Research Paper ■

A Logical Foundation for Representation of Clinical Data

KEITH E. CAMPBELL, MD, AMAR K. DAS, MARK A. MUSEN, MD, PHD

Abstract **Objective**: A general framework for representation of clinical data that provides a declarative semantics of terms and that allows developers to define explicitly the relationships among both terms and combinations of terms.

Design: Use of conceptual graphs as a standard representation of logic and of an existing standardized vocabulary, the Systematized Nomenclature of Medicine (SNOMED International), for lexical elements. Concepts such as time, anatomy, and uncertainty must be modeled explicitly in a way that allows relation of these foundational concepts to surface-level clinical descriptions in a uniform manner.

Results: The proposed framework was used to model a simple radiology report, which included temporal references.

Conclusion: Formal logic provides a framework for formalizing the representation of medical concepts. Actual implementations will be required to evaluate the practicality of this approach.

■ **J Am Med Informatics Assoc.** 1994;1:218–232.

1. Clinical Data Representation

Collection of clinical data is expensive and time-consuming. Currently, only minimal data are encoded routinely. Any analysis requiring new data elements consumes significant resources to develop a sampling plan for the data, to develop data-collection instruments, to train the data abstracters, to collect the data, and to analyze the data.

In today's cost- and quality-conscious atmosphere, there is increasing interest in encoding more data on a routine basis, such as the data contained in admission histories and physical examinations, progress notes, and discharge summaries. Some researchers have been working on applications to collect these data as part of the process of care.[1-6] Other researchers have been developing applications to extract these data from narrative text.[7-10] Agencies such as the Health Care Financing Administration (HCFA) recognize that the process of standardizing and automating representation of medical information is essential for increasing the efficiency of their programs.[11,12]

Today, there is no existing standard capable of representing the detailed clinical data contained within histories and physical examinations, progress notes, and discharge summaries.

In this paper, we describe a model for how these clinical data (including their temporal dimensions) might be represented. We start by describing the problems with existing medical terminologies, and by giving a brief historical perspective on concept classification. Next, we describe our rationales for using formal logic as a foundation and conceptual graphs as a standard representation of logic, and provide an example of how a simple radiology report could be represented using conceptual graphs. Later, we describe the need for foundational models to represent topics such as anatomy and time, and we use

Affiliation of the authors: Stanford University School of Medicine, Stanford, CA.

Supported in part by the Agency for Health Care Policy and Research under grant HS06330, by the National Library of Medicine under grants LM05208 and LM07033, and by gifts from the Computer-Based Assessment Project of the American Board of Family Practice.

Presented in part at the annual Symposium on Computer Applications in Medical Care, 1992.

Correspondence and reprints: Keith E. Campbell, MD, Section on Medical Informatics, Medical School Office Building X-215, Stanford University School of Medicine, Stanford, CA 94305-5479.

Received for publication: 11/12/93; accepted for publication: 1/14/94.

time as an example to show how foundational models can be represented using formal logic. Next, we describe how an existing terminology system, the Systematized Nomenclature of Medicine (SNOMED International),[13] can be enhanced to fit within our framework. Finally, we provide a critical discussion of our work.

2. Common Terminology Structure

Typical medical terminologies, such as SNOMED International[13] and the International Classification of Diseases, 9th ed., with Clinical Modifications (ICD-9-CM),[14] use a hierarchical structure that organizes the concepts into *type hierarchies*. As an example, we show SNOMED and ICD-9-CM classifications of "pleural effusion" in Figure 1.

These type hierarchies provide a mechanism for indexing terminology within the system; however, such hierarchies have significant shortcomings. For example, this simple categorization neither defines sufficiently what a term represents nor tells how one term differs from another. It simply indicates that terms with the same parents are related, and that they are also different in that they were not given the same term code. This kind of classification scheme works well for a system in which a term needs only one parent, in which only terminology that is unambiguously interpreted by anyone using the type hierarchy is used, and, finally, in which minimal automated processing of the terms in the type hierarchy is required. In other cases, this mechanism has serious limitations. These limitations may contribute to the significant random errors in ICD-9-CM classification of hospital discharge diagnoses. Error rates between 20% and 29% have been reported in the literature.[15-17] If a terminology lent itself to automated consistency checking, perhaps these errors could be reduced.

The most serious limitation of vocabularies that use only type hierarchies is their lack of a *formal definition* for each term. If each term in the terminology were defined formally, it could be used in a more consistent manner. Some medical-terminology systems, such as the Medical Subject Headings (MeSH)[18] and the Diagnostic and Statistical Manual of Mental Disorders (DSM-III),[19] have textual definitions intended for the human reader. If these definitions also could be processed by a machine, then we could develop tools to process the terminology in a consistent way.

3. Historical Perspective

Type hierarchies, similar to the ones used by most medical-terminology systems, were first introduced by Aristotle, around 300 B.C., with his theory of categories.[20] Aristotle also developed a method for defining new types within the type hierarchy by *genus* (the category of classification for a term) and *differentia* (the elements, features, or factors that distinguish one term from another), and for using deductive arguments to analyze the inheritance of proper-

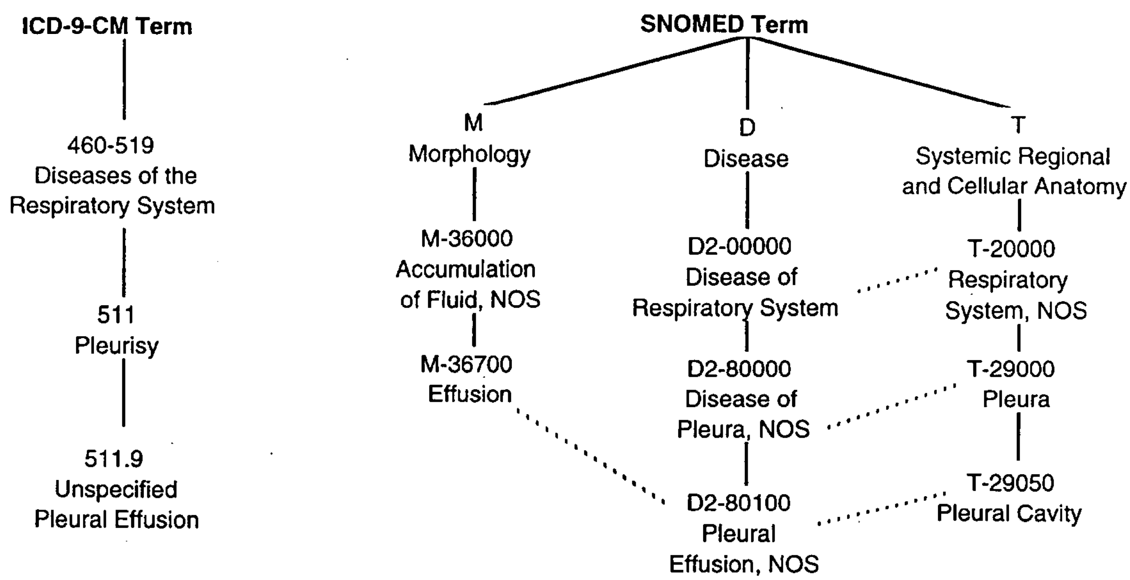

Figure 1 ICD-9-CM type hierarchy (left) and SNOMED type hierarchy (right) showing classification of "pleural effusion." The dashed lines represent cross-reference links provided by SNOMED. There were no cross-references for the ICD-9-CM terms. The SNOMED cross-references are discussed in Section 8.2.

ties of these new types. This kind of type hierarchy is called *Aristotelian*. The type hierarchies in Figure 1 already define the genus of each term by the lines connecting each term to its *parent* or genus. To make these type hierarchies Aristotelian, we need to specify differentia for each term. An example of a definition that contains both the genus and differentia for the term PLEURAL-EFFUSION can be represented in English by the phrase "a pleural effusion is an effusion located in the pleural cavity and caused by a disease." The genus is represented by the first portion, "a pleural effusion is an effusion," and the differentia is represented by the last portion, "located in the pleural cavity and caused by a disease."

A little over 300 years ago, Leibniz developed the first system capable of *computing* the elementary concepts of which more complex terms are composed. This system, the Universal Characteristic, represented *primitive concepts* as prime numbers. It created *compound concepts* by multiplying primitive concepts together. If PLEURAL were represented by 3 and EFFUSION were represented by 7, their product, 21, would represent PLEURAL-EFFUSION. Leibniz's system was certainly visionary, and led to his development of the first mechanical computer capable of performing multiplication and division.

The Universal Characteristic was essentially the first mechanical implementation of a multiple-inheritance type hierarchy. Because it implements only a type hierarchy, and is not able to represent machine-processible definitions by means of differentia, it suffers from the same limitations of type hierarchies described in Section 2. In fact, the only logical relations permissible in such a system are *conjunctions* of primitives. To represent the logical relations necessary to define fully a system for medical concept representation, we need more complex relations, such as *negation* and *disjunction*, as well as the ability to use *defined relations*, such as IS-A relationships and PART-OF relationships.

Although Leibniz was limited by having only mechanical computers capable of performing multiplication and division, today we have far fewer constraints on what we are able to compute reasonably. Modern computing power makes it possible to implement concept-representation systems that define terms by genus and differentia, which in turn will allow the development of automated methodologies to process the terminology. The more complete the differentia defined by such a system, the more powerful the tools that can be developed. The ideal system would allow the differentia to be defined in a complete system of logic.

Although comprehensive logic-based approaches have not yet affected our common medical naming systems, other researchers have been exploring how formal systems could be used for medical terminology. Cimino and colleagues have constructed the Medical Entities Dictionary using a semantic network.[21] Masarie and colleagues described a frame-based approach to map equivalent concepts between clinical vocabularies.[22] In another project, Rector and colleagues have described an approach that uses their Structured Meta Knowledge (SMK) formalism.[23] These systems can also be described using first-order logic. If we adopt logic as a common foundation for representation of medical concepts, all applications will benefit.

4. Logical Foundation

Predicate logic is relevant to any area of reasoning. It is *topic neutral*.[24] Logic can be applied to problems in any domain with equal validity. In certain fields, such as linguistics and knowledge representation, practitioners use logic extensively. Computers use logic at their lowest level, in the switching of their circuitry, and at their highest level, in logic programming languages such as PROLOG and to represent knowledge in expert-system shells such as KL-ONE.[25]

We use logic as the basis for our proposal, but we realize that logic is not without its critics. Early artificial-intelligence systems ignored logical soundness. Some early researchers ignored logical soundness because they were overwhelmed by problems of syntax and semantics.[26] Other early researchers lacked understanding about what logic is, confusing logic with a particular programming system of syntax.[27] Today, logic has been making a steady resurgence with the development of a family of *term-subsumption languages*.[28–30] In addition to these systems, new standards for interchanging information are based on logic such as the Knowledge Interchange Format (KIF)[31] being developed with support from the Advanced Research Project Agency (ARPA), and the Information Resource Dictionary Standard (IRDS) developed by the American National Standards Institute (ANSI).[32]

Logic's notation makes it possible to formalize relationships between terms in a system, so that applications can make valid inferences using those relationships. These formal relationships are the most powerful argument in logic's favor. Without this formality, we must use ad hoc approaches that may result in questionable conclusions because the methods of inference cannot be proved sound.

5. Conceptual Graphs

Conceptual graphs are a system of logic that is able to represent complete first-order, modal, and higher-order logics. They were developed as a more intuitive notation for logic. To illustrate the benefits of this notation, we shall give an example that uses different notations for logic. Using logic, we can state the definition of the term PLEURAL-EFFUSION given in Section 3 as:

> For all x, x is a pleural effusion if x is an effusion, and for some pleural cavity y and for some disease z, x is located in y and x is caused by z.

Or, using predicate-calculus notation, we restate the definition as follows:

$$\forall x \text{ PLEURAL-EFFUSION}(x) \equiv \text{EFFUSION}(x) \wedge \exists y \text{ PLEURAL-CAVITY}(y) \wedge \exists z \text{ DISEASE}(z) \wedge \text{LOCATED-IN}(x,y) \wedge \text{CAUSED-BY}(x,z).$$

The predicate-calculus notation for logic was developed by mathematicians, who patterned it after algebra. The awkwardness of this notation was noted by one of its original developers, Charles Saunders Peirce, who, in 1897, developed a graphic notation for logic that he called *existential graphs*.[33] The modern descendant of existential graphs is *conceptual graphs*, as described by Sowa.[26]

The attractive features of conceptual graphs have been noted previously by other medical-informatics researchers who are using them in applications.[7,34–41] These features include the ability to represent complex relationships among entities; to express selection constraints for any given entity; to map conceptual graphs onto database representations; and to map onto other formal systems, such as first-order predicate calculus. In addition to these advantages, conceptual graphs also have an intuitive structure that can be understood easily. The same definition for a PLEURAL-EFFUSION represented in predicate calculus can be represented using conceptual-graph notation, as shown in Figure 2.

Conceptual-graph notation is relatively simple. A conceptual graph is a finite, connected graph with two kinds of nodes: *concept nodes* denoted by *boxes*, and *concept-relation nodes* denoted by *circles*. Conceptual nodes can be connected to only concept-relation nodes, and vice versa, making the graph *bipartite*.

There is also an alternate notation for conceptual graphs, called the *linear form*. The box-and-circle notation, shown in Figure 2, is not machine readable and often takes up excessive space on the page. The linear form uses only standard machine-readable characters and therefore is suitable as an interchange format. The linear notation uses square brackets for the boxes and parentheses for the circles. The definition for pleural effusion can be represented in the linear form as follows:

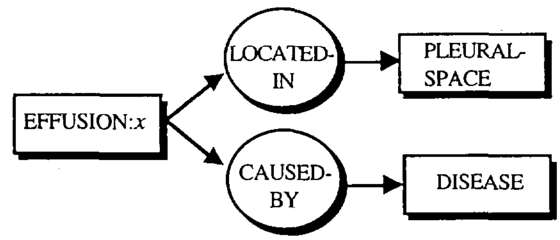

Figure 2 Conceptual-graph definition of "pleural effusion."

type PLEURAL-EFFUSION(x) *is*
  [EFFUSION: x]-
    (LOCATED-IN)→[PLEURAL-CAVITY]
    (CAUSED-BY)→[DISEASE],.

A conceptual node can refer to a general concept, or to an *instance* of a general concept. When a node refers to an instance, its label is followed by a colon and the node is marked with a symbol. This symbol is called the node's *referent*.

Some referants are marked with the "#" symbol followed by a *serial number*, representing a specific instance. Other referents are *variables*, represented by a string of characters that does not start with the "#" symbol. The serial number of a graph is valid for the entire collection of graphs. If two nodes have the same serial number, then they refer to the same instance. Variables are only local in scope. Two graphs that are not connected, yet share the same variable label, may not refer to the same instance.

Every conceptual-relation node has one or more arcs, each of which must be linked to some concept. Within a node, braces "{ }" denote a *set* of concepts, and a set that contains an asterisk "{*}" denotes a set of zero or more elements. When a set is preceded by a term with a colon, this term specifies selection constraints for the set of elements. All elements of the set must then be a specialization of the preceding term (all terms are considered specializations of themselves). A *collective set* can be represented by separating elements with commas: $\{i_1, i_2, i_3\}$. We can use this set notation to specify allowable values for laterality as follows:

[LATERALITY: {LEFT, RIGHT}]

This brief introduction to conceptual graphs is sufficient for the examples in this paper, but does not describe the conceptual-graph syntax completely. Additional details are discussed by Sowa.[26]

5.1 Canonical Graphs

A developer can create conceptual graphs by joining concept nodes and relation nodes together, as described earlier. It is possible to combine these nodes in ways that do not make sense. To distinguish graphs that actually represent true situations, developers declare certain graphs to be *canonical*, using their insight and perceptions. These graphs form the basis for creation of other graphs that also represent true situations. Developers must create an initial set of canonical graphs from their insight and their perceptions regarding clinically relevant information. Once the initial set of canonical graphs is defined, developers can derive new canonical graphs from existing canonical graphs by applying the rules copy, restrict, simplify, and join:

- *Copy*. Any graph may be duplicated.
- *Restrict*. Any concept label may be replaced by the label of a subtype as long as the referent of the node conforms to the type before and after the change.
- *Simplify*. If two conceptual relations in a graph are duplicated, then one of them may be deleted from the graph together with all its arcs, thus simplifying the graph.
- *Join*. If a concept in one graph is identical to a concept in another graph, then a new graph can be created by deletion of the concept in one graph and linkage of all the arcs from the concept to the identical concept in the other graph.

We shall give examples that use these canonical formation rules in Section 6.

5.2 Subsumption

Formal rules for *subsumption* are an important feature of logic, and are also essential for meaningful aggregation of clinical data. By using formal subsumption rules, we make it possible to ask the question "How many cases of pleural effusion are in the data set?" of a database and to have the database count properly all the concepts *subsumed* by "pleural effusion," such as "right-sided pleural effusion," "left-sided pleural effusion," and "an effusion located in the pleural cavity."

There are several ways that subsumption can be implemented. Conceptual-graph applications can implement subsumption using *structural subsumption*. Structural subsumption can be implemented using techniques of *subgraph isomorphism* that check to see whether the structure of one graph is completely contained within that of another. If one graph completely contains another, it is said to *subsume* the other graph. Efficient algorithms have been developed to support subsumption over certain classes of conceptual graphs.[42,43]

6. Proposed Model

Our proposed model uses conceptual graphs as a standard representation of logic to implement an Aristotelian classification of medical concepts. A variety of knowledge-representation systems, including PROLOG and KL-ONE, can implement this method of classifying medical concepts. We shall show how to use the set of concepts that we define in Table 1, together with the canonical formation rules decribed in Section 5.1, to represent the report of a posterior–anterior (PA)* chest radiograph. We shall start with the hypothetical example of patient John Doe (patient ID = 123-45-67). The chest radiograph was read by Dr. Smith (provider ID = MD-56789), who noted a left pleural effusion.

6.1 Type Definitions

In all type hierarchies, there must be some starting point from which all other concepts are derived. This concept is usually called the *top concept;* we refer to it simply as T. To create an Aristotelian hierarchy, we must specify differentia for every term. We are using conceptual graphs to specify these differentia by defining, for each term, relations and selection constraints that are appropriate for the term, but that are not appropriate for the term's immediate ancestor(s). These differentia are contained within the conceptual-graph *type definition*, a statement that incorporates both the genus and differentia of each term. The definitions necessary to represent the hypothetical example are presented in Table 1.

The type definition for the term DIAGNOSTIC-RADIOLOGY-PROCEDURE is prototypical. The genus of DIAGNOSTIC-RADIOLOGY-PROCEDURE is RADIOLOGY-PROCEDURE. The differentia of diagnostic radiology procedures is that an image is always created from which a report can be generated. We can define formally the genus and differentia of DIAGNOSTIC-RADIOLOGY-PROCEDURE by specifying the following type definition:

*The orientation of a chest radiograph where the patient is upright and the image is created by the x-rays entering the chest from the posterior, exposing the film placed on the anterior of the chest.

type DIAGNOSTIC-RADIOLOGY-PROCEDURE(x) *is*
[RADIOLOGIC-PROCEDURE: x]-
(RESULTS-IN)→[REPORT]
(CREATES)→[IMAGE],.

For all the type definitions that we describe, the term's label is tied to the term's genus by the referent "x." These referents are local in scope, and, when we join these graphs with other graphs, as we do in Section 6.3, we may need to rename the referent to prevent conflicts.

In some cases, we cannot adequately define a term by relating it to other terms within the type hierarchy. In these cases, the concept is considered *primitive*. In our type definitions, we indicate primitives by defining the genus of the concept as we normally would, but we define the differentia to include a relation labeled GD, which stands for *generic difference*. The generic difference is followed by a concept node given a label that is identical to the label of the type definition, except that the label is preceded by an underscore. An example defined this way would be the term LEFT:

type LEFT(x) *is*
[LATERALITY: x]→(GD)→[_LEFT].

This method for defining terms ensures that each primitive concept is related to a unique term that is

Table 1 ■

Concepts and Their Corresponding Type Definitions*

PROCEDURE	[T: x]→(PERFORMED-ON)→[PATIENT].
RADIOLOGIC-PROCEDURE	[PROCEDURE: x]→(USES)→[IONIZING-RADIATION].
DIAGNOSTIC-RADIOGRAPHY-PROCEDURE	[RADIOLOGIC-PROCEDURE: x]- (RESULTS-IN)→[REPORT] (CREATES)→[IMAGE],.
PA-CHEST-X-RAY	[RADIOLOGIC-PROCEDURE: x]- (RESULTS-IN)→[REPORT] (CREATES)→[IMAGE]- (OF-LOCATION)→[CHEST] (HAS-ORIENTATION)→[PA],,.
IMAGE	[T: x]- (CREATES)←[PROCEDURE] (OF-LOCATION)→[BODY-REGION] (HAS-ORIENTATION)→[ORIENTATION],.
OBSERVATION	[T: x]→(GD)→[_OBSERVATION].
LOCALIZED-OBSERVATION	[OBSERVATION:x]→(LOCATED-IN)→[LOCATION].
LOCATION	[T: x]→(GD)→[_LOCATION].
BILATERALLY-SYMMETRIC-LOCATION	[LOCATION: x]→(HAS-LATERALITY)→[LATERALITY].
LATERALITY	[T: x]→(GD)→[_LATERALITY].
RIGHT	[LATERALITY: x]→(GD)→[_RIGHT].
LEFT	[LATERALITY: x]→(GD)→[_LEFT].
PATIENT	[PERSON: x]→(HAS-PATIENT-ID)→[IDENTIFIER].
PROVIDER	[PERSON: x]→(HAS-PROVIDER-ID)→[IDENTIFIER].
EFFUSION	[LOCALIZED-OBSERVATION: x]→(GD)→[_EFFUSION].
PLEURAL-CAVITY	[BILATERALLY-SYMMETRIC-LOCATION: x]- (GD)→[_PLEURAL-CAVITY],.
CHEST	[LOCATION:x]→(GD)→[_CHEST].
PERSON	[T: x]→(HAS-NAME)→[string].
REPORT	[T: x]- (CREATES)←[PROVIDER] (HAS-OBSERVATIONS)→[OBSERVATION: {*}],.

GD = generic difference.
*See Section 6 for further explanation.

not within the type hierarchy. This method formally requires that any *primitive* concept is different from its parent term in a way that is not expressible within the system. By defining primitive concepts in this way, we can maintain the properties of an Aristotelian type hierarchy. Other examples of such primitive terms include RIGHT and LOCATION, as shown in Table 1.

Medical-terminology systems must allow definition of primitive concepts; however, primitive definitions should be avoided when practically possible. At least part of a primitive definition cannot be expressed within the system, therefore the term cannot be classified completely; thus, the extent to which applications can process terms is limited.

6.2 Canonical-Graph Derivation

The canonical graph for a term can be derived from the term's type definition, joined with the type definitions of all the ancestors of the type definition's referent node. This derivation first restricts the referent nodes of each of the type definitions to the type label that is the subject of the derivation. Then, the definitions are joined on identical nodes. If the definition of an ancestor is specialized by a child, then the definition is overridden. The process of recognizing specialization occurs by a combination of restrict, join, and simplify.

We shall illustrate, by example, how canonical-graph derivation is used. To derive the canonical graph for PA-CHEST-X-RAY, we first *copy* the type definition for the term PA-CHEST-X-RAY:

```
[RADIOLOGIC-PROCEDURE: x]-
  (RESULTS-IN)→[REPORT]
  (CREATES)→[IMAGE]-
    (OF-LOCATION)→[CHEST]-
      (HAS-ORIENTATION)→[PA].
```

Second, we *copy* all the type definitions for all the ancestors of PA-CHEST-X-RAY. The set of graphs now includes the definitions for PROCEDURE, RADIOLOGIC-PROCEDURE, and DIAGNOSTIC-RADIOGRAPHY-PROCEDURE:

```
[T:x]→(PERFORMED-ON)→[PATIENT].

[PROCEDURE: x]→(USES)→[IONIZING-RADIATION].

[RADIOLOGIC-PROCEDURE: x]-
    (RESULTS-IN)→[REPORT]
    (CREATES)→[IMAGE],.

[RADIOLOGIC-PROCEDURE: x]-
  (RESULTS-IN)→[REPORT]
  (CREATES)→[IMAGE]-
    (OF-LOCATION)→[CHEST]
    (HAS-ORIENTATION)→[PA],,.
```

Third, we *restrict* all the concept nodes with the referent "*x*" to the type PA-CHEST-X-RAY:

```
[PA-CHEST-X-RAY: x]→(PERFORMED-ON)→[PATIENT].

[PA-CHEST-X-RAY: x]→(USES)→[IONIZING-RADIATION].

[PA-CHEST-X-RAY: x]-
    (RESULTS-IN)→[REPORT]
    (CREATES)→[IMAGE],.

[PA-CHEST-X-RAY: x]-
  (RESULTS-IN)→[REPORT]
  (CREATES)→[IMAGE]-
    (OF-LOCATION)→[CHEST]
    (HAS-ORIENTATION)→[PA],,.
```

Fourth, we *join* the graphs on identical concepts, starting with the nodes with the referent "*x*," resulting in the following graph:

```
[PA-CHEST-X-RAY: x]-
  (PERFORMED-ON)→[PATIENT]
  (USES)→[IONIZING-RADIATION]
  (RESULTS-IN)→[REPORT]
  (CREATES)→[IMAGE]
  (RESULTS-IN)→[REPORT]
  (CREATES)→[IMAGE]-
    (OF-LOCATION)→[CHEST]
    (HAS-ORIENTATION)→[PA],,.
```

Finally, we *extend* this join to be *maximal* by doing a combination of *join* and *simplify*, until all duplicate concepts and relation are removed. This procedure results in the following graph:

```
[PA-CHEST-X-RAY: x]-
  (PERFORMED-ON)→[PATIENT]
  (USES)→[IONIZING RADIATION]
  (RESULTS-IN)→[REPORT]
  (CREATES)→[IMAGE]-
    (OF-LOCATION)→[CHEST]
    (HAS-ORIENTATION)→[PA],,.
```

6.3 Addition of the Data

Using the same steps as we used to create the canonical graph for PA-CHEST-X-RAY, we can derive the following canonical graphs by starting with the type definitions for PATIENT, REPORT, PROVIDER, and LOCALIZED-OBSERVATION presented in Table 1:

```
[PATIENT: x]-
  (HAS-PATIENT-ID)→[123-45-67]
  (HAS-NAME)→[John Doe],.
```

```
[REPORT: x]-
  (CREATES)←[PROVIDER]
  (HAS-OBSERVATIONS)→[OBSERVATION: {*}],.
[PROVIDER: x]-
  (HAS-PROVIDER-ID)→[MD-56789]
  (HAS-NAME)→[Dr. Smith],.
[EFFUSION: x]-
  (LOCATED-IN)→[PLEURAL-CAVITY]-
    (HAS-LATERALITY)→[LEFT],,.
```

Substituting these graphs into the graph for the procedure PA-CHEST-X-RAY yields:

```
[PA-CHEST-X-RAY: x]-
  (PERFORMED-ON)→[PATIENT: a]-
    (HAS-PATIENT-ID)→[123-45-67]
    (HAS-NAME)→[John Doe],
  (USES)→[IONIZING-RADIATION]
  (RESULTS-IN)→[REPORT: b]-
    (CREATES)←[PROVIDER: c]-
      (HAS-PROVIDER-ID)→[MD-56789]
      (HAS-NAME)→[Dr. Smith],
    (HAS-OBSERVATIONS)-
      [EFFUSION: d]-
        (LOCATED IN)→[PLEURAL-CAVITY]-
          (HAS-LATERALITY)→[LEFT],,,
  (CREATES)→[IMAGE]-
      (OF-LOCATION)→[CHEST]
      (HAS-ORIENTATION)→[PA],,.
```

The preceding conceptual graph represents the example in a reproducible manner; however, there are no temporal references describing when the film was taken or when the film was read by the radiologist. In addition, the anatomical model used to represent the pleural cavity and laterality has no formal foundation. To provide a sound foundation, we need to develop detailed models to represent specific topics, such as the temporal and anatomical dimensions of the report. In Section 7, we describe how such a model can be developed for time.

7. Foundational Models

Conceptual graphs emphasize—in an intuitive, visual manner—the logical constructs needed to represent clinical concepts. In Section 6, we defined a small set of relations necessary to represent the structure of a simple radiology report. We can also incorporate the existing models contained in a standard terminology system—such as SNOMED—by using their existing labels and defined relationships to populate our conceptual model. We show how this incorporation might be done in Section 8. The set of concepts and relations in existing coding schemes is limited, and represents only the beginning of the work needed to develop a comprehensive medical terminology.

Implicit in standard lexicons such as SNOMED are a number of *foundational assumptions*. SNOMED, for example, includes in its modifier and linkage axis (the G axis) anatomical concepts, such as *anterior* and *adjacent*, temporal concepts, such as *subacute* and *relapsing*, and measures of uncertainty, such as *possible* and *cannot exclude*. These terms for representing temporal relationships, anatomical relationships, and uncertain relationships reflect underlying models of time, of anatomy, and of uncertainty that users can apply to almost any concept description created out of SNOMED codes. Nonaxial coding schemes, such as ICD-9-CM, also embody such *foundational models*, but the models are reflected only in the distinctions made by the surface-level terms in the lexicon. For example, by making a distinction between the code 410 (*myocardial infarction, acute*) and the code 412 (*myocardial infarction, old*), ICD-9-CM offers a choice of disease codes that reflects its rather simple underlying model of time.

SNOMED offers an advance over lexicons such as ICD-9-CM by separating out and making explicit sets of modifiers for anatomical, temporal, and probabilistic concepts. SNOMED, however, offers only a list of terms without clarifying the relationships among those terms. Our proposal for representation of clinical data requires that we create reusable models of these foundational concepts—models that allow us to encode arbitrary medical descriptions so that the attendant anatomical, temporal, and probabilistic distinctions can be represented in a uniform manner with precise semantics.

In this section, we show how we can create a foundational model to represent *temporal relations*. Time is an essential context for representing all clinical data, since automated methods to manage and interpret patient data require the temporal dimension of those data.[44]

Existing medical-terminology systems model the temporal context of clinical data using the *time-stamping* schemes used in clinical databases; these databases simply provide an instant time stamp to record the date and time at which the datum was observed. Such schemes are not able to represent temporal information commonly found in clinical reports. For example, observations such as "left pleural effusion increased over previous examination" require representation of the interval of time between the two examinations. Since current instant-stamping schemes do not readily capture the uncertainty associated with such interval-based observations, we require a new formalism to model the semantics of time references.

To model time with conceptual-graph representations of clinical data, we must define a model of time appropriate for clinical data and a standard syntax for time references.

7.1 Model of Time

Several methods of time have been proposed for the computer-based representation and manipulation of clinical data. Kahn[45] has differentiated these temporal models of clinical data by two different aspects: (1) whether time consists of a *continuous flow* or a set of *discernible values*, and (2) whether the primitive element is the *time point* or the *interval*. Since modern database technology requires discrete time-stamp values for efficient and reliable storage, indexing, and processing, we chose a metric model of time that time stamps clinical data with explicit calendar-date or clock-time values. We use two special symbols, $-\infty$ and $+\infty$, to represent the lower and upper bounds, respectively, of all other time points. We require that all time stamps have an interval-based representation, thus allowing representation of incomplete temporal references, such as "chest x-ray taken 9/12/93" or "the problem lasted for 2 to 3 days," using these intervals. The necessity of modeling such temporal uncertainty for clinical data has been recognized by Console and colleagues,[46] by Das and colleagues,[47] and by Kohane and Haimowitz.[48]

Our proposed temporal model can represent uniformly three general types of time-stamped clinical data: *instant based, interval based,* and *time invariant.* Most discrete time models use a single time stamp with varying granularities. The coarser the granularity of the time stamp, the greater the temporal uncertainty associated with the datum. Our temporal model uses time stamps with a single granularity, and represents uncertainty with two time points that represent the lower and upper bounds of the closed *interval of uncertainty* (IOU) during which the datum occurred. Our representation of instant-based data offers two distinct advantages over representations that use different granularities. First, we are not limited to storing a period of uncertainty for instant-based data that corresponds to real-world calendar or clock units. For example, we can store the interval of uncertainty for the previous examination as 36 hours, instead of simply 1 day. Second, when comparing instant-based data with different intervals of uncertainty, our time-stamping scheme does not require time-point approximations, which temporal models storing various granularities need to make.

7.2 Standard Syntax for Time

To incorporate our temporal model into a conceptual-graph representation of clinical data, we must create a standardized syntax for temporal references. We specify conceptual terms for time stamps, and conceptual links to joint clinical-finding terms to their temporal dimension.

We provide the following canonical graph for the IOU representation for *instant-based* time stamps in our temporal model:

```
[TIME-STAMP: x]-
  (POINT-START)→[date time]
  (POINT-END)→[date time],.
```

The date–time box could be filled with any representation for time points; we choose the format MM/DD/YY HHMM. For example, the conceptual-graph representation of the time stamp of a chest x-ray image taken 9/12/93 would be:

```
[TIME-STAMP: x]-
  (POINT-START)→[9/12/93 0000]
  (POINT-END)→[9/12/93 2359],.
```

This graph states formally that the previous examination occurred at some time after midnight 9/11/93 and before midnight 9/12/93. If we wish to represent a second time stamp of another x-ray image taken sometime in January 1993, we can do so as:

```
[TIME-STAMP: x]-
  (POINT-START)→[1/1/93 0000]
  (POINT-END)→[1/31/93 2359],.
```

To represent the time elapsed between two different examinations, we must first establish the representation of time stamps for *interval-based* information. As we described in our temporal model, our time-stamping scheme can capture the uncertainty associated with interval-based data by a pair of intervals, which represent the start and end IOUs. We can thus model interval-based data using the following canonical graph:

```
[INTERVAL: x]-
  (INTERVAL-START)→[TIME-STAMP]
  (INTERVAL-END)→[TIME-STAMP],.
```

By substituting the value of the time stamps corresponding to the time stamps of the previous and the current films, we would obtain the following interval of comparison:

```
[INTERVAL: x]-
  (INTERVAL-START)→[TIME-STAMP: y]-
    (POINT-START)→[1/1/93 0000]
    (POINT-END)→[1/31/93 2359],
  (INTERVAL-END)→[TIME-STAMP: z]-
    (POINT-START)→[9/12/93 0000]
    (POINT-END)→[9/12/93 2359],,.
```

Now that we have developed a general method for representing time, we can modify our initial type definitions to include time. To represent the time that a procedure was performed, we must add the following definition to those given in Table 1:

```
type TIME-STAMP(x) is
  [T: x]-
    (POINT-START)→[date time]
    (POINT-END)→[date time],.
```

Second, we must change the definition of procedure as follows:

```
type PROCEDURE(x) is
  [T: x]-
    (PERFORMED-ON)→[PATIENT]
    (PERFORMED-AT-TIME)→[TIME-STAMP],.
```

If we use the preceding type definitions, and specify that the x-ray examination was performed on 9/12/93, the graph that represents our simple example in Section 6.3 becomes:

```
[PA-CHEST-X-RAY: x]-
  (PERFORMED-ON)→[PATIENT: a]-
    (HAS-PATIENT-ID)→[123-45-67]
    (HAS-NAME)→[John Doe],
    (PERFORMED-AT-TIME)→[TIME-STAMP: e]-
      (POINT-START)→[9/12/93 0000]
      (POINT-END)→[9/12/93 2359],
  (USES)→[IONIZING-RADIATION]
  (RESULTS-IN)→[REPORT: b]-
    (CREATES)←[PROVIDER: c]-
      (HAS-PROVIDER-ID)→[MD-56789]
      (HAS-NAME)→[Dr. Smith],
    (HAS-OBSERVATIONS)-
      [EFFUSION: d]-
        (LOCATED IN)→[PLEURAL-CAVITY]-
          (HAS-LATERALITY)→[LEFT],,,
  (CREATES)→[IMAGE]-
      (OF-LOCATION)→[CHEST]
      (HAS-ORIENTATION)→[PA],,.
```

In a way similar to how procedures are changed to represent time, findings themselves can be changed to represent time including interval-based observations, instant-based observations, and time-relative observations.

We have described how a logical foundation can be developed for representation of clinical data. We have described how a simple radiology report, including the report's temporal references, can be represented. We have not described how a sound model for representing the other detailed foundational models relevant to this report should be represented. Detailed foundational models, such as a model able to represent the anatomy of the pleural cavities, will need to be developed.

SNOMED has a set of concepts and relations that can provide a basis to represent such anatomical concepts. In Section 8, we shall describe how SNOMED might be enhanced, in an evolutionary way, to fit within our framework.

8. SNOMED International

We have described previously the desirable features and limitations of SNOMED.[49] Perhaps the most important feature of SNOMED is its relative *domain completeness*. SNOMED has over 120,000 terms that represent concepts commonly found in clinical medicine.

8.1 Conceptual Graphs and SNOMED

Because of SNOMED's size and the numerous person-years that have already been invested in its development, and the existing infrastructure and implementations of SNOMED-based systems, it makes sense to enhance SNOMED rather than to start from scratch to develop a new standard. SNOMED has been evaluated for its ability to represent nursing concepts in the patient record, and was found more complete than existing nursing-specific classifications.[50] In addition, the Board of Directors of the American Medical Informatics Association has published a position paper in which it suggests SNOMED as a standard to represent diagnoses, symptoms and findings, microbes and etiologies, and anatomic locations.[51]

SNOMED has no standardized syntax to allow construction of statements with complex interrelationships. For example, to encode that a patient had "left pleural effusion," SNOMED provides all the needed functional and topographic terms, as shown in Table 2. However, having all the terms is not in itself sufficient. SNOMED provides no standard syntax to combine the terms into a statement.

Before conceptual-graph formalisms can be applied to SNOMED, an initial set of graphs must be ap-

Table 2 ■

SNOMED Codes and Their Corresponding Terms

SNOMED Codes	Terms
A-81000	Ionizing radiation
D2-80100	Pleural effusion
G-A100	Right
G-A101	Left
G-C220	With laterality
P0-00000	Procedure, NOS
P5-00000	Radiologic procedure, NOS
P5-00010	Diagnostic radiologic examination, NOS
P5-20040	Diagnostic radiography of chest, PA

NOS = not otherwise specified; PA = posterior–anterior.

proved officially or *canonized*. Once a set of canonical graphs has been developed, additional graphs can be derived from formation rules. Using these canonical graphs and formation rules, we can create a standard syntax for SNOMED codes, and can enhance the overall system to include important concepts not included in SNOMED, such as the model of time that we described in Section 7.

8.2 Migration Path for SNOMED

The first step in recasting SNOMED to fit within a logical framework is to create an Aristotelian type hierarchy from the existing codes, taking advantage of SNOMED's hierarchy and cross-reference links. Initially, most of the SNOMED codes will have to be declared as *primitive*, as we described in Section 6.1. Once this initial Aristotelian type hierarchy is created, the differentia of each term can be enhanced in an evolutionary way until only a minimum set of terms are primitives.

Initially, we can use SNOMED's existing cross-reference links to develop these enhancements. Each SNOMED term has an associated cross-reference field where one or more related terms may be linked by including the relevant term code in this field. Figure 1 illustrates these cross-references for the terms "disease of respiratory system," "disease of pleura, NOS,"† and "pleural effusion, NOS." The type of cross-reference link often can be inferred based on the terms that are cross-referenced. For example, a term from the *diagnosis* axis may be cross-referenced to terms in the *morphologic* axis that represent the morphologic changes associated with that particular disease. It can be inferred that the linkage is that the change represented by the morphology term is CAUSED-BY the disease represented by the disease term.

†NOS is an abbreviation for "not otherwise specified."

We can use these cross-reference links as a first approximation to create the differentia for each term. We can assume that a probable relationship between a morphology term and a disease term is that the morphological change is CAUSED-BY the disease, and that a probable relationship between a disease and a location is that the disease is LOCATED-IN the location. We can use these probable relations, the existing SNOMED cross-references for the term PLEURAL-EFFUSION, and the hierarchical parent for this term (hierarchical parent and cross-references are shown in Fig. 1) to suggest the following definition for pleural effusion:

type PLEURAL-EFFUSION(x) *is*
  [PLEURAL-DISEASE: x]-
    (CAUSED-BY)→[EFFUSION]
    (LOCATED-IN)→[PLEURAL-CAVITY],.

We do not believe that this definition is the correct definition for PLEURAL-EFFUSION, however. It would be a simple matter for a developer to rearrange the definition to a more correct one by using a "smart" graphical editing environment:

type PLEURAL-EFFUSION(x) *is*
  [EFFUSION: x]-
    (CAUSED-BY)→[DISEASE]
    (LOCATED-IN)→[PLEURAL-CAVITY],.

Cimino[21] has discussed other desirable features of knowledge-based editing environments that can be used to maintain and enhance a controlled clinical terminology. He describes how, by using a knowledge-based approach, the quality of Columbia–Presbyterian Medical Center's laboratory and pharmacy terminology has improved, and the maintenance efforts have been reduced by 90%.

9. Discussion

We have presented a framework for formalizing the representation of medical concepts, using conceptual graphs as a standard notation for logic. The conceptual-graph representations can be applied broadly to several different practical applications and research areas, providing a unifying framework. There are, however, important limitations to this approach.

9.1 Logic-based Approach

In this paper, we have suggested the use of conceptual graphs for representation of the logical relationships among the terms used to represent medical concepts. As mentioned in Section 5, conceptual-graph representations have a number of desirable features, including the ease with which they may be read by humans and parsed by machines. We emphasize, however, that conceptual graphs constitute but one syntax for writing expressions in logic. Although we believe that it is essential to represent medical-concept descriptions using a form of logic, the conceptual-graph notation is only one of several equivalent formalisms. We might just as easily represent medical concepts using either the common algebraic notation for logical propositions, or any one of several popular knowledge-representation languages.

Fortunately, researchers in the computer-science community are working to overcome the problems of multiple competing syntaxes for knowledge representation. Members of the ARPA-sponsored knowledge-sharing project[52] have created the Knowledge Interchange Format (KIF)[31] to serve as a means to interconvert knowledge bases created in a variety of representations. In the same way that Microsoft's rich-text format (RTF) allows users of different word-processing programs to translate documents from the format used by one proprietary program to another, KIF permits concept descriptions created in a variety of languages (including conceptual graphs) to be interchanged. Thus, KIF allows medical concepts described using a combination of SNOMED and conceptual graphs to be translated directly into a knowledge-representation system such as KL-ONE. Similarly, medical concepts represented by an object-oriented language such as MODEL (used to represent knowledge in our laboratory's PROTÉGÉ-II system[53]) can be translated via KIF into conceptual-graph notation. Thus, conceptual graphs provide a convenient mechanism for writing and reviewing standardized descriptions of medical concepts, but do not need to be the vehicle for representing those concepts within end-user applications.

This flexibility is critical. Whereas conceptual graphs capture the full semantics of first-order logic, and thus offer enormous expressive power, proving theorems in first-order logic is, in the worst case, computationally intractable. Logic, while supporting the rich semantics that we seek to represent nuances in clinical findings, is in its full form too computationally expensive for most practical applications. Developers of knowledge-representation systems consequently make deliberate decisions about which elements of first-order logic they may safely omit to improve run-time performance.[54] For example, common object-oriented languages do not allow the expression of concepts such as negation or disjunction. No practical knowledge-representation system is perfect; each embodies a particular set of trade-offs. Our approach thus is to use complete first-order logic (as reflected in the syntax of conceptual graphs) for *canonical* representations of clinical concepts, and to translate those representations into the particular knowledge-representation systems that are most suitable for individual applications.

9.2 Foundational Models

Because logic is topic neutral, the choice of conceptual graphs as a method for medical-concept representation does not imply any particular model of the medical domain. Indeed, to capture the rich semantics of clinical data and to formalize clinical concepts and their relationships, we must rely on foundational models in medicine. Existing medical nomenclature systems (such as SNOMED) provide a set of labels and defined relationships that can serve as the basis for a comprehensive terminology; yet, the foundational models embodied in such nomenclature are either limited in their scope or encoded implicitly. Therefore, developers of logic-based conceptual models face the challenge of formalizing such foundational models.

In this paper, we have presented in detail one simple but essential foundational model: a representation for the temporal dimension of clinical data. We have derived our model of time references from our research on a uniform temporal data model for clinical databases.[47] Our temporal representation shares the same formal underpinnings as Shoham's formal interval-based logic[55]; both models rely on time points as primitive temporal elements, but permit temporal assertions only over intervals. The main obstacle to creating a foundational model of time references, however, is representing the temporal indeterminacy of clinical observations. By measuring within a bounded IOU the instant at which an observation was valid, we can define an interval-based logic that captures the semantics of clinical data time stamped at different granularities. Our temporal model thus can avoid the information loss that occurs when providers enter time stamps that are more or less precise than can be justified by their knowledge.

A foundational model of the time stamps of clinical findings is only one of many aspects of time that a provider may want to encode. Our simple temporal model, for example, does not assume any particular density distribution of the instance's occurrence during the IOU. Consequently, a provider could not use our proposed temporal representation to model the occurrence of a datum as more probable in the first 6 hours of a given date than in the last 18 hours. If a principled model of uncertainty was added to our conceptual-graph representation, we could assert various types of density distributions (such as uniform distributions) with IOUs. In addition, the relations we have described here are not sufficient to specify the ordering of events when only their sequences are known. We should thus anticipate the need to merge different foundational models (such as time, uncertainty, and anatomy) to develop a comprehensive conceptual representation of clinical observations. We should further expect a range in the complexity of the foundational models from simple time references in radiology reports to spatiotemporal changes in an embryo.

Determining whether our foundational model of time references (or any other foundational model) is sufficiently expressive for encoding clinical findings requires empirical evaluation. Like our uniform temporal data model for clinical databases, foundational models that we choose to incorporate into our logic-based representation may already be implemented in data-management systems. In such cases, the limitations of the foundational models may be well understood. In other cases (such as the model of anatomy in GALEN[56]), we will not be able to evaluate the appropriateness of such models until we use the models in data acquisition or in data encoding. Therefore, to establish the suite of foundational models that is necessary for a logic-based clinical representation, we need (1) to incorporate current or developing formal models of specific topics whenever possible, (2) to extend or combine these models as necessary, and (3) to instantiate the concepts and relations in these models with standardized codes from existing medical nomenclatures.

9.3 Evolutionary Enhancement

The evolution of a standard representation for clinical data will be a challenging task to manage. This evolution will have to be guided by empirical studies of real-world applications. These applications may require local enhancements to function properly. If these enhancements are useful, they should eventually be incorporated into the standard so that other applications can use them. The technical and managerial aspects of how this process should be managed are poorly understood.

Tuttle and colleagues[57] have described how developers incur a penalty for creating local enhancements. This penalty occurs when developers try to synchronize their enhancements with similar changes that may be incorporated into the Unified Medical Language System.[58] If participation by more than one institution in the development of the standard is desired, then we must find a way to incorporate local enhancements from multiple sites, without penalizing institutions for their efforts.

We have been looking at how the concurrency-control models of advanced-database applications[59] might be used to manage local enhancements in a way that eliminates the penalty created by local enhancements. If we are able to develop a proper concurrency-control model for managing evolutionary enhancement, national participation in the evolution of a clinical data-representation standard may become a reality. Without such a concurrency-control model, the problems of managing local enhancements will discourage participation by all but the most determined developers.

9.4 Conclusion

This paper described the beginning stages of work on our medical-terminology systems. Significant challenges remain for the development of sound foundational models for the various aspects of clinical care. If such models can be developed, they could ease the task of developing decision-support and epidemiological tools for use in complex patient situations.

The models that we propose will have to evolve over time. Because our underlying framework is based in logic, we are confident that the representation is both general enough and sufficiently expressive to represent the foundational models as they are developed. However, we still must prove that such a model can be practically implemented. It is crucial that we progress to actual implementations so that the practicality of this approach can be evaluated formally.

Finally, the ultimate success or failure of a clinical data-representation standard, such as the one that we propose, is not based solely on the technical merits of the underlying representation. Important problems of how to manage the development process, how to meet the needs of specific applications, how to coordinate evolutionary enhancements, and how to develop a political consensus have yet to be solved.

The authors thank Roger Côté for allowing us to use a preliminary version of SNOMED International, John Sowa for generously giving us his time and insight during many discussions of this work, and Lyn Dupré for her editorial assistance.

References ■

1. Campbell KE, Wieckert K, Fagan LM, Musen MA. A computer-based tool for generation of progress notes. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on

Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:284–8.
2. Naeymi-Rad F, Almeida F, Trace D. IMR-entry (intelligent medical record-entry). In: Frisse ME, ed. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care, Baltimore, MD. New York: McGraw-Hill, 1992:783–4.
3. Rector AL, Nowlan WA, Kay S. Foundations for an electronic medical record. Methods Inf Med. 1991;30(3):179–86.
4. Bell DS, Greenes RA, Doubilet P. Form-based clinical input from a structured vocabulary: initial application in ultrasound reporting. In: Frisse ME, ed. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care, Baltimore, MD. New York: McGraw-Hill, 1992:789–90.
5. Benoit RG, Canfield K, Teitelbaum S, et al. Frame-based clinical data entry: design issues for physician acceptance. In: Lun KC, Degoulet P, Piemme TE, Rienhoff O, eds. Proceedings of MEDINFO 92, Geneva, Switzerland. North Holland: Elsevier Science Publishers, 1992:1280–5.
6. Kuhn K, Zemmler MR, Heinlein DR. Structured data collection and knowledge-based user guidance for abdominal ultrasound reporting. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:311–5.
7. Baud R, Lovis C, Alpay L, et al. Modelling for natural language understanding. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993: 289–3.
8. Lenert LA, Tovar M. Automated linkage of free-text descriptions of patients with a practice guideline. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:274–8.
9. Sager N, Lynman M, Tick LJ, Ngô TN, Bucknall CE. Natural language processing of asthma discharge summaries for the monitoring of patient care. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:265–8.
10. Lamiell JM, Zbigniew MW, Isaacks J. Computer auditing of surgical operative reports written in English. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:269–73.
11. United States General Accounting Office. Standards for Automated Medical Records. Report to Congress GAO/IMTEC-93-17, 1993.
12. Audet A-M, Scott HD. The Uniform Clinical Data Set: An evaluation of the proposed national database for Medicare's quality review program. Ann Intern Med. 1993;119(12): 1209–13.
13. Côté RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L, eds. The Systematized Nomenclature of Medicine: SNOMED International. Northfield, IL: College of American Pathologists, 1993.
14. World Health Organization. International Classification of Diseases: Ninth Revision. Geneva: World Health Organization, 1977.
15. Smith MW. Hospital discharge diagnoses: how accurate are they and their international classification of diseases (ICD) codes? N Z Med J. 1989;102(876):507–8.
16. Hsia DC, Ahern CA, Ritchie BP, Moscoe LM, Krushat WM. Medicare reimbursement accuracy under the prospective payment system, 1985 to 1988. JAMA. 1992;268(7):896–9.
17. van Walraven C, Wang B, Ugnat AM, Naylor CD. False-positive coding for acute myocardial infarction on hospital discharge records: chart audit results from a tertiary centre. Can J Cardiol. 1990;6(9):383–6.
18. National Library of Medicine. Medical Subject Headings. NLM, NTIS NLM-MED-92-01, 1992.
19. American Psychiatric Association Task Force on Nomenclature and Statistics, Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III). Washington, DC: American Psychiatric Association, 1980.
20. Hutchins RM, ed. The works of Aristotle: volume I. In: Hutchins RM, ed. Great Books of the Western World. Chicago: Encyclopedia Britannica, 1952:726.
21. Cimino JJ, Clayton PD, Hriscsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Informatics Assoc. 1994;1(1): 35–50.
22. Masarie FE, Miller RA, Bouhaddou O, Nunzia BG, Warner HR. An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies. Comput Biomed Res. 1991;24(4):379–400.
23. Rector AL, Nowlan WA, Kay S, Goble CA, Howkins TJ. A framework for modeling the electronic medical record. Methods Inf Med. 1993;32(2):109–19.
24. Spencer-Smith R. Logic and Prolog. Hertfordshire, Great Britain: Harvester Wheatsheaf, 1991.
25. Brachman RJ, Schmolze JG. An overview of the KL-ONE knowledge representation system. Cogn Sci. 1985;9(2):171–202.
26. Sowa JF. Conceptual Structures. Reading, MA: Addison-Wesley, 1984.
27. Hayes PJ. In defense of logic. In: Reddy R, ed. International Joint Conference on Artificial Intelligence-1977, Cambridge, Massachusetts. Pittsburgh: Carnegie-Mellon University, 1977:559–65.
28. Brachman RJ, Fikes RE, Levesque HJ. Krypton: a functional approach to knowledge representation. Computer. 1983;16(10): 67–73.
29. Moser MG. An overview of NIKL: the new implementation of KL-ONE. Cambridge, MA: Bolt, Beranek and Newman, Technical Report 5421, 1983.
30. Brachman RJ, McGuinness DL, Patel-Schneider PF, Resnick LA, Borgida A. Living with a classic: when and how to use a KL-ONE-like language. In: Sowa JF, ed. Principles of Semantic Networks: Explorations in the Representation of Knowledge. San Mateo, CA: Maurgan Kaufman Publishers, 1991:401–56.
31. Genesereth MR, Fikes RE. Knowledge Interchange Format, Version 3.0 Reference Manual. Technical Report Logic-92-1. Stanford, CA: Stanford University, 1992.
32. Perez S, Sarris A. Information Resource Dictionary System (IRDS) Conceptual Schema (CS). American National Standards Institute X3H4 IRDS and the International Organization for Standardization, ANSI X3H4/92-003, ISO/IEC JTC1/SC21 N7486, 1993.
33. Peirce CS. Manuscripts on existential graphs. In: Burks AW, ed. Collected Papers of Charles Sanders Peirce, vol. 4. Cambridge, MA: Harvard University Press, 1897–1906:320–410.
34. Bernauer J. Conceptual graphs as an operational model for descriptive findings. In: Clayton PD, ed. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1991:214–8.
35. Schröder M. Knowledge based analysis of radiological reports using conceptual graphs. In: Pfeiffer HD, ed. Proceedings of the Seventh Annual Workshop on Conceptual Graphs, Las Cruces, New Mexico. Berlin: Springer-Verlag, 1992:213–22.
36. Rassinoux AM, Baud RH, Scherrer JR. Conceptual graphs model

extension for knowledge representation of medical texts. In: Lun KC, Degoulet P, Piemme TE, Rienhoff O, eds. Proceedings of MEDINFO 92, Geneva, Switzerland. North Holland: Elsevier Science Publishers, 1992:1368–74.
37. Schröder M. Knowledge-based processing of medical language: a language engineering approach. In: Ohlbach H-J, ed. Proceedings of GWAI-92: German Conference on Artificial Intelligence. Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag, 1993:190–9.
38. Johnson SB, Aguirre A, Peng P, Cimino J. Interpreting natural language queries using the UMLS. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:294–8.
39. Peng P, Aguirre A, Johnson SB, Cimino JJ. Generating MEDLINE search strategies using a librarian knowledge-based system. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:596–600.
40. Joubert M, Fieschi M, Robert JJ. A conceptual model for information retrieval with UMLS. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:715–9.
41. Friedman C, Cimino JJ, Johnson SB. A conceptual model for clinical radiology reports. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:829–33.
42. Levinson R. Pattern associativity and the retrieval of semantic networks. Comput Math Applications J. 1992;23(2):573–600.
43. Ellis G. Efficient retrieval from hierarchies of objects using lattice operations. In: Mineau G, Moulin B, Sowa J, eds. Lecture Notes in Artificial Intelligence 699: Conceptual Graphs for Knowledge Representation. Berlin: Springer-Verlag, 1993:274–93.
44. Shortliffe EH, Barnett OG. Medical data: their acquisition, storage, and use. In: Shortliffe EH, Perreault LE, eds. Medical Informatics: Computer Applications in Health Care. Reading, MA: Addison-Wesley, 1990:39.
45. Kahn MG. Modeling time in medical decision support programs. Med Decis Making. 1991;11(4):249–64.
46. Console L, Furno A, Torasso P. Dealing with time in diagnostic reasoning based on causal models. In: Ras Z, Saitta L, eds. Methodologies for Intelligent Systems, vol. 3. Amsterdam: North-Holland Publishing Co, 1988.
47. Das AK, Tu SW, Purcell GP, Musen MA. An extended SQL for temporal data management in clinical decision-support systems. In: Frisse ME, ed. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care, Baltimore, MD. New York: McGraw-Hill, 1992:128–32.
48. Kohane IS, Haimowitz IJ. Hypothesis-driven data abstraction with trend templates. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:444–8.
49. Campbell KE, Musen MA. Representation of clinical data using SNOMED III and conceptual graphs. In: Frisse ME, ed. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care, Baltimore, MD. New York: McGraw-Hill, 1992:354–8.
50. Henry SB, Holzemer WL, Reilly CA, Campbell KE. Terms used by nurses to describe patient problems: can SNOMED III represent nursing concepts in the patient record? J Am Med Informatics Assoc. 1994;1(1):61–74.
51. Board of Directors of the American Medical Informatics Association. Standards for medical identifiers, codes, and messages needed to create an efficient computer-stored medical record. J Am Med Informatics Assoc. 1994;1(1):1–7.
52. Neches R, Fikes RE, Finin T, et al. Enabling technology for knowledge sharing. AI Magazine. 1991;12(3):16–36.
53. Musen MA. Dimensions of knowledge sharing and reuse. Comput Biomed Res. 1992;25:435–67.
54. Levesque HJ, Brachman RJ. A fundamental tradeoff in knowledge representation and reasoning (revised version). In: Brachman RJ, Levesque HJ, eds. Readings in Knowledge Representation. San Francisco: Morgan Kaufman, 1985:42–70.
55. Shoham Y. Temporal logic in AI: semantical and ontological considerations. Artif Intelligence. 1987;33(1):89–104.
56. Rector AL, Nowlan WA, Glowinski A. Goals for concept representation in the GALEN project. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1993:414–8.
57. Tuttle MS, Sherertz DD, Erlbaum MS, et al. Adding your terms and relationships to the UMLS metathesaurus. In: Clayton PD, ed. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care, Washington, DC. New York: McGraw-Hill, 1991:219–23.
58. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med. 1993;32:281–91.
59. Barghouti NS, Kaiser GE. Concurrency control in advanced database applications. ACM Computing Surveys. 1991;23(3):269–317.