994; 496–500.
45. Joubert M, Fieschi M, Robert JJ. A conceptual model for information retrieval with UMLS. Proc Annu Symp Comput Appl Med Care 1993; 715–9.
46. Kingsland LC 3d, Harbourt AM, Syed EJ, Schuyler PL. Coach: applying UMLS knowledge sources in an expert searcher environment. Bull Med Libr Assoc 1993; 81(2):178–83.
47. Miller RA, Gieszczykiewicz FM, Vries JK, Cooper GF. Chartline: providing bibliographic references relevant to patient charts using the UMLS metathesaurus knowledge sources. Proc Annu Symp Comput Appl Med Care 1992; 86–90.
48. Nelson SJ, Fuller LF, Erlbaum MS, Tuttle MS, Sherertz DD, Olson NE. The semantic structure of the UMLS metathesaurus. Proc Annu Symp Comput Appl Med Care 1992; 649–53.
49. Payne TH, Martin DR. How useful is the UMLS metathesaurus in developing a controlled vocabulary for an automated problem list? Proc Annu Symp Comput Appl Med Care 1993; 705–9.
50. Pratt W. Dynamic organization of search results using the UMLS. Proc AMIA Annu Fall Symp 1997; 480–4.
51. Schuyler PL, Hole WT, Tuttle MS, Sherertz DD. The UMLS metathesaurus: representing different views of biomedical concepts. Bull Med Libr Assoc 1993; 81(2):217–22.
52. Suarez HH, Hao X, Chang IF. Searching for information on the internet using the UMLS and medical world search. Proc AMIA Annu Fall Symp 1997; 824–8.
53. Tuttle MS, Nelson SJ. The role of the UMLS in 'storing' and 'sharing' across systems. Int J Biomed Comput 1994; 34(1–4):207–37.
54. Cimino JJ, Sideli RV. Using the UMLS to bring the library to the bedside. Med Decis Making 1991; 11(4 Suppl):S116–20.