Methods of Information in Medicine

J. J. Cimino

Use of the Unified Medical Language System in Patient Care at the Columbia-Presbyterian Medical Center

Columbia University Center
for Medical Informatics,
New York, USA

Abstract: The Unified Medical Language System (UMLS) project at the United States National Library of Medicine contains and organizes a large number of terms from controlled medical vocabularies. This study examines the suitability of the UMLS for representing patient care information as it exists in the Columbia-Presbyterian Medical Center (CPMC) clinical information system. Comparisons were made between the semantic types, semantic relations and medical concepts of the UMLS and the data model entities, semantic classes, semantic relations and concepts in the CPMC system. Results of the comparison demonstrate that the UMLS structural model is appropriate for representing CPMC vocabularies and patient data and that the UMLS concepts provide excellent coverage of CPMC concepts in many areas. Recommendations are made for enhancing UMLS structure to provide additional coverage of the CPMC model. It is concluded that content expansion to provide better coverage of clinical terminology is possible within the current UMLS model.

Keywords: Nomenclature, Unified Medical Language System, Hospital Information Systems

1. Introduction

The United States National Library of Medicine initiated the Unified Medical Language System (UMLS) project in 1986 to develop methods for facilitating access to on-line medical information [1]. The major products of the work have been a UMLS Semantic Network (USN) [2] for classifying medical concepts and inter-concept relations, a Metathesaurus (Meta) to serve as a repository for bringing together concepts from disparate controlled vocabularies [3], and an Information Sources Map (ISM) for describing information sources accessible through use of controlled vocabularies [4]. Recent research on computer-based medical records in both the United States and in Europe has focused on the need for a comprehensive controlled clinical vocabulary. This has naturally led to interest in potential roles for the USN and Meta, especially since the UMLS is readily available in machine-readable form and subsumes many relevant vocabularies.

In the past five years, there have been a large number of papers which examine different aspects of the UMLS. Most of these have dealt with the usefulness of the UMLS in retrievals of online medical information. A very small percentage of the reports have addressed the suitability of UMLS contents for actually representing patient information [5–12]. This paper adds to the latter category by reviewing the role of the UMLS in the large hospital-based clinical information system at the Columbia-Presbyterian Medical Center (CPMC). This study sought to determine the relevance of UMLS structures to clinical information and the degree to which it might be encoded through UMLS content.

This work was supported by the National Library of Medicine and the IBM Corporation.

2. Background

2.1. UMLS Knowledge Sources

At the time of this writing, the UMLS is in its fourth experimental version [13] (all references to the UMLS in this paper will refer to this fourth version). The USN contains 132 semantic types (such as "substance" or "pathologic function") and 46 semantic relations (such as "causes", as in "substance causes pathologic function"). The types are arranged into a strict hierarchy of is-a links. Semantic relations associate pairs of types in a variety of ways, with relationships inherited through the hierarchy. This inheritance allows links to propagate such that the USN actually contains some 6,233 relationships between semantic types ("alkaloid causes congenital abnormality", where "alkaloid" is a descendant of "substance" and "congenital abnormality" is a descendant of "pathologic function"). Meta contains names and associated information about 279,237 terms drawn from twenty-two controlled medical vocabularies. These terms are grouped together based on equivalent meaning into 152,444 concepts. The concepts are, in turn, assigned one or more semantic types, corresponding to those in the USN.

Concepts in Meta are almost always related to one or more other concepts. Usually, these relations are of a broader/narrower or parent/child nature, though in some cases the relationship is simply given as "other". In a small proportion of cases, the relationship is made explicit, based on the semantic relationships allowed in the USN for Meta concepts involved.

2.2. The CPMC Clinical Information System (CIS) and the Medical Entities Dictionary

The CPMC Clinical Information System (CIS) is based on a large relational clinical database which was designed using a conceptual model of patient information [14]. Interaction with the database occurs through a set of data access modules (DAMs). The DAMs accept data from ancillary systems to be stored in the database and respond to queries against the database.

Coded data from ancillary systems are translated into a central coding scheme using the Medical Entities Dictionary (MED) [15]. The MED consists of a semantic network with a directed acyclic graph to provide for multiple, coexisting hierarchies. The semantic network was originally based on the semantic types and relations in the USN and has since evolved as needed to accommodate controlled vocabularies from various clinical systems. Terms in the MED were taken from CPMC ancillary systems that include laboratory, pharmacy, and discharge coding, as well as other applications such as an automated decision support system [16] and an automated literature retrieval system [17]. Concepts in the MED are taken primarily from the ancillary systems; however, in many cases additional terms are needed to provide adequate descriptive information. These additional terms were drawn, when available, from Meta.

At this writing, the MED contains over 35,000 terms, including over 2,000 laboratory tests, 3,000 medications, and 20,000 terms from the International Classification of Diseases, Ninth Edition, with Clinical Modifications (ICD9-CM) [18].

3. Methods

Because the MED is based on the UMLS model, a direct side-by-side comparison of the two systems was possible. Both include semantic types, semantic relations and medical concepts. In the MED, these are included in a single database, while the UMLS places semantic types in the USN and medical concepts in Meta. UMLS semantic relations can be found in both knowledge sources.

The specific comparisons were:

a. MED database entities vs USN semantic types
b. MED classes of coded data vs USN semantic types
c. MED semantic links with USN semantic relations
d. MED concepts from selected clinical domains vs Meta concepts
e. Specific interconcept semantic links in the MED vs specific interconcept relations in Meta

The comparisons were done through a combination of manual and semiautomated methods. When the numbers of items to be compared was small (such as in tasks a-c), simple manual comparison was sufficient. Matching between MED and Meta concepts was conducted through either manual review of "small" domains (such as laboratory and diagnostic tests, of which Meta has only 1,209) or through automated lexical matching with manual review of retrievals (such as for matching chemical names).

Lexical matching was carried out through the use of a word index supplied with the UMLS and a keyword synonym list constructed through analysis of word substitution across term synonyms. In this analysis, multiword Meta terms were examined to find synonymous terms (as assigned in Meta) which shared some common words. When shared words were removed, the remaining words were proposed as possible synonyms. For example, the terms "atrial fibrillation" and "auricular fibrillation" are listed as synonyms in Meta. After removal of "fibrillation", the words "atrial" and "auricular" can be extracted as synonyms.

Subsequent lexical matching made use of these word synonyms to expand the number of words searched in the word index and, thus, the number of Meta concepts retrieved.

4. Results

4.1. Comparison of Database Entities with the USN

Patient information in the CPMC clinical database is organized into general classes: patient demographics (e. g., name, date of birth, address), provider (e. g., doctors, nurses, technicians), organizations, locations and events. Events capture patient data in terms of their occurrence at some point in time. Typical events include: hospital admission, medication order and laboratory result. The USN provides a strict hierarchy of 132 semantic types. They are organized into two trees, headed by the types *Event* ("a broad type for grouping activities, processes and states") and *Entity* ("a physical or conceptual entity"). Most of the CPMC model maps well into the USN, as shown in Table 1.

A closer inspection of Event reveals an interesting distinction between the items classified as events in the CPMC and USN models. In the USN, events are broadly defined to be, in essence, any action. In the CPMC model, events are the occurrences of actions. This difference has some implications for the way that concepts are classified. In the CPMC model, an appendectomy is considered a procedure and is therefore an "entity". It has no time or person associated with it. If an appendectomy is

Table 1 Comparison of CPMC database entities with USN Semantic Types.

CPMC Entity	USN Type
Patient	Human
Patient Demographics	Organism Attribute
Provider	Professional or Occupational Group
Organization	Organization
Location	Geographic Location
Event	Event

Table 2 Comparison of MED classes of coded data from applications with USN Semantic Type.

Application	MED Class	USN Type
Clinical Laboratory	Laboratory Procedure	Laboratory Procedure and Diagnostic Procedure
Clinical Laboratory	Laboratory Test	Laboratory Procedure and Diagnostic Procedure
Clinical Laboratory	Specimen	<none>
Clinical Laboratory	Finding	Laboratory or Test Result
Clinical Laboratory	Organism	Organism
Pharmacy	Medication	<none>
Pharmacy	Drug Form	<none>
Discharge Coding	ICD9 Disease	Disease or Syndrome
Discharge Coding	ICD9 Procedure	Diagnostic Procedure and Therapeutic Procedure
Clinical Information	HL7 Event	Event
Decision Support	Alert	Event
Medline Button	Generic Query	Idea or Concept
*	Measurable Entity	<none>
	Chemical	Chemical
	Body Substance	Body Substance
	Cell	Cell
	Organism	Organism
	Organ or Tissue Function	Organ or Tissue Function
	Spatial Concept	Spatial Concept
	Temporal Concept	Temporal Concept
	Qualitative Concept	Qualitative Concept
	Quantitative Concept	Quantitative Concept
*	Samplable Entity	<none>
	Body Substance	Body Substance
	Body Space or Junction	Body Space or Junction
	Anatomical Structure	Anatomical Structure
*	Etiologic Agent	<none>
	Chemical	Chemical
	Organism	Organism
	Procedure	Healthcare Activity
*	Anatomic Site	Anatomical Structure

* indicates classes added to describe ancillary terms

performed, then there is an event which has, as the performed activity, the procedure appendectomy. Semantic classes such as procedures can participate in a number of roles in this way. Thus, rather than having terms such as "history of appendectomy", "order an appendectomy", "appendectomy event", the CPMC model uses the more parsimonious "history of <procedure>", "order a <procedure>" and "<procedure> event". Using this view, the CPMC model considers procedures and other activities to be conceptual entities. Events are also considered to be conceptual entities.

4.2. Comparison of Classes of Coded Data with Semantic Net

Coded data in the CPMC clinical database are derived largely from ancillary systems. In order to incorporate these terms into the MED, each term needs an assigned location in the MED hierarchy. Representation of ancillary terminology in the MED often requires the addition of descriptive terms. For example, in order to describe laboratory tests, it was necessary to add chemical terms to provide information about what the tests measure. The ancillary terms and their descriptors fall into natural classes. Table 2 shows the classes of terms in the MED and how these classes correspond to USN Types. The USN provided good coverage for classes of terms used at CPMC; however, there were several situations in which the coverage did not occur.

First, the USN (and Meta) does not distinguish between medications and chemicals. In the MED, medications are modeled as entities which are related to chemicals as pharmaceutic components. This distinction is used to separate the properties of medications (manufacturer, pill size, etc.) from those of the chemicals themselves.

Second, some classes (Specimen and Drug Form) are missing from the USN which, in the MED, contain substantial numbers of specific concepts. Terms corresponding to these concepts can be found in Meta, however. "Specimen" terms in Meta are classified as having the USN type "body substances". The MED distinguishes specimens from body substances since specimens can have attributes (collection method, tube type, etc.) which are not properties of the body substance itself. "Drug Form" terms in Meta are classified as having the USN type "Biomedical or Dental Material"; this appears to be an acceptable broad class in which drug form terms could be classified. In fact, a concept called "Dosage Forms" exists in Meta (with semantic type "Biomedical or Dental Material") which is identified as a parent of eleven more specific forms. Thus, the classification for these concepts exists, but is relegated to the level of Meta, rather than the USN.

Finally, some classes have been created in the MED to group together other classes with some common property. Specifically, the MED provides for the semantic classes Measurable Entity, Samplable Entity and Etiologic Agent. Each of these three primary classes is a collection of three or more subordinate classes. As can be seen in Table 2, while the primary classes do not exist in the USN, there are good corresponding types in the USN for each of the subordinate classes. The principal reason for the creation of these primary classes is due to a design consideration in the MED: whenever a relationship is needed between classes, it may be introduced at one and only one point in the hierarchy. Thus, in order for organisms, chemicals and procedures to participate as etiologic agents in disease term descriptions, they must all be included in the class (Etiologic Agent) which introduces the relationship (Etiology). Since the USN does not include this restriction (i. e., the same semantic relation can be defined between multiple pairs of semantic types) there is less need for a unifying class.

4.3. Comparison of Semantic Relations in MED with UMLS Semantic Relations

4.3.1. HAS-PARTS // PART-OF

The MED introduces a nonspecific partitive relationship at the highest level in the hierarchy (i. e., at "Medical Entity"). Thus, all concepts in the MED inherit the attributes "has part" and "part of". With this arrangement, anything can have anything else as a part or be a part of anything. Actual instantiations of this relationship occur primarily in two places: specific Laboratory Procedures are linked to specific Laboratory Tests and specific events are linked to other specific events (known in the MED as event components).

The USN provides the relationships "part_of" and "conceptual_part_of"; however, it does not accommodate the notion that laboratory tests, or any other kind of procedure, may have component procedures. Interestingly, the USN does allow "Laboratory or Test Results" to be part of each other. The USN also allows one procedure to be a method of another procedure, such as "molecular cloning is a method of biochemical genetics". Current efforts are underway to develop a national standard for test naming conventions which includes such information (McDonald, CM: personal communication); however, the MED does not require such information, nor is it readily available from the CPMC laboratory system.

The USN does not allow events to have other events as parts. However, it does allow temporal concepts to have other temporal concepts as conceptual parts. Since, in the MED, events are in effect temporal occurrences of activities (such as procedures), the USN would allow events to have component events if they were treated as temporal concepts.

4.3.2. SUBSTANCE-MEASURED // MEASURED-BY

The USN recognizes that laboratory and diagnostic procedures "measure" other entities, including: Body Substance, Chemical, Biologically Active Substance, Biomedical or Dental Material, Natural Phenomenon or Process, Organism Attribute, Organism Function, Quantitative Concept and Temporal Concept. As mentioned above, the MED includes the concept "measurable entity" which, in turn, subsumes many of these classes. However, it also includes Cell, Organism, Spatial Concept, and Qualitative Concept. Thus, the USN does not recognize that diagnostic or laboratory tests can measure leukocytes, bacteria, nucleotide sequences (a spatial concept) and color.

4.3.3. RESULT-TYPE // RESULT-OF-TESTS

The USN recognizes that laboratory and diagnostic procedures can have other entities as "results", including: Acquired Abnormality, Congenital Abnormality, Injury or Poisoning, Laboratory or Test Result and Pathologic Function. At present in the MED, test results include findings, cells and organisms. As the MED expands, the use of diseases as test results can be anticipated.

4.3.4. SPECIMEN-OF // SPECIMEN

The MED (and most laboratory systems) includes a class of terms representing the specimens on which laboratory tests are performed. The USN does not include this class; thus, there is no corresponding "specimen" relation in the USN. (Meta included the term "stool specimen", but it is unrelated to any test terms.)

4.3.5. SUBSTANCE-SAMPLED // SAMPLED-BY

As mentioned above, the MED models specimens as a distinct class of entities related to other entities which represent the material from which the specimen is drawn. The USN does not model specimens in this way, nor is there any relation in the USN which could accommodate relations between tests and substances analyzed (such as body fluids). [Meta "stool specimen" concept is listed as a body substance but is unrelated to the term "feces" (the preferred name for "stool")].

4.3.6. PHARMACEUTIC-COMPONENT // PHARMACEUTIC-COMPONENT-OF

The USN does not have a semantic type specifically for medications. Instead, it provides a hierarchy for chemicals which includes the type "Pharmacologic Substance". Meta includes many medications; they are considered to be concepts which are "narrower than" the concept for the chemical they contain. This arrangement permits the establishment of chemical/drug relationships as in the MED.

4.3.7. ETIOLOGY // CAUSES-DISEASES

The MED provides a single class – Etiologic Agent – under which to collect all concepts which may be "Disease or Syndrome" etiologies. At present it includes chemicals, organisms and procedures. The USN includes the same relationship, but uses it to connect different semantic types, as in Table 3. In some cases, the USN is much more specific than the MED (for example, only certain organisms can be listed as etiologic agents). The USN is also broader than the MED in some places (for example, "substances" can cause any "pathologic function", not just "disease or syndrome").

4.3.8. SITE // SITE-OF-DISEASE

The USN contains the relation "location of" which allows connection between disease concepts and anatomic concepts.

4.3.9. Events

As discussed above, events in the MED are modeled as discrete episodes of some activity in a particular time. The USN model basically considers the activity itself to be an event. The MED includes seventeen different relation-

Table 3 UMLS semantic types corresponding to agents which have the "causes" relation to some disorder.

Agent	Disorder
Fungus	Pathologic Function
Virus	Pathologic Function
Rickettsia or Chlamydia	Pathologic Function
Bacterium	Pathologic Function
Invertebrate	Pathologic Function
Substance	Pathologic Function
Substance	Congenital Abnormality
Substance	Acquired Abnormality
Substance	Injury or Poisoning
Manufactured Object	Pathologic Function
Manufactured Object	Congenital Abnormality
Manufactured Object	Acquired Abnormality
Manufactured Object	Injury or Poisoning

Table 4 MED concepts with corresponding UMLS concepts.

MED Class	Number in MED	Number in UMLS (%)
Specimen	43	1 (2.3)
Laboratory Procedures	59	15 (25.4)
Laboratory Tests	416	52 (12.5)
ICD9 Procedures	7,218	3,498 (48.5)[a]
ICD9 Diseases	12,656	12,523 (98.95)[a]
Organisms	617	345 (55.9)
Medications[b]	642	642 (100)
Measurable Entities	443	334 (75.4)
Samplable Entities	108	72 (66.7)

[a] UMLS content to include 100% of ICD9 Diseases and Procedures in 1994
[b] UMLS Chemical concepts were used since UMLS does not distinguish between chemicals and medications

Table 5 MED concepts with semantic relations in Meta.

MED Class	# MED[a]	# Meta[b]	# SR[c]	Predominant classes	# OR[d]	Predominant classes
Medication (chemical)	642	642	0/0	–	367/532	Chemicals and Poisonings
Measurable Entity	443	334	18/60	Cells related to body parts	176/401	Chemicals and Tests
Samplable Entity	108	72	48/221	Body parts related to other body parts	56/202	Body parts to tests
Lab Tests and Procedures	475	77	7/14	Other lab tests	77/253	Chemicals, lab tests, findings, diseases

[a] number of terms in the MED
[b] number of corresponding Meta terms
[c] number of terms having semantic relations/total number of semantic relations
[d] number of terms having "other" relations/total number of "other" relations

ships to other concept classes which are derived from Health Level 7 (HL7) [19] and deal largely with who performed what activity where, when and on whom. The USN relates Events to Occupation or Discipline with the semantic relation "issue in" (defined as "an issue in or a point of discussion, study, debate, or dispute"). This does not correspond to any of the seventeen relations in the MED.

4.4. Comparison of Coded Classes with Meta Content

The CPMC clinical laboratory system provides terms for 787 tests, 604 procedures (which are collections of tests), 179 specimens, and 970 findings (110 microscopy smear findings, 828 culture findings, and 32 modifiers of culture findings). Most of these terms are institution-specific. For example, the list of procedures includes "Presbyterian Hospital Chem-7", "Allen Pavilion Chem-7" and "Chem-7 Profile". The MED collects similar institution-specific terms into more generic classes. The procedures in the preceding example are all included in the class "CHEM-7". In all, there are 416 classes for tests, 59 classes for procedures, and 43 classes for specimens.

Meta contains 1,209 concepts which are assigned one or both of the semantic types "laboratory procedure" or "diagnostic procedure"; 475 of these correspond to procedures or tests in the MED. As previously noted, the USN does not provide a semantic type for specimen (although the concept "stool specimen" is present in Meta).

Meta does not contain terms which relate to the microscopy smear findings and culture finding modifiers from the laboratory system. In addition, many of the culture findings are messages such as "No Growth to Date" or "Specimen Not Properly Collected". However, 617 of these findings are specific organisms. Meta includes 345 of these. Of the 272 others, the genus, but not the species, in present in Meta.

The 179 specimen concepts are represented in the MED through links to 108 concepts which correspond to the "substance sampled" by the specimens; 72 are found in Meta. The remainder include a number of specific body fluids (such as pancreatic cyst fluid) and lesions (such as lip lesion). Also missing from Meta are sampled substances "serum" and "fomite".

The 787 laboratory tests are represented through links to 443 concepts which correspond to the "substance measured" by the tests; 334 are found in Meta. The remainder include a variety of cells (e. g., target cells) isoenzymes (e. g., LDH 1-5) and specific antibodies (e. g., lyme IgG antibody). The results of the comparison of Meta concepts with the laboratory class terms are summarized in Table 4.

4.5 Comparison of MED Links with Meta Links

Meta includes over one million relationships between concepts. Most of these are parent-child or broader-narrower links. A small proportion are identified as semantic links as allowed by the USN. In addition, concepts are often interrelated through nonspecific "other" links. For example, among the 334 measurable entities found in Meta, 18 concepts (all of semantic type "cell") participate in a total of 60 semantic links to other body parts while another 176 concepts (most of semantic type "chemical") are linked to 401 other concepts, predominantly chemicals and tests.

5. Discussion

This study examines degree to which the UMLS covers CPMC clinical information. Since the existence of the former had a strong influence on the development of the latter, a certain degree of similarity is not unexpected. However, the CPMC CIS is not a toy model of a clinical system based on the UMLS. The primary driving force in the design of the CPMC information model and in the building of the MED has been the information which is supplied to the CIS by a variety of applications which are *not* based on the UMLS. In fact, the UMLS is primarily a collection of controlled vocabularies which, for the most part, are not intended for clinical systems. Thus, the degree to which the UMLS can cover the CPMC CIS structures and content is an encouraging sign.

The UMLS can be considered at three different levels. At the highest "meta" level, is the general arrangement of the UMLS into semantic types, semantic relations, medical concepts and interconcept relations. Each of these aspects is completely consonant with the arrangement of CPMC. One difference worth noting is that the MED does not distinguish in any formal way between semantic classes and concepts; that is, in the MED, a semantic class is also a concept. Although this distinction is made in the UMLS design, in reality the separation is blurred, since many USN semantic types can be found as concepts in Meta.

The next level of consideration is the conceptual content of both the USN and Meta. For the most part, the classes of data in the CPMC CIS and the important classes and concepts of the MED are well represented in the UMLS. When dissimilarities occur, they are either due to simple philosophical differences of organization (for example, the MED groups all things measured by tests into "Measurable Entity") or due to some incompleteness of content in the UMLS. Differences in organization are not surprising, since the USN is based on the kinds of concept classes and interrelationships which occur in medical information sources (such as Medline), rather than those which occur in clinical databases. Such differences can be subsumed in the MED, where multiple hierarchies can be incorporated. The USN does not, at present, include multiple hierarchies, although they may be allowed. Differences in content reflect inadequacies of the UMLS source vocabularies, not the UMLS itself.

The final level of consideration is the specific interrelationships in the UMLS (among semantic types and among concepts). As mentioned above, the USN was not designed to meet the needs of clinical information systems. Nevertheless, the specific links in the USN match those in the MED rather well. On the other hand, the links between concepts in Meta are rather poor in relation to those needed in the MED. Specific semantic links (such as "glucose test measures glucose") are rare in Meta. The "other" links are more prevalent and may be of use for helping human vocabulary builders to create interconcept links. However, further expansion of the semantic links in the Meta would improve its vocabulary coverage. A first step, for example, would be to identify a semantic relation for each "other" link.

6. Recommendations

Based on these considerations, and the specific results of this study, the following recommendations have been made to the NLM:

6.1. Make "Event" a Temporal Concept and Relate it to Various Activities

At CPMC, an event is a concept which involves an action, a time and one or more participants, such as ordering a test, performing a test, or reporting a test result. The test itself is not considered an event, but rather is a procedure which has a role in an event.

6.2. Consider Making the USN a Directed Acyclic Graph

In the MED, semantic relations are introduced at a single point in the hierarchy. This is extremely convenient for managing the links. However, each link might suggest its own grouping. For example, the "substance measured" link suggests a "measurable entities" class which may not be mutually inclusive or exclusive of all other classes. The use of a directed acyclic graph in the MED allows for these classes to exist in harmony.

6.3. Reconsider the Conceptual Entity vs Physical Entity Polarization

There are many places where adherence to this view of the world causes difficulty. For example, there is no single class which subsumes all anatomic entities, since some, such as body surface, are conceptual while the rest are physical. Making the USN a directed acyclic graph would allow the addition of "anatomic entity" to subsume these while still enabling the physical and conceptual to be separated.

6.4. Recognize that Tests have Specimens

This is a fundamental aspect of clinical medicine.

6.5. Recognize that Tests have Parts

This, too, is a fundamental aspect of clinical medicine.

6.6. Consider Separating Medications from Chemicals

In the CPMC CIS, the notion of "medication" is useful for providing ways to assign specific attributes to drugs which can not be associated with chemicals (e. g., form, size, etc.). It seems natural to consider chemicals as substances which can be measured by tests. The role of chemicals as ingredients in drugs seems comparable. This simplifies, for example, the modeling of combination drugs which have multiple ingredients.

6.7. Consider Expansion of Allowable Relations

The results in Section 4.3 and in Table 3 provide specific examples of places in the USN where additional links should be added.

7. Conclusion

The UMLS structure appears to fit well with the CPMC CIS. For the most part, areas of content discrepancy are more a matter of prioritizing the ongoing vocabulary construction efforts. It is worth noting that the UMLS is far larger than the MED and is likely to hold clinical concepts which have not yet been needed at CPMC. It is also worth noting that the MED has dealt primarily with coded data and is only now beginning to venture into the domains which are dominated by natural language representation. The representation of heretofore uncoded information will be crucial to the success of comprehensive medical records systems. This study does not attempt to evaluate the suitability of the UMLS for such a purpose.

REFERENCES

1. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth Inf Med 1993; 32: 281–91.
2. McCray AT. The UMLS Semantic Network. In: Kingsland LC, ed. *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press 1989; 503–7.
3. Tuttle MS, Sherertz D, Erlbaum M, Olson N, Nelson SJ. Implementing Meta-1: the first version of the UMLS Metathesaurus. In: Kingsland LC, ed. *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press 1989; 483–7.
4. Masys DR. An evaluation of the source selection elements of the prototype UMLS Information Sources Map. In: Frisse ME, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1992; 295–8.
5. Cimino JJ, Hripcsak G, Johnson SB, Friedman C, Fink DJ, Clayton PD. UMLS as knowledge base – a rule-based expert system approach to controlled medical vocabulary management. In: Miller RA, ed. *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press 1990; 175–80.
6. Cimino JJ. Representation of clinical laboratory terminology in the Unified Medical Language System. In. Clayton PD, ed. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1991; 199–203.
7. Fu LS, Huff S, Bouhaddou O, Bray B, Warner H. Estimating frequency of disease findings from combined hospital databases: a UMLS project. In: Clayton PD, ed. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1991, 373–7.
8. Friedman C. The UMLS coverage of clinical radiology. In: Frisse ME, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1992: 309–13.
9. Greenes RA, McClure RC, Pattison-Gordon E, Sato L. The findings-diagnosis continuum: implications for image descriptions and clinical databases. In: Frisse ME, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1992; 383–7.
10. Sato L, McClure RC, Rouse RL, Schatz CA, Greenes RA. Enhancing the Metathesaurus with clinically relevant concepts: anatomic representations. In: Frisse ME, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1992; 388–91.
11. Zielstorff RD, Cimino C, Barnett GO, Hassan L, Blewett DR. Representation of nursing terminology in the UMLS Metathesaurus: a pilot study. In: Frisse ME, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1992; 392–6.
12. Payne TH, Martin DR. How useful is the UMLS Metathesaurus in developing a controlled vocabulary for an automated problem list? In: Safran C, ed. *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1993; 705–9.
13. National Library of Medicine. UMLS *Knowledge Sources – Fourth Experimental Edition Documentation.* Bethesda, Md: The Library, April 1993.
14. Johnson SB, Friedman C, Cimino JJ, Clark T, Hripcsak G, Clayton PD. Conceptual data model for a central patient database. In: Clayton PD, ed. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1991; 381–5.
15. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. JAMIA 1994; 1: 35–50.
16. Hripcsak G, Cimino JJ, Johnson SB, Clayton PD. The Columbia-Presbyterian Medical Center decision-support system as a model for implementing the Arden Syntax. In: Clayton PD, ed. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1991; 248–52.
17. Cimino JJ, Johnson SB, Aguirre A, Roderer N, Clayton PD. The Medline Button. In: Frisse ME, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1992; 81–5.
18. United States National Center for Health Statistics. *International Classification of Diseases, Ninth Revision, with Clinical Modifications,* Washington, DC, 1980.
19. Health Level Seven. *An Application Protocol for Electronic Data Exchange in Healthcare Environments – Version 2.1.* Chicago: Health Level Seven, Inc, 1990.

Address of the author:
James J. Cimino, M. D.,
Center for Medical Informatics,
Atchley Pavilion, Room 1310,
Columbia-Presbyterian Medical Center,
New York, NY, 10032
USA