Research Paper ■

An Event Model of Medical Information Representation

STANLEY M. HUFF, MD, ROBERTO A. ROCHA, MD, BRUCE E. BRAY, MD, HOMER R. WARNER, MD, PHD, PETER J. HAUG, MD

Abstract **Objective:** Develop a model for structured and encoded representation of medical information that supports human review, decision support applications, ad hoc queries, statistical analysis, and natural-language processing.

Design: A medical information representation model was developed from manual and semiautomated analysis of patient data. The key assumption of the model is that medical information can be represented as a series of linked events. The event representation has two main components. The first component is a frame or template definition that specifies the attributes of the event. The second component is a structured vocabulary, the terms of which are taken as the values of the slots in the event template structure. Individual event instances are linked by specific named relationships.

Results: The proposed model was used to represent a chest-radiograph report.

Conclusions: The event model of medical information representation provides a mechanism for formal definition of the logical structure of medical data and allows explicit time-oriented and associative relationships between event instances.

■ **J Am Med Informatics Assoc.** 1995;2:116–134.

One of the fundamental challenges in medical informatics is the development of a consistent and comprehensive scheme for representing clinical data. A common representation scheme is an essential foundation for exchanging data between heterogeneous systems, pooling data across institutions for clinical and outcomes research, sharing medical decision support logic, and sharing patient care applications. Issues related to the development of a scheme for medical-concept representation have been discussed by the Canon Group,[1] and individuals participating in the Canon Group and other research groups have recently presented specific models for medical-concept representation.[2–7] In conjunction with these efforts, we have been working on an event-based model of medical information representation, which is described in this article. Before we describe the model, we describe the context in which the model was developed and give a brief summary of our expectations and requirements for the model.

Affiliations of the authors: Department of Medical Informatics, University of Utah College of Medicine (SMH, RAR, BEB, HRW, PJH), and Medical Informatics (SMH, PJH), Intermountain Health Care, Salt Lake City, UT.

Dr. Rocha is supported by a scholarship from the National Council for Scientific and Technological Development (CNPq), Secretary for Science and Technology, Brazil.

Correspondence and reprints: Stanley M. Huff, MD, Department of Medical Informatics, Primary Children's Medical Center, 100 North Medical Drive, Salt Lake City, UT 84113. e-mail: smhuff@cc.utah.edu

Received for publication: 8/09/94; accepted for publication: 11/03/94.

Background

The HELP[8–11] computerized hospital information system has been in constant development at LDS Hospital for over 20 years. One tenet of that development has been the concept of a comprehensive centralized "computable" patient database. By computable we mean that data in the database are represented internally by codes, and that the structure of the database allows medical decision support processes to evaluate rules using the coded data. The hierarchical coding scheme used in the HELP system is called PTXT[10] (Pointer-to-Text). The types of decision support functions that are currently supported include critiquing of blood bank orders,[12,13] pharmacy and laboratory alerting,[14–16] nosocomial infection monitoring,[17] execution of adult respiratory distress syndrome (ARDS) protocols,[18] detection of adverse drug events,[19] monitoring appropriateness of antibiotic use,[20] and utilization review.[21]

The centralized patient database is usually the primary source of data for direct patient care, and applications for collecting patient care data are an integral part of the patient care process. Previous work on user interfaces focused on structured input, including fill-in-the-blank forms, pick lists, and selection sets.[22] The input is transformed into codes, and the codes are stored in the database. The strength of a structured user interface is that validation and quality assurance knowledge can be incorporated in the input process to enforce use of uniform vocabulary and to ensure that all essential data are entered.

However, we have seen an increasing need to obtain information for incorporation in the HELP system database by analysis of free-text documents rather than from structured user interfaces. This is because structured user interfaces are often difficult to create and are not always as fast or efficient as other forms of data entry. A typical example is medical imaging. While we continue to pursue even-better mechanisms for structured data entry of the findings from medical imaging studies, reports for all chest radiographs taken at LDS Hospital are currently dictated, transcribed, reviewed, finalized, and then stored as free-text documents in the HELP database. A few years ago one of us (PJH) initiated research into natural-language processing (NLP) of chest radiographs[23,24] under the assumption that important information could be extracted from the free-text documents and placed in the encoded patient database. This work proved successful, and coded data extracted from chest radiographs are now an important contribution to the nosocomial infection monitoring database of the HELP system.[17]

The development of the NLP software had the additional effect of exposing needed enhancements to the structure and content of the PTXT encoding system. First of all, we recognized a need to record lexical information to support parsing of free text. This information includes parts of speech, tense, varying word forms, plural forms, etc. Second, we needed to be able to link synonyms. We use the word synonyms to describe terms that have *exactly* the same meaning, which is more restrictive than some definitions that would call similar but not identical terms synonyms. PTXT currently has the ability to link synonyms by using one term as a pointer (keyword) to another, but this is not sufficient for NLP needs because the keywords are also used for representing nonsynonymous relationships. Third, we recognized a need to add a more formal model to describe the relationship between complex medical phrases and the individual terms from which they are composed. Fourth, though negation is handled reasonably well in some parts of the HELP system, we saw a need for a more general strategy. We needed the ability to explicitly state that findings were not present without creating thousands of new terms with the word "not" as their only distinguishing characteristic. We also recognized that these new features could be supported only if we added definitions and usage notes to our vocabulary, so that the meanings of the words would be explicit and would not change with time.

Besides support of NLP, we wanted to have a data representation model that would support data exchange and knowledge exchange between heterogeneous clinical information systems. Recent studies have indicated that a common data model is essential to knowledge exchange if laborious reworking of the decision logic is to be avoided.[25]

Another assumption is that coded data in our system are the primary medical record. That is, the coded database must be of sufficient detail and accuracy to serve as the primary basis for diagnostic and therapeutic decision making for the patient. This holds true whether the decision process is computerized or occurs in a clinician's brain.

Finally, our vision of the role the model plays in the collection and use of medical information is shown in Figure 1. The objective is to create an accurate

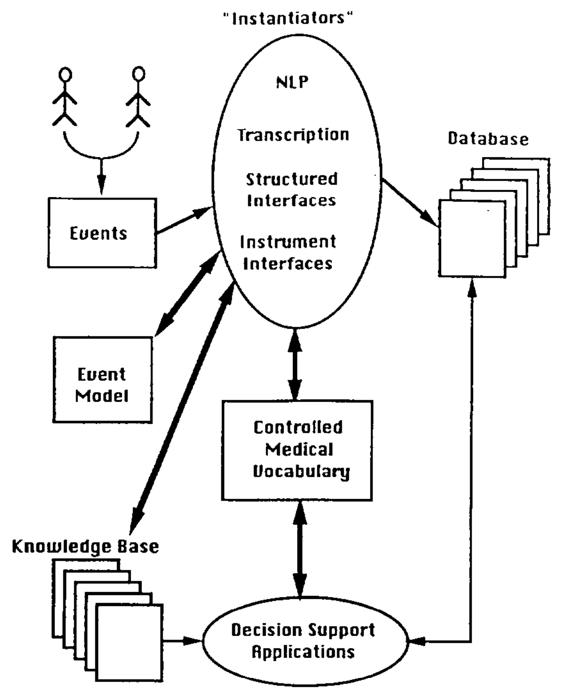

Figure 1 Interaction of the event model with applications, a clinical database, and decision support processes.

coded database representation of any event that can occur in the real world. As noted above, the methods used to record the clinical events vary and include use of structured user interfaces, processing data from automated instruments, and processing user-entered free text. Regardless of the particular method used to record events, the process is guided by the model of information representation. The model also serves as a guide to user and decision support applications that access and analyze the recorded information.

Model Development

The model was derived by manual and semiautomated analysis of patient data. The process was informal and consisted of 1) looking for patterns in data in paper medical records, 2) using Unix word processing tools to analyze the structure of dictated medical reports, and 3) examining the structure of coded data in the HELP system. All types of medical data were examined, including dictated history and physical examinations, laboratory data, admit diagnoses, problem lists, anatomic pathology reports, echocardiography reports, and dictated radiology reports. Chest radiology was chosen as the focus of our current development because it has all the semantic complexity of other domains in medicine, but the set of concepts that are needed to represent the findings is small enough to allow rapid prototyping. Another reason for choosing chest radiology was that it was chosen as the focus of work by the Canon Group. Using the same domain for our own work made collaboration and comparison much easier. Thus, the examples of the event model presented in this article are from the area of chest radiology, but our model was designed as a very general data representation scheme. We have used the event model extensively outside the domain of radiology, and brief descriptions of projects that have used the model are provided later in this article.

The model considers the real world to consist of objects (or entities), where an object is "any part of the perceivable or conceivable world."[26] Real-world objects involved with medicine include individual people such as Steve Roper and Jack Armstrong, institutions such as LDS Hospital and Primary Children's Medical Center, and physical objects such as a thermometer, a blood pressure cuff, or a tablet of penicillin. Objects interact with other objects and can be associated with other objects by relationships. Thus, Jack Armstrong (a patient) is admitted to LDS Hospital (a hospital) by Steve Roper (a physician). In our model, objects are described by their particular attributes or characteristics. For example, most people have a name, a home address, a father, and a mother. Each medication given to a patient has a name, a dose, a route of administration, and a date and time of administration. Objects that have characteristics in common and serve a common function are called "object types." Important medical object types include patients, nurses, physicians, medications, beds, rooms, etc. When two or more objects interact in the real world, an "event" is said to have occurred. Events, like objects, can be classified by their particular attributes. A set of events that have common characteristics can be called "event types."

Events are the central theme of our model. Events are of two major types. One type of event pertains to general activities such as ordering floor stock medications, while the second type of event is directly related to a specific patient. Patient-oriented events are named actions or states of being that exist in relationship to a patient. The following are examples of patient-related events:

1. Placing an order for a medication
2. The birth of a child
3. Performing a surgical procedure
4. The review of the patient's chart by a physician
5. Posting of a charge to a patient's bill
6. The occurrence of chest pain
7. Ordering a chest radiograph
8. Reading and interpreting a chest radiograph
9. Dictating a report for a chest radiograph
10. Transcribing the report of a chest radiograph
11. Storing a transcribed report in the clinical database

Events are described by their attributes or characteristics. For example, the birth of a child might have as attributes the mother's name, the attending physician's name, the sex of the child, the weight of the child, the 5- and 10-minute Apgar scores, and the date and time of birth. Since patient care events take place at a specific time and in relationship to a specific patient, two implied attributes of all patient events are the time the event occurred and a patient identifier. The temporal and associative aspects of this model are similar to those proposed by Rector et al.[6,27] In many situations, the time that an event takes place and the time that the event is recorded in the system are different, so a third common attribute of most events is the time that a record of the event was stored in the database. In situations where the date and time are not known exactly, the time of occurrence can be represented using the techniques described by Campbell et al.[2]

The model can be further understood by examining the four instantiated chest radiograph events shown in Tables 1 through 4. In this example, a chest radiograph was obtained on 10-23-91 and a radiologist read the film and dictated a report. The four event instances were created to represent the information contained in the following two sentences: PA view is compared to the previous examination dated 10-22-91. Surgical clips are again seen along the right mediastinum and right hilar region.

Each event is represented as a frame, where the slots in the frame are typical name–value pairs. The attribute identifiers do not run consecutively because some attributes that can optionally exist do not exist in this example. The full text of the report and its event instantiations are shown in the appendix.

The first event instance (Table 1) records that a chest x-ray was obtained. All appropriate attributes related to that event are noted. From the text, we know that the current examination has been compared with a previous examination that was done on 10-22-91. The second procedure instance (Table 2) represents the information we know about the previous examination. This event instance is short because all we know about the previous examination is the day that it was done. A "compared to" relationship has been created between the two procedures by use of the "semantic link" construct. Note that in a normal inpatient environment, the previous examination would have already existed in the database and the "compared to" link would have been created to the preexisting record.

Table 1 ■

Procedure Event Instance Created to Represent Information in the Sentence *PA view is compared to the previous examination dated 10-22-91**

Label	Attribute	Value
a.1	Patient Identifier	44312897
a.2	Event Family	diagnostic procedures
a.3	Event Template ID	003
a.4	Event Instance ID	1001
a.5	Event Date and Time	10-23-91.14:25
a.6	Event Store Date and Time	10-23-91.16:43
b.3	Procedure Type	chest x-ray
c.2	Technique	PA
c.3	Number of Views	1 (view)
c.4	Film Identification	BW report #22
f	Semantic Link (instance 1)	
f.1	Semantic Relationship	compared to
f.2	Linked Event Template ID	003
f.3	Linked Event Instance ID	222
f	Semantic Link (instances 2–16)	
f.1	Semantic Relationship	has observation
f.2	Linked Event Template ID	005
f.3	Linked Event Instance ID	1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1008, 1009, 1010, 1011, 1012, 1013, 1014

*This is the primary procedure event and it has links to all of the observation event instances represented in the full text of the report.

Table 2 ■

Second Procedure Event Instance Representing the *Previous Examination* Referred to in the First Procedure

Label	Attribute	Value
a.1	Patient Identifier	44312897
a.2	Event Family	diagnostic procedures
a.3	Event Template ID	003
a.4	Event Instance ID	222
a.5	Event Date and Time (previous examination)	10-22-91
a.6	Event Store Date and Time	10-23-91.16:43
b.3	Procedure Type	chest x-ray
f	Semantic Link (instance 1)	
f.1	Semantic Relationship	(compared to)*
f.2	Linked Event Template ID	003
f.3	Linked Event Instance ID	1001

*This denotes the inverse of the relationship stated.

The third and fourth event instances (Tables 3 and 4) are observation event instances. They record specific findings that the radiologist reported as being present on the chest x-ray. These observation events have an inverse "has observation" link back to the first procedure event instance. The two observation events are also linked to one another by a Boolean "and" link because the surgical staples are found in the right hilar region *and* the right mediastinum. The ability to link event instances by Boolean operators is essential when "or" relationships exist between findings if the meaning is to be represented accurately. A graphic representation of how the event instances are linked is shown in Figure 2.

Representation of the Model

We now describe how the event model is defined and supported. Event instances are created from event templates (ETs). The ET is a pattern or frame that describes the logical structure of data that are to be stored in the clinical database. An ET is created by giving an event a name and a unique identifier, and then specifying the attributes that characterize the event. An ET is a named structure whose meaning is derived from the set of attributes that it contains. In this sense, ETs are similar to message formats defined in data exchange standards such as HL7[28] (Health Level Seven) and ASTM E1238.[29] These standards specify the name of a message and define which data fields must be sent when transmitting data between computer systems. An example ET is shown

Table 3 ■

First Observation Event Instance Created from the Statement *Surgical clips are again seen along the right mediastinum**

Label	Attribute	Value
a.1	Patient Identifier	44312897
a.2	Event Family	observations
a.3	Event Template ID	005
a.4	Event Instance ID	1002
a.5	Event Date and Time	10-23-91.14:25
a.6	Event Store Date and Time	10-23-91.16:43
d.1	Observation Source	human input
d.2	Observation Type	present (are seen)
d.3	Observation	surgical clips
e.14	Observation Novelty	again
e.19	Observation Relationship	along
e.21	Location	mediastinum
e.22	Location Side	right
f	Semantic Link (instance 1)	
f.1	Semantic Relationship	and
f.2	Linked Event Template ID	005
f.3	Linked Event Instance ID	1003
f	Semantic Link (instance 2)	
f.1	Semantic Relationship	⟨has observation⟩†
f.2	Linked Event Template ID	003
f.3	Linked Event Instance ID	1001

*Note the "and" relationship between this observation and that shown in Table 4.
†This denotes the inverse of the relationship stated.

Table 4 ■

Second Observation Event Instance Created to Represent the Phrase *surgical clips are again seen . . . right hilar region**

Label	Attribute	Value
a.1	Patient Identifier	44312897
a.2	Event Family	observations
a.3	Event Template ID	005
a.4	Event Instance ID	1003
a.5	Event Date and Time	10-23-91.14:25
a.6	Event Store Date and Time	10-23-91.16:43
d.1	Observation Source	human input
d.2	Observation Type	present (are seen)
d.3	Observation	surgical clips
e.14	Observation Novelty	again
e.19	Observation Relationship	along
e.21	Location	hilar
e.22	Location Side	right
e.23	Location Area	region
f	Semantic Link (instance 1)	
f.1	Semantic Relationship	and
f.2	Linked Event Template ID	005
f.3	Linked Event Instance ID	1002
f	Semantic Link (instance 2)	
f.1	Semantic Relationship	⟨has observation⟩†
f.2	Linked Event Template ID	003
f.3	Linked Event Instance ID	1001

*Note the "and" relationship between this observation and that shown in Table 3.
†This denotes the inverse of the relationship stated.

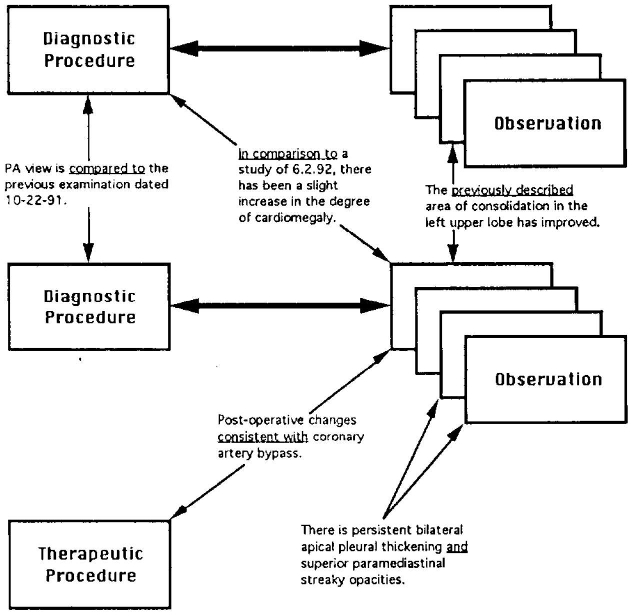

Figure 2 The linking of event instances based on relationships that exist between the events.

in Table 5. The example ET is for a chest x-ray procedure event, but ETs have also been created for laboratory, pharmacy, and physical examination events.

There are usually more details associated with real-world events than can be managed easily in a database. Selecting which attributes should be represented in the model is guided by the purpose for which the data are being collected. The event model, like any model, is only an approximation of the real world. The creator of the event definition must choose which aspects of a real-world event to represent in the model, and which aspects to ignore. However, the model can be made as complex as needed to fulfill the requirements of the people and processes that use the data.

Each attribute (or slot) in the ET has a specific set of possible values that constitute the domain of that attribute for a given event. An event instance is created by selecting a single value for each attribute of the ET and then creating a record in the database with those values. The values are selected so that the instantiated event represents the real-world event.

ETs are created by proceeding from the general to the specific, following the principles of object-oriented modeling. The starting set of attributes contains attributes that pertain to all events. All events, for instance, have a date and time of occurrence. Patient care events are a specialization of general events, and by definition must have an additional attribute that identifies the patient involved in the event. Specialization continues by adding additional attributes as needed to fully describe all the desired characteristics of the real-world event that is being modeled. In the case of the ET shown in Table 5, attributes have been included that specialize from a patient event, to a procedure event, and, finally, to a chest x-ray procedure event.

Table 5 ■

Chest X-ray Procedure Event Template

Label*	Attribute	Possible Values (Domain)†
a.1	Patient Identifier	a unique patient identifier
a.2	Event Family	diagnostic procedures, therapeutic procedures, observations
a.3	Event Template ID	001 (root), 002 (procedure), 003 (chest x-ray procedure), 004 (observation), 005 (chest x-ray observation)
a.4	Event Instance Unique ID	sequentially assigned integer
a.5	Event Date and Time	any valid Julian date and time
a.6	Event Store Date and Time	any valid Julian date and time
b.3	Procedure Type	chest x-ray, coronary artery bypass graft, lobectomy
c.1	Type	portable, serial, . . .
c.2	Technique	PA, AP, lateral, . . .
c.3	Number of Views	a non-zero integer
c.4	Film Identification	a unique alphanumeric identifier of the film
c.5	Radiology Technician	technician identifier
c.6	Radiologist	radiologist identifier
c.7	Ordered by	ordered by identifier
c.8	Date/Time Ordered	any valid Julian date and time
c.9	Date/Time Transcribed	any valid Julian date and time
f	Semantic Link	(repeating set)
f.1	Semantic Relationship	and, or, has observation, compared to, caused by, associated with
f.2	Linked Event Template ID	number, same domain as a.3
f.3	Linked Event Instance ID	number, same domain as a.4

*The alphabetic part of this field indicates the level of specialization of the attribute. Attributes that have labels starting with "a" are common to all patient events. Attributes that have labels starting with "b" are common to all procedure events. Attributes that have labels starting with "c" are used only in chest x-ray procedure events.

†The possible values for each attribute were selected to illustrate the unspecialized domain of the attribute, but it should be understood that the procedure attributes exist in an event template only if "procedure" was selected as the value of "Event Family." Similarly, as values are selected for the higher level coded attributes, the range of possible values for the remaining attributes becomes more restricted.

Table 6 ■

Sample Rows in the Event Body Table*

Event ID	Attribute ID	Label
003 (x-ray procedure)	101 (Event Family)	a.1
003 (x-ray procedure)	289 (Event Template ID)	a.2
003 (x-ray procedure)	734 (Event Unique ID)	a.3
003 (x-ray procedure)	532 (Event Date)	a.4
003 (x-ray procedure)	234 (Proc. Family)	b.1
003 (x-ray procedure)	997 (Proc. SubFamily)	b.2
003 (x-ray procedure)	439 (Proc. Type)	b.3

*The event identifier and the attribute identifier are stored as numbers, but the translation in words is shown in parentheses. The Event Body table makes associations between an event and the attributes it contains.

The process of specialization is reflected in the labels of the attributes as shown in Table 5. Attributes that have labels starting with "a" are common to all patient events. Attributes that have labels starting with "b" are common to all procedure events. Attributes that have labels starting with "c" are specific to chest x-ray procedure events. Example event instances in the appendix use "d" as the label for attributes associated with observations and "e" as the label for chest x-ray–specific observations. The possible values for coded attributes in Table 5 were selected to illustrate the unspecialized domains of the attributes. In the real implementation, the domains of the attributes are more restricted. Procedure attributes exist in an ET only if "procedure" was selected as the value of the "Event Family" attribute. Similarly, when values are selected for any of the higher level coded attributes, the range of possible values for the remaining attributes becomes restricted to the context implied by the earlier choices.

The semantic link attribute (label f) is unique in that it represents a set of attributes (f.1 to f.3) that allow one event instance to be linked by a named relationship to other event instances. The semantic link is also unique in that it is the only attribute that is allowed to repeat. It can be used as many times as necessary to represent all relationships between a given event and its related events.

ETs are recorded in the system using two tables. The Event Head table makes an association between an event identifier and a physical file where event instances are stored. (This association is unnecessary if events are implemented in an object-oriented database, since the objects are stored without reference to a particular file.) The two columns in the Event Head table are the event identifier and the file identifier. The Event Body table is used to make the association between the event identifier and the attributes that are part of the ET. The Event Body table has four columns: the event identifier, the attribute identifier, a hierarchical string, and a domain restriction. Examples of rows in the Event Body table (with meanings shown in parentheses, and ignoring domain restrictions) are shown in Table 6. The details of the attributes themselves are not present in the Event Body table. This is so that the attributes (slots) can be defined independently and used in many different ETs. For example, the attribute "Drug" would be defined once in the attribute table, but it could be used in a drug allergy event, a drug prescription event, a drug administration event, and a microbiology susceptibility testing event.

The values that fill the slots of events can be of many different fundamental types. For instance, patient identifiers may be alphanumeric character strings, blood pressures might be integers, a hepatitis surface antigen test could have a titer, a microscopic urine cell count could be a range of the numbers of cells per high-power field, and the date and time of birth could be a date/time element. There can be as many kinds of values for attributes as are useful in the system, including binary data strings, graphic images, icons, and bar codes.

Many attributes within medicine have values that are expressed using words or text rather than numbers. Attributes such as patient names, symptom names, drug names, disease names, and anatomic locations are examples of these types of textual attributes. Note, however, that the possible values for an attribute such as "drug name" usually come from a finite set of possible choices. This means that it is possible to represent these terms by codes rather than by their full-text representation. This ability to encode the value of textual attributes has important implications in the physical database.

Attributes are defined in the attribute table. The attribute table defines the "event independent" characteristics of a slot, i.e., the characteristics of the slot that are universally true for the attribute regardless of the event in which it may be participating. The "Domain Restriction" field in the event body file is used to further specialize or restrict the domain of an attribute when it is used in a specific event. Sample rows from the attribute table are shown in Table 7. Ranges (domains) for numeric variables are specified as being bounded by high and low values. (Variations on absolute high and low values can be accommodated by adding a few more columns to the table.) However, the domains of coded attributes are specified by pointing the attributes to a finite set of concepts that have been defined within the vocabulary component of the model. For instance, anatomic locations have been previously defined as part of SNOMED International,[30] and the Location attribute domain can be assigned by pointing the attribute to the "T" axis in the SNOMED part of the vocabulary. All coded attributes have their domains assigned in this same manner. This provides the logical link between the ET and the terms in the vocabulary. The relationship between the ETs, their associated attributes, and the concepts in the vocabulary is shown in Figure 3. A more complete description of how domains are created from the concepts in the vocabulary is included in the discussion of the vocabulary component of the system.

To handle complex issues such as negation and probability, each attribute has its own set of implicit or internal characteristics. These implicit attributes could be handled explicitly as slots in the ET itself, but their frequent use and general applicability favor making them a built-in (implicit) part of each event slot. Implicit attributes are used to modify or specify more exactly the meaning of each regular attribute (slot) in an event instance. The implicit attributes of each slot that we have defined are described below, and all examples of their use refer to Figure 4.

Value: Contains the actual value of the slot or attribute. Example #1 shows how the attribute "systolic blood pressure" would be instantiated if it existed in a "vital signs" ET and had the value of "less than 60."

Relational Operator: Symbols for relational operators (>, <, >=, <=, =, !=) are used to indicate open-

Table 7 ■

Sample of Several Rows from the Attribute Table*

Attribute ID	Type	High	Low	Domain
101 (Event Family)	Coded	—	—	Event Families
289 (Event Template ID)	Coded	—	—	Events
734 (Event Unique ID)	Integer	—	—	—
532 (Event Date)	Date	—	12/92	—
729 (Observation)	Coded	—	—	Observations
285 (Obs. Novelty)	Coded	—	—	Novelties
322 (Obs. Relationship)	Coded	—	—	Relationships
482 (Location)	Coded	—	—	SNOMED T axis
167 (Location Side)	Coded	—	—	Sides

*The attribute table defines the context-independent characteristics of the attributes. It makes the association between a coded attribute and its semantic domain. Domains for numeric variables are defined as the highest and lowest possible values for the attribute. Attribute identifiers are stored as numeric values, but the names are shown in words in parentheses.

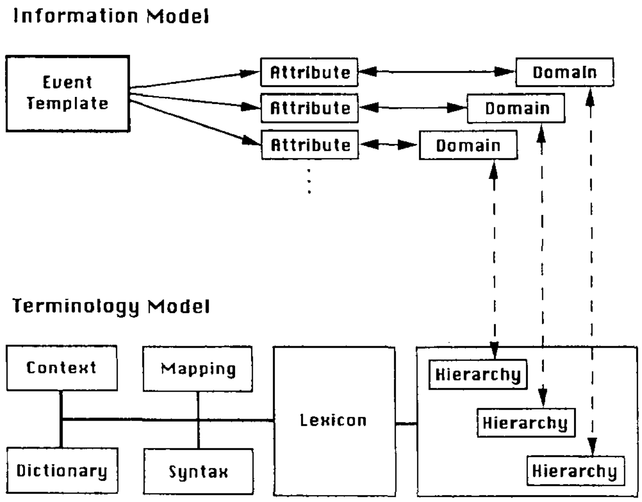

Figure 3 The relationship of the vocabulary component of the system to the event template.

Example #1
Attribute: Systolic Blood Pressure
  Value: 60
  Value Format: Integer
  Units: mmHg
  Relational Operator: <

Example #2
Attribute: Finding
  Value: 367 (pleural effusion)
  Value Format: code
  Fuzzy Probability: not present

Example #3
Attribute: Finding
  Value: 983 (malignancy)
  Value Format: code
  Numeric Probability: 0.35

Example #4
Attribute: Finding
  Value: 367 (pleural effusion)
  Value Format: code
  Fuzzy Probability: probable
  Intensity Modifier: slight

Example #5
Attribute: Finding
  Value: 673 (dyspnea)
  Value Format: code
  Fuzzy Probability: present
  Intensity Modifier: mild
  Specific Synonym Used: 273 (shortness of breath)

Figure 4 Examples of how internal or implicit attributes are used to amplify or modify the meaning of a primary attribute.

ended uncertainty when the value of an attribute is a quantitative variable. See Example #1.

Value Format: As shown in all examples in Figure 4, value format indicates whether the value of the regular attribute is a real number, an integer, a coded value, a sound bite, an icon, an image, etc.

Units: The units of measure are specified as appropriate for any quantitative attribute. See Example #1.

Fuzzy Probability: This implicit attribute allows any coded attribute to be negated or qualified. It specifically handles representation of negation. Thus, the statement "there is no pleural effusion" would be instantiated as shown in Example #2 of Figure 4.

Numeric Probability: Each regular attribute can have a user-stated numeric probability. Example #3 shows the instantiation of the statement "there is a 35% chance of malignancy."

Machine-generated probability: When the event instance is generated from an automated instantiation process, this implicit slot represents the calculated probability that a given attribute is really present (i.e., has been instantiated accurately by the instantiating process). No example is shown for this case, but it is represented in the same manner that numeric probability is in Example #3.

Intensity Modifier: This implicit attribute qualifies the intensity of a coded attribute. Thus, the statement "there is probably a slight pleural effusion" would be instantiated as shown in Example #4.

Specific Synonym Used: This implicit attribute indicates, for coded attributes, which particular synonym was selected during data input. For instance, if the vocabulary had "shortness of breath" and "dyspnea" defined as synonyms and a clinician was using a structured user interface to enter a patient's symptoms and chose "mild shortness of breath" to represent the patient's complaint, the information would be instantiated as shown in Example #5. From the instantiated event one knows that dyspnea is present and that what the user saw when he or she entered the data was "shortness of breath." Storing both codes in an event instance also makes it simple to recover if it is later determined that "dyspnea" and "shortness of breath" are really not synonyms.

Vocabulary

Cimino et al. have previously described the requirements of a good vocabulary.[31,32] We describe only the features of our vocabulary structures that are essential to the event model.

The components of the event model can be viewed as objects. Important objects in the event model are the ET, its attributes, and the terms that may be taken as the values of a coded attribute. The objects in the event model are related in specific ways. For example, an ET "contains" attributes, and a coded attribute such as "Finding" is related to a hierarchy of vocabulary terms by a "has domain" relation. There is a set of characteristics common to all objects in the model, and these common characteristics are catalogued in the master object index (MOI). The common characteristics of all objects in the model are:

1. Name
2. Unique numeric identifier
3. Synonyms
4. Audit or life cycle information, including:

Table 8 ■

Sample Rows from the Master Object Index*

ID	Concept	Language	Syn†	Grp‡	Representation
101	43	English	1	1	Goodpasture's Syndrome
102	43	English	1	2	Syndrome, Goodpasture's
923	43	English	2	1	Lung Purpura with Nephritis
201	43	Portuguese	1	1	Síndrome de Goodpasture
522	43	French	1	3	GOODPASTURE, SYNDROME
109	43	PTXT	1	1	20.1.154.3.159.28.2.0
467	43	SNOMED II	1	1	D-6740
867	43	ICD9CM	1	1	446.21
111	43	Meta 1.4	1	1	C0018085.L0018085.S0045699
112	43	Meta 1.4	1	2	C0018085.L0018085.S0222427
113	43	Meta 1.4	2	1	C0018085.L0024124.S0058478

*The use of the concept identifier to represent a unique "meaning" in the vocabulary while the object identifier is used to tag each unique surface form (string representation) of the concept is shown.
†Syn = synonym.
‡Grp = group.

a. Creation date
b. Created by (a person identifier)
c. Status (proposed, active, obsolete, or inactive)

Each object has a concept identifier that is unique to that object. There is one and only one concept identifier for each "meaning" in the vocabulary. If a given object has many names, or surface forms,[33] each of the various names for the object shares the unique concept identifier. For instance, referring to Table 8, all entries belong to the concept "Goodpasture's Syndrome," and the various names or textual descriptions of "Goodpasture's Syndrome" have the same concept identifier, "43." Each synonym, or different representation of an object, is distinguished by a unique identifier. The concept identifier represents the meaning of the object and the unique identifier represents the particular name used in the instantiation. A similar method of representing synonym classes is used by the UMLS metathesaurus.[34]

Each synonym has an associated language, such as English, French, or Portuguese. Rossi-Mori et al. have provided important evidence that the basic concepts or meanings needed to represent medical information appear to be language independent.[35] Besides natural languages, coding schemes such as PTXT, SNOMED International, the UMLS metathesaurus, and ICD9-CM are also considered languages. Thus, D-6740 is the name for Goodpasture's syndrome in the SNOMED International language. Synonyms are not restricted to textual names. They can be any symbolic representation that identifies the concept, including a sound bite, an icon, an image, or a bar code.

Before the vocabulary concepts can be linked to specific attributes in an ET, they must be organized into sets or domains. There are many potential sources for defining the domains of coded attributes. The UMLS metathesaurus has established semantic types for all of its concepts,[34] which provides a good starting point for defining the domains needed in the event model. The separate axes of SNOMED (i.e., Topography, Morphology, and Procedures) also define domains that are a very good starting point for the domains needed in the event model. However, it has been our experience that the domains needed by the event model are generally more specific than are those that are available in existing vocabularies. The SNOMED T axis happens to fit very well as the domain of "Location" in the example shown in Table 7, but this is the exception rather than the rule. Our strategy is to use existing vocabularies as the source for terms used in the model, but to organize the terms into more specific domains as required. When domains are defined, they become objects in the MOI with specific names and concept identifiers. Once the domain exists in the MOI, it can be associated with an attribute in the attribute table. The relationship between the vocabulary, the attributes, and the ETs is summarized in Figure 3.

Use of the Model

The event model has been a useful construct in several previously reported projects. The first implementation used ETs to construct a coded record of echocardiography findings in a relational table from free-text reports.[36] This early work showed that 90% of echocardiographic findings could be represented as event instances using as few as six attributes in the ET. Independent concepts could be adequately represented, but relationships between findings were not included in this model. This deficiency identified the need for the semantic links present in the current model.

A second project successfully used the event model to map between the vocabularies of two independent hospital information systems.[37] Using the map created, patient data from the two architecturally different systems were combined in a common database that was used for clinical research.

A third project used the ETs for doing automated mapping between the Iliad diagnostic system's dictionary and the UMLS metathesaurus.[38] This project investigated the use of the event model to integrate data stored using different controlled medical vocabularies. For this project, the ETs were used as an "interlingua," i.e., a common underlying structure, where the two medical vocabularies were represented. This experiment was limited to the domain of chest radiology, but the authors were able to determine that the computerized mapping method was comparable to manual mapping performed by a physician.

Discussion

The event model was first conceived in 1986[39] and was inspired by the multiaxial approach of SNOMED[40] and the work of Sager et al.[41] and Flavin.[42] The model builds on the SNOMED concept of combining terms from semantically typed axes to make an aggregate expression that describes a complex medical finding. The event model extends this strategy by adding the context of a particular named event as the focus of the description, and adds fields that represent the time that the event occurred and the time that it was recorded. In this sense, ETs define context-specific subsets of the global SNOMED model. Sager's initial work with the Linguistic String Project (LSP)[43] was syntactically based, as compared with the semantic classes of SNOMED. In the model, this syntactic influence leads to the inclusion of implicit attributes, which play the same role in the model that adjectives and adverbs play in human speech. Flavin's work outlines the principles of information modeling that guide the creation of data structures in database design. He asserts that different classes of real-world entities can be distinguished by the attributes that are used to describe them, and that new classes of entities are needed when a member of a class requires a new attribute to fully describe its role. These principles lead to the creation of classes of events based on shared attributes, with specialization of events occurring by the addition of new attributes to a base ET.

Since the event model was initially proposed, it has been influenced heavily by other models. The event model parallels many of the ideas of case frame theory,[44–47] but adds the idea that slots in the frame can be associated with domains derived from a coded medical vocabulary. The representation of synonyms and homonyms used in the vocabulary component of the model is nearly identical to the strategy used in the UMLS.[48–50] Other models that have been developed by UMLS contractors, including Masarie et al.'s work on generic frames,[51] and the interlingua concepts described by Cimino and Barnett,[52] have contributed to the evolution of the model. We have also gained insights from daily exposure to problems with PTXT and the HELP system database, and from early prototypes of the model.[36,37,53]

The event model is only one of several models that have been recently proposed.[2–7] The event model shares many characteristics with one or more of these models, including:

1. Complex concepts are represented by a combination of atomic terms that have been semantically grouped or typed.
2. The concepts used in the event instances come from a coded medical vocabulary.
3. The atomic components of the model are unique *meanings*, not words or phrases.
4. The temporal progression of findings can be represented.
5. A finding is attributable to a specific observer at a specific date and time.
6. Fuzzy probability and negation are represented.

Areas where we feel that the event model enhances or extends existing models are:

1. Semantic links can be created between event instances. These links allow representation of complex Boolean, causative, and associative relationships that commonly occur in clinical data.
2. The idea of synonyms has been broadened to include coding schemes, icons, sound bites, and images as valid surface forms of a concept.
3. Storage of object (string) identifiers along with concept identifiers in the database provides both the meaning of user input (concept used) and the information about the particular form of expression used (specific synonym used).
4. User-stated numeric probability and instantiation-associated probability (machine-generated probability) have been modeled, in addition to user-stated fuzzy probability.

Areas where our model is lacking, and for which solutions have been proposed by other groups, include:

1. We are currently using a simple model of time. We have not used interval-based logic as proposed by Campbell et al.[2] We feel that the interval-based model is correct, but we have not yet needed the flexibility of the more complex model.
2. While we use a formal frame-oriented notation for representing the event model, it is unique to our project. This is in distinction to using conceptual graph notation as proposed by Campbell et al.[2]

and Friedman et al.,[3] the Structured Meta Knowledge (SMK) system described by Rector et al.,[27] or the Model for Representation of Semantics in Medicine (MOSE) proposed by Rossi-Mori et al.[7]
3. We have not attempted to create a generative grammar. This is in distinction to the work of Rector et al.[27] and Nowlan and Rector,[54] where data representation is restricted to "what can sensibly be said." It is possible to create event instances that are not sensible. In our current implementation, the user is responsible for not constructing nonsensible event instances.
4. We do not have explicitly stated criteria for making modeling decisions as described by Bell et al.[4]
5. We have not tested our model as extensively or applied it as broadly as have Sager et al.[5] or Friedman et al.[3]

One concern that has been raised about the event model is its complexity. Application developers and others are initially overwhelmed by events, ETS, attributes, values, domains, synonyms, concept identifiers, etc. However, the complexity of the model is directly related to the requirements of the applications that use it. Each facet of the model is dictated by some function that we want the model to perform. If a given application does not need some of the functions, then some of the complexity can be avoided. For instance, if synonyms are not needed in the user interface, then several columns in the MOI can be eliminated. Likewise, if multiple languages are not required, several more columns in the MOI can be eliminated. In general, the complexity of the model is dictated by the complexity of the information we are trying to represent. The closer the model approximates the real world, the more complex it becomes. We have not been able to simplify the model without decreasing its functionality.

Another concern expressed about the model is the reuse of attributes in different event definitions. If a "Drug" attribute is used in pharmacy order events, allergy events, and drug level events, doesn't this become confusing? The answer is: it would if the attributes were viewed out of context. However, the system is implemented so that the attributes are always seen in context. For instance, let's assume that a user wants to find patients who are taking penicillin. He or she would enter "penicillin" as a keyword into the system. The system knows that penicillin is a vocabulary concept because it is listed in the MOI, and it also knows the domains in which penicillin participates. Because domains are linked to attributes, and attributes are linked to ETs, the system knows all ETs in which penicillin can participate. The system returns to the user a list of all applicable ETs, and the user is asked to select the type of data he or she would like to see. If the user selects the allergy event, then an unambiguous context of penicillin has been established. One of the underlying assumptions of the model is that there is no clinical meaning without establishing context. Until context is established, penicillin just means a particular chemical substance. However, if penicillin is taken as the value of the "Drug" attribute in an allergy event, the event instance now means that the patient has an allergic reaction to penicillin.

There are several theoretical unknowns about the model. It is unclear how many families of ETs are needed to represent all of medicine. We have already seen the need for procedure events, observation events, history events, family history events, and diagnosis events. We conjecture that the number of families is small. However, each event family gives rise to many specializations. Within the procedure event family, there are chest x-ray procedures, laboratory procedures, surgical procedures, etc. The model will be unsupportable if the number of event definitions increases without bound as the domain of the model increases. We hope to investigate this question in the future.

A final problem that we are trying to solve is related to evaluation of the model. Methods have been described by Sager et al.[5] for evaluating processes that instantiate data records from free text. There are also obvious methods for evaluating the completeness of a coded vocabulary. However, we are unaware of any standard methodology for determining the quality of the model itself. We are currently conducting an experiment to test a possible method for evaluating the quality of the model. As previously reported,[55] the proposed method involves measuring information loss when clinical data are captured in a structured database. The quality of the representation is measured by comparing the number of relevant clinical questions that can be answered from the structured patient database with the number of questions that can be answered directly from the original free-text (natural language) report of the same information. We hope to have the results of this experiment in the near future.

Conclusion

This article describes a model for representing medical information in a database system. The model is organized around events, which are characterized by attributes linked to domains in a controlled medical vocabulary. The model was applied in the domain of chest radiology, and a single chest x-ray was instantiated as an example of how the model can be used. We have implemented a prototype of the model in a relational database.

The event model makes very few assumptions about the physical database. The only elements that must exist in the system are records, fields, and values, or some other set of corresponding data structures that allow an instance of a patient care event to be stored in a database. Any health care facility that was implementing this model would use the most cost-effective and appropriate hardware and software solution for its specific circumstance. The elements that must be common across independent systems if information is to be transferred without complex translations are: 1) the representational model with its associated structure, 2) a structured vocabulary that has been organized by domains, and 3) links that bind the domains of the vocabulary to attributes of the model.

The event model has shown utility in early prototypes for representing clinical data and for mapping between structured clinical vocabularies. However, the model needs to be used and rigorously evaluated in other domains of medicine. It is as yet unclear how many ETs will be needed to cover all of medicine, and whether the number of ETs needed will be manageable. A serious problem in evaluating this model, or any model of medical-concept representation, is the lack of any generally accepted method for measuring the quality of the representation. A second major challenge is developing a scientific and reproducible method for the process of model construction. Now that many different groups have proposed medical-concept representation models, the next step should be collaboration among the independent groups. The focus of the collaboration would be to establish a standard vocabulary and associated functions, develop a common methodology for model construction, adopt a standard nomenclature for model descriptions, and establish guidelines for evaluating the quality of competing models.

References ■

1. Evans DA, Cimino JJ, Hersh WR, Huff SM, Bell DS. Toward a medical-concept representation language. J Am Med Informatics Assoc. 1994;1:207–17.
2. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. J Am Med Informatics Assoc. 1994;1:218–32.
3. Friedman C, Cimino JJ, Johnson SB. A schema for representing medical language applied to clinical radiology. J Am Med Informatics Assoc. 1994;1:233–48.
4. Bell DS, Pattison-Gordon E, Greenes RA. Experiments in concept modeling for radiographic image reports. J Am Med Informatics Assoc. 1994;1:249–59.
5. Sager N, Lyman M, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. J Am Med Informatics Assoc. 1994;1:142–60.
6. Rector AL, Nowlan WA, Kay S. Foundations for an electronic medical record. Methods Inf Med. 1991;30:179–86.
7. Rossi-Mori A, Bernauer J, Pakarinen V, et al. CEN/TC251/PT003: models for representation of terminologies and coding systems in medicine. In: Proceedings of the Seminar, Opportunities for European and U.S. Cooperation in Standardization in Health Care Informatics, Geneva, 1992.
8. Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP System. J Med Syst. 1983;7(2):87–101.
9. Pryor TA. The HELP medical record system. MD Comput. 1988;5:22–33.
10. Kuperman GJ, Gardner RM, Pryor TA. HELP: A Dynamic Hospital Information System. New York: Springer-Verlag, 1991.
11. Warner HR. Computer Assisted Medical Decision-making. New York: Academic Press, 1979.
12. LePage EF, Gardner RM, Laub RM, Golubjatnikov OK. Improving blood transfusion practice: role of a computerized hospital information system. Transfusion. 1992;32:253–8.
13. Gardner RM, Laub RM, Golubjatnikov OK, Evans RS, Jacobsen JA. Computer critiqued blood ordering using the HELP system. Comput Biomed Res. 1990;23:514–28.
14. Tate KE, Gardner RM, Weaver LK. Impact of a computerized laboratory alerting system on patient care and outcome. MD Comput. 1990;7:296–301.
15. Hulse RK, Clark SJ, Jackson JC, Warner HR, Gardner RM. Computerized medication monitoring system. Am J Hosp Pharm. 1976;33:1061–4.
16. Gardner RM, Hulse RK, Larsen KG. Assessing the effectiveness of a computerized pharmacy system. In: Miller RA, ed. Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care. Los Alamitas, CA: IEEE Computer Society Press, 1990:99–105.
17. Evans RS, Larsen RA, Burke JP, et al. Computer surveillance of hospital-acquired infections and antibiotic use. JAMA. 1986;256:1007–11.
18. Thomsen GE, Pope D, East TE, et al. Clinical performance of a rule-based decision support system for mechanical ventilation of ARDS patients. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care. New York: McGraw-Hill, 1993:339–43.
19. Evans RS, Pestotnik SL, Classen DC, et al. Development of a computerized adverse drug event monitor. In: Clayton PD, ed. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care. New York: McGraw-Hill, 1991:23–7.
20. Evans RS, Burke JS, Pestotnik SL, Classen DC, Menlove RL, Gardner RM. Prediction of hospital infections and selection of antibiotics using an automated hospital database. In: Miller RA, ed. Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care. Los Alamitas, CA: IEEE Computer Society Press, 1990:663–7.
21. Nelson BD, Gardner RM, Hedrick G, Gould P. Computerized decision support for concurrent utilization review using the HELP system. J Am Med Informatics Assoc. 1994;1:339–52.
22. Huff SM, Pryor TA, Tebbs RD. Pick From Thousands: a collaborative processing model for coded data entry. In: Frisse ME, ed. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care. New York: McGraw-Hill, 1992:104–8.
23. Haug PJ, Ranum DL, Frederick PR. Computerized extraction of coded findings from free-text radiology reports. Radiology.

1990;174:543–8.
24. Ranum DL. Knowledge based understanding of radiology text. In: Greenes R, ed. Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care. Silver Spring, MD: IEEE Computer Society Press, 1988:141–5.
25. Pryor TA, Hripcsak G. Sharing NLM's: an experiment between Columbia-Presbyterian and LDS Hospital. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care. New York: McGraw-Hill, 1993:399–403.
26. ISO 1087. Terminology—Vocabulary. International Standards Organization.
27. Rector AL, Nowlan WA, Kay S. Unifying medical information using an architecture based on descriptions. In: Miller RA, ed. Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care. Los Alamitas, CA: IEEE Computer Society Press, 1990:190–4.
28. Health Level 7. Ann Arbor, MI: Health Level Seven.
29. ASTM E1238 Clinical Data Interchange Standard. Philadelphia: American Society of Testing and Materials.
30. Cote RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L, eds. The Systematized Nomenclature of Medicine: SNOMED International. Northfield, IL: College of American Pathologists, 1993.
31. Cimino JJ, Hripcsak G, Johnson SB, Clayton PD. Designing an introspective, multipurpose, controlled medical vocabulary. In: Kingsland LC, ed. Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care. Washington, DC: IEEE Computer Society Press, 1989:513–8.
32. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Informatics Assoc. 1994;1:35–50.
33. Evans DA, Rothwell DJ, Monarch IA, Lefferts RG, Cote RA. Towards representations for medical concepts. Medical Decis Making. 1991;11:S102–15.
34. UMLS Knowledge Sources, 5th ed. Bethesda, MD: US Department of Health and Human Services, National Institutes of Health, National Library of Medicine, 1994.
35. Rossi-Mori A, Thornton AM, Gangemi A. An entity relationship model for a European machine-dictionary of medicine. In: Miller RA, ed. Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care. Los Alamitas, CA: IEEE Computer Society Press, 1990:185–9.
36. Canfield K, Bray B, Huff SM, Warner HR. Database capture of natural language echocardiology reports. In: Kingsland LW, ed. Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care. Washington, DC: IEEE Computer Society Press, 1989:559–63.
37. Fu LS, Bouhaddou O, Huff SM, Sorenson DK, Warner HR. Toward a public domain UMLS patient database. In: Miller RA, ed. Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care. Los Alamitas, CA: IEEE Computer Society Press, 1990:170–4.
38. Rocha RA, Rocha BHSC, Huff SM. Automated translation between medical vocabularies using a frame-based interlingua. In: Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care. New York: McGraw-Hill, 1993:690–4.
39. Huff SM, Craig RB, Gould BL, Castagno DL, Smilan RE. Medical data dictionary for decision support applications. In: Stead WW, ed. Proceedings of the Eleventh Annual Symposium on Computer Applications in Medical Care. Los Angeles: IEEE Computer Society Press, 1987:310–7.
40. Cote RA, ed. SNOMED: Systematized Nomenclature of Medicine. Northfield, IL: College of American Pathologists, 1979.
41. Sager N, Friedman C, Lyman MS, et al. Medical Language Processing. New York: Addison Wesley, 1987.
42. Flavin M. Fundamental Concepts of Information Modeling. New York: Yourdon Press Computer Series, 1981.
43. Sager N. Syntactic analysis of natural language. In: Advances in Computers, vol. 8. New York: Academic Press, 1967:153–88.
44. Fillmore C. The case for case. In: Bach E, Harms R, eds. Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston, 1968:1–90.
45. Simmons RF. Semantic networks: their computation and use for understanding English sentences. In: Schank RC, Colby KM, eds. Computer Models of Thought and Language. San Francisco: W. H. Freeman, 1973.
46. Schank RC, Colby KM, eds. Computer Models of Thought and Language. San Francisco: W. H. Freeman, 1973.
47. Riesbeck CR, Schank RC. Comprehension by Computer: Expectation-Based Analysis of Sentences in Context. New Haven, CT: Yale University, 1976.
48. Humphreys BL, Lindberg DAB. Building the Unified Medical Language System. In: Kingsland LC, ed. Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care. Washington, DC: IEEE Computer Society Press, 1989:475–80.
49. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med. 1993;32:281–91.
50. Evans DA. Final Report on the MedSORT-II Project: Developing and Managing Medical Thesauri. Technical Report No. CMU-LCL-87-3. Pittsburgh, PA: Laboratory for Computational Linguistics, Carnegie Mellon University, 1987.
51. Masarie FE, Miller RA, Bouhaddou O, Giuse NB, Warner HR. An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies. Comput Biomed Res. 1991;24:379–400.
52. Cimino JJ, Barnett GO. Automated translation between medical terminologies using semantic definitions. MD Comput. 1990;7:104–9.
53. Canfield K, Bray B, Huff S. Representation and database design for clinical information. In: Miller RA, ed. Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care. Los Alamitas, CA: IEEE Computer Society Press, 1990:350–3.
54. Nowlan WA, Rector AL. Medical knowledge representation and predictive data entry. In: Stefanelli S, Hasman A, Fieschi M, Talmon J, eds. AIME91: Lecture Notes in Medical Informatics, no. 44. Berlin: Springer-Verlag, 1991:105–16.
55. Huff SM, Warner HR, Bray BE, Sorenson DK, Haug PJ. Assessing the accuracy and completeness of structured clinical databases using questions and answers and expert systems (abstract). In: Ackerman MJ, ed. Proceedings of the 1992 Spring Congress of the American Medical Informatics Association. Bethesda, MD: American Medical Informatics Association, 1992:35.

Appendix

Full Text of the Chest-radiograph Report and Its Event Instantiations

Assume that the patient identifier was 44312897 and that the radiograph was obtained on 10-23-91 at 14:25. After the first report, header information is omitted. After the second report, all "has observation" semantic links between procedure instance 1001 and its observations have been omitted. Relationships in brackets are the inverses of the relationships stated.

Full text of radiography report:

BW report #22:

PA view is compared to the previous examination dated 10-22-91.

Surgical clips are again seen along the right mediastinum and right hilar region. There are new surgical clips in the distribution of the circumflex artery, as well as four intact sternotomy wires. There is persistent increased right paramediastinal opacity, possibly related to previous radiation therapy. New plate-like opacities are seen in the left and mid lower lung zones, compatible with atelectasis. There has been some interval improvement in the left pleural effusion.

1. Slight interval decrease in left pleural effusion.
2. Left lower lobe atelectasis.
3. Postoperative changes consistent with coronary artery bypass graft, as well as previous lobectomy on the right.

PA view is compared to the previous examination dated 10-22-91.

(a.1)	Patient Identifier	44312897
(a.2)	Event Family	diagnostic procedures
(a.3)	Event Template ID	003
(a.4)	Event Instance ID	1001
(a.5)	Event Date and Time	10-23-91.14:25
(a.6)	Event Store Date and Time	10-23-91.16:43
(b.3)	Procedure Type	chest x-ray
(c.2)	Technique	PA
(c.3)	Number of Views	1 (view)
(c.4)	Film Identification	BW report #22
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	compared to
(f.2)	Linked Event Template ID	003
(f.3)	Linked Event Instance ID	222
(f)	Semantic Link (instances 2–16)	
(f.1)	Semantic Relationship	has observation
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1008, 1009, 1010, 1011, 1012, 1013, 1014

(a.2)	Event Family	diagnostic procedures
(a.3)	Event Template ID	003
(a.4)	Event Instance ID	222
(a.5)	Event Date and Time (previous examination)	10-22-91
(a.6)	Event Store Date and Time	10-23-91.16:43

(b.3)	Procedure Type	chest x-ray
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	⟨compared to⟩
(f.2)	Linked Event Template ID	003
(f.3)	Linked Event Instance ID	1001

Surgical clips area gain seen along the right mediastinum and right hilar region.

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1002
(d.1)	Observation Source	human input
(d.2)	Observation Type	present (are seen)
(d.3)	Observation	surgical clips
(e.14)	Observation Novelty	again
(e.19)	Observation Relationship	along
(e.21)	Location	mediastinum
(e.22)	Location Side	right
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	and
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1003
(f)	Semantic Link (instance 2)	
(f.1)	Semantic Relationship	⟨has observation⟩
(f.2)	Linked Event Template ID	003
(f.3)	Linked Event Instance ID	1001

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1003
(d.1)	Observation Source	human input
(d.2)	Observation Type	present (are seen)
(d.3)	Observation	surgical clips
(e.14)	Observation Novelty	again
(e.19)	Observation Relationship	along
(e.21)	Location	hilar
(e.22)	Location Side	right
(e.23)	Location Area	region
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	and
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1002

There are new surgical clips in the distribution of the circumflex artery, as well as four intact sternotomy wires.

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1004
(d.1)	Observation Source	human input
(d.2)	Observation Type	present (there are)
(d.3)	Observation	surgical clips
(e.14)	Observation Novelty	new
(e.19)	Observation Relationship	in
(e.21)	Location	circumflex artery
(e.23)	Location Area	distribution
(f)	Semantic Link (instance 1)	

(f.1)	Semantic Relationship	and (as well as)
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1005

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1005
(d.1)	Observation Source	human input
(d.2)	Observation Type	present (there are)
(d.3)	Observation	sternotomy wires
(e.4)	Observation Mod. 4 (appearance)	intact
(e.15)	Observation Numeric Quantity	4
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	and (as well as)
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1004

There is persistent increased right paramediastinal opacity, possibly related to previous radiation therapy.

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1006
(d.1)	Observation Source	human input
(d.2)	Observation Type	present (there is)
(d.3)	Observation	opacity
(e.14)	Observation Novelty	persistent (continuing)
(e.19)	Observation Relationship	(para-)
(e.31)	Progression	increased
(e.21)	Location	mediastinal
(e.22)	Location Side	right
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	related to
	User Fuzzy Probability	possibly
(f.2)	Linked Event Template ID	002
(f.3)	Linked Event Instance ID	333

(a.2)	Event Family	therapeutic procedures
(a.3)	Event Template ID	002
(a.4)	Event Instance ID	333
(a.5)	Event Date and Time	10-23-91.14:25
	Relational Operator	⟨(previous)
(b.1)	Procedure Family	radiation therapy
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	⟨related to⟩
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1006

New plate-like opacities are seen in the left and mid lower lung zones, compatible with atelectasis.

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1007
(d.1)	Observation Source	human input
(d.2)	Observation Type	present (are seen)
(d.3)	Observation	opacities
(e.14)	Observation Novelty	new

(e.4)	Observation Modifier 4 (appearance)	plate like
(e.19)	Observation Relationship	in
(e.21)	Location	lung
(e.22)	Location Side	left
(e.23)	Location Area	zone
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	and
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1008
(f)	Semantic Link (instance 2)	
(f.1)	Semantic Relationship	compatible with
(f.2)	Linked Event Template ID	004
(f.3)	Linked Event Instance ID	1009

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1008
(d.1)	Observation Source	human input
(d.2)	Observation Type	present (are seen)
(d.3)	Observation	opacities
(e.14)	Observation Novelty	new
(e.4)	Observation Modifier 4 (appearance)	plate like
(e.19)	Observation Relationship	in
(e.21)	Location	lung
(e.24)	Location Modifier 1	mid lower
(e.23)	Location Area	zone
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	and
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1007
(f)	Semantic Link (instance 2)	
(f.1)	Semantic Relationship	compatible with
(f.2)	Linked Event Template ID	004
(f.3)	Linked Event Instance ID	1009

(a.2)	Event Family	observations
(a.3)	Event Template ID	004
(a.4)	Event Instance ID	1009
(d.1)	Observation Source	human input
(d.2)	Observation Type	present (are seen)
(d.3)	Observation	atelectasis
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	⟨compatible with⟩
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1007
(f)	Semantic Link (instance 2)	
(f.1)	Semantic Relationship	⟨compatible with⟩
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1008

There has been some interval improvement in the left pleural effusion.

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1010

(d.1)	Observation Source	human input
(d.2)	Observation Type	present (there has been)
(d.3)	Observation	effusion
(e.19)	Observation Relationship	in
(e.21)	Location	pleural
(e.22)	Location Side	left
(e.31)	Progression	improvement
	Intensity Modifier	some
(e.33)	Progression Relation	interval

1. *Slight interval decrease in left pleural effusion.*

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1011
(d.1)	Observation Source	human input
(d.3)	Observation	effusion
(e.19)	Observation Relationship	in
(e.21)	Location	pleural
(e.22)	Location Side	left
(e.31)	Progression	decrease
	Intensity Modifier	slight
(e.33)	Progression Relation	interval

2. *Left lower lobe atelectasis.*

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1012
(d.1)	Observation Source	human input
(d.3)	Observation	atelectasis
(e.21)	Location	lobe
(e.22)	Location Side	left
(e.24)	Location Modifier 1	lower

3. *Postoperative changes consistent with coronary artery bypass graft, as well as previous lobectomy on the right.*

(a.2)	Event Family	observations
(a.3)	Event Template ID	005
(a.4)	Event Instance ID	1013
(d.1)	Observation Source	human input
(d.3)	Observation	post-operative changes
(e.21)	Location	(chest)
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	consistent with
(f.2)	Linked Event Template ID	002
(f.3)	Linked Event Instance ID	444
(f)	Semantic Link (instance 2)	
(f.1)	Semantic Relationship	consistent with
(f.2)	Linked Event Template ID	002
(f.3)	Linked Event Instance ID	555

(a.2)	Event Family	therapeutic procedures
(a.3)	Event Template ID	002
(a.4)	Event Instance ID	444

(a.5)	Event Date and Time	10-23-91
	Relational Operator	⟨(previous)
(b.3)	Procedure Type	coronary artery bypass graft
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	⟨consistent with⟩
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1013

(a.2)	Event Family	therapeutic procedures
(a.3)	Event Template ID	002
(a.4)	Event Instance ID	555
(a.5)	Event Date and Time	10-23-91
	Relational Operator	⟨(previous)
(b.3)	Procedure Type	lobectomy on the right
(f)	Semantic Link (instance 1)	
(f.1)	Semantic Relationship	⟨consistent with⟩
(f.2)	Linked Event Template ID	005
(f.3)	Linked Event Instance ID	1014