MEDINFO 98
B. Cesnik et al. (Eds)
Amsterdam: IOS Press

An Evaluation of UMLS as a Controlled Terminology for the Problem List Toolkit

Howard Goldberg[a], Denise Goldsmith[a], Vincent Law[a], Kevin Keck[b], Mark Tuttle[b], Charles Safran[a]

[a]*Division Of Clinical Computing, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA*
[b]*Lexical Technology, Incorporated, Alameda, California, USA*

Abstract

We are developing a set of software components—the Problem List Toolkit (PL-Tk) -- to support operations on clinical problem labels. An adaptation of the National Library of Medicine's Unified Medical Language System (UMLS) provides general vocabulary services to domain-specific software components. Our initial investigation centers on the inclusion in UMLS of problem labels used in the Beth Israel Deaconess Medical Center's Online Medical Record (OMR). We also explore the semantic typing of problem labels matched in UMLS. We have operationally defined a clinical problem to derive its semantic type from classes of terms representing findings or processes typically requiring diagnostic evaluation or therapeutic management in clinical practice. Of 1262 unique OMR problem labels, 999 terms (79%) have matches in UMLS. 986 of 999 terms (99%) map to the UMLS concept of the corresponding lexical match. 952 of 999 terms (95%) have semantic types that comply with our operational definition of clinical problems. These 952 terms (75%) constitute Version 1.0 of the problem list vocabulary BI96. Matching terms with inappropriate semantic types raise issues regarding requirements for PL-Tk, typing of existing UMLS terms, and the adequacy of our operational definition for clinical problems. UMLS provides a large repertoire of pre-coordinated terms that are used as problem labels in a heavily used computer-based patient record system. The semantic type hierarchy provides a framework for the consistent use of clinical concepts in problem lists such that clinical problem labels represent "good" clinical problems.

Keywords

Controlled Clinical Terminology; Problem List; Unified Medical Language System

Introduction

Facilities to assign standardized, high fidelity labels to patients' clinical problems are an important requirement in the ongoing evolution of clinical information systems. Labels that capture the "state" of the patient are fundamental data in both clinical practice and research. Problem labels are used to stratify individual patients for diagnostic evaluation and therapeutic management, and subsequently utilized to stratify cohorts of patients for the purposes of outcome measurements. One would assume that the fidelity of the label to the actual state of the patient must affect the quality of the next clinical decision or the validity of a research study. As contemporary clinical information systems are comprised of clinical documentation systems in which problems labels are generated, decision support systems which monitor real-time patient management, and clinical repositories which support aggregate outcome analysis, it is crucial for the information system to provide vocabulary services for selecting, generating, and administering the terminology of clinical problems.

This report describes our initial research into problem-based vocabulary services which we call the "Problem List Toolkit" (PL-Tk). Our long-term goal in the development of the PL-Tk is to develop a comprehensive set of software services for use by application developers to operate on clinical problem labels. The space of clinical problems circumscribes a distinct set of functionality--what is a clinical problem, what are specific interrelationships between clinical problems or relationships between clinical problems and other clinical concepts--that leverages and specializes functionality provided by a "generalized" vocabulary server. We intend to leverage the capabilities of a "generalized" vocabulary server--an adaptation of the National Library of Medicine's Unified Medical Language System (UMLS)-- to provide lexical matching and semantic typing, and will layer our domain-specific functionality on top these foundational services.

Our short-term goal in the development of PL-Tk is to develop an application programming interface (API) to support the selection of standardized problem labels within clinical documentation applications, e.g., progress notes, discharge summaries, etc. This functionality will support the modification of existing applications to capture problems from the standardized vocabulary as well as the development of utilities to standardize existing problems in an automated or semi-automated way. Our initial requirements are as follows:

- The toolkit must be able to generate the entire repertoire of authoritative problem labels, either returning valid atomic terms from the underlying dataset or constructing valid compositional terms from accepted problem label templates, e.g., *<disease><location>* as in *Right Leg Cellulitis*

- The API provides entry points to problem labels via both lexical matching and user navigation. There should exist API functions to
  - return all valid problem labels matching a given input string
  - return a navigable relationship leading from the input string to valid problem labels. For example, given the string "*chlamydia*", an API function would return the relationship "*causes*" to support navigation from a type of *<organism>* to a related *<disease or syndrome>* recognized as a problem label.
  - Modify problems with operators such as "*history of*" or "*status post*"
  - Compose problems using operators such as "*associated with*" or "*secondary to*"

Our primary focus has been on evaluating the coverage of UMLS of our existing legacy problem list vocabulary. A separate but related task has been creating and refining a machine-implementable definition of ***clinical problem.*** We further report on the extent of coverage of our existing problem list vocabulary within UMLS and subsequent application of our definitional filter in determining whether a problem exists in the standardized vocabulary, and, if indeed, it is truly a "good" problem.

Materials and Methods

The UMLS Knowledge Sources are a group of datasets and utility programs "designed to facilitate the retrieval and integration of information from multiple machine-readable biomedical information sources".[1,2] Components of importance to the current work include

- Metathesaurus (Meta) concept database and the associated normalized word and string indices into the database
- Semantic Network concept type hierarchy
- Specialist Lexicon including the utility program "norm" which generates a canonical representation of input strings to serve as entry points into the normalized string index [3]

The Metaphrase™ API, under development by Lexical Technology, Incorporated, adapts the resources of UMLS described above, and provides a set of functions that match phrases against Meta concepts, and retrieve various concept attributes given a concept identifier. Storage for UMLS datasets is provided via the Berkeley Database Package, a freely available programmatic toolkit supporting embedded database development.[1] Metaphrase functions are currently implemented as Perl CGI scripts executing on a web server.

The Metaphrase API extends the UMLS toolkit by providing additional information about the quality of a phrase match between an input string and existing concepts. In particular, the Metaphrase function V1 takes an input string and returns a collection of candidate matches from Meta, indicating in a bit mask which lexemes matched as well as a score indicating the fit of the match.

The 1270 candidate terms for the problem list vocabulary originate in the Online Medical Record (OMR), a heavily-used, ambulatory computer-based patient record system.[4] 37 of 76 outpatient clinics on the East Campus of the Beth Israel Deaconess Medical Center (the former Beth Israel Hospital) actively use OMR, and have entered over 140,000 problems entered between 1989 and July, 1996. The candidate terms comprise the problem list data dictionary, a compilation of problem labels developed over the years of operation of OMR. This source list includes ICD-9 labels, acronyms, and other user-defined problem phrases.[5]

The following working definition is used for clinical problems: *A clinical problem is a finding or process that requires consideration for diagnostic evaluation or therapeutic intervention.* To operationally implement the definition, a problem label must be a member of the subset of UMLS semantic types that represent findings or processes that meet the criteria set forward in the working definition. Those UMLS semantic types and their subtypes that comprise the type *<clinical problem>* include

- <finding>
- <pathologic function>
- <injury or poisoning>
- <anatomic abnormality>
- <individual behavior>
- <therapeutic or preventative procedure>

Matching was attempted for all 1270 candidate terms. Terms found to have either exact lexical matches with Meta concepts or to be synonymous through the Large-Scale Vocabulary Test methodology were considered for inclusion in the final vocabulary, termed BI96.[6] The semantic type of matched terms was reviewed against the semantic types comprising the type *<clinical problem>* to determine inclusion in version 1.0 of BI96.

Results

Eight of the original 1270 terms varied from other terms by insignificant white space and were eliminated as duplicates. A total of 75% of the unique candidate terms qualified for entry into version 1.0 of BI96. Of the 1262 unique candidate terms, 909 terms had unambiguous lexical matches to Meta concept strings and an additional 90 terms had synonymy facts from the Large-Scale Vocabulary trial, totaling 999 matches (79%). The remaining 263 terms (21%) were found to have either no lexical matches or were partial matches. From the 999 matching terms, 952 terms (95%) had semantic types in the subset of types comprising *<clinical problem>*, 13 terms (2%) did not map to the UMLS concepts of their lexical matches, and 34 terms (3%) had other non-complying types.

Of the 263 non/partially-matching terms, 143 (54%) were acronyms or abbreviations with no corresponding string entries in

1. Sleepycat Software Inc. 394 E. Riding Dr. Carlisle, MA 01741 http://www.sleepycat.com/

Meta. The 120 remaining terms (46%) were phrases with, at best, only partial matches. Of the 13 terms not mapping to the appropriate UMLS concept, 8 terms were acronyms with multiple meanings, 4 terms had "+" characters ignored by the matching routines (e.g., *hiv+),* and 1 term *(Paget's Disease)* was ambiguous in its meaning. Examples of the 34 matching terms not having semantic types compliant with *<clinical problem>* are presented in Table 1.

Table 1 - UMLS-matched terms with non-complying semantic types

UMLS Semantic Type	OMR Term
Subtype of <Organism>	chlamydia, hsv, human immunodeficiency virus, isospora
<Body Substance>	gallstones
<Substance>	chewing tobacco, cigarettes
<Immunologic Factor>	pneumovax
<Amino Acid, Peptide, Or Protein>	heptavax b, hgb ss
<Health Care Activity>	cardiovascular, health maintenance healthcare maintenance
<Population Group>	smoker, former smoker, hepatitis b carrier
<Organism Function>	pregnancy, menopause
<Mental Process>	anxiety, concentration, grief reaction

Discussion

A majority of the problem labels used in the OMR system exist as UMLS concepts. Because a significant number of candidate terms could not be matched because of the inclusion of acronyms or abbreviations, coverage of our problem labels by UMLS is likely to be higher than reported. Almost all of our matching terms had types consistent with our operational definition of *<clinical problem>*.

Candidate terms not recognized as "good" clinical problems raise interesting issues regarding requirements for PL-Tk, typing of existing UMLS terms, and the adequacy of our operational definition for clinical problems. Table 1 reveals that occasionally clinicians will use proxy terms as problem labels, i.e., organism for disease, substances for individual behavior, etc. While these labels are adequate for provider communication, they will be unacceptable as data elements for decision support or data analysis. Therefore, navigational support will be important within PL-Tk for application programmers to lead end-users from a non-problem entry point to a relevant problem term.

Obviously, PL-Tk relies heavily on the semantic typing of UMLS concepts in identifying clinical problems. The set of types needed to identify clinical problems is deliberately small in order to eliminate false-positives problems--terms with the correct semantic typing not employed as problem labels in clinical practice. False negative terms, therefore, must be evaluated for the adequacy of their typing within UMLS to determine whether the terms have been inappropriately typed or a deficiency exists within the operational definition for clinical problems. For example, we have currently excluded the commonly referenced concept *health care maintenance,* because it carries the broad type *<health care activity>*. However, because health care maintenance, well visits, routine care, etc., are best thought of as clinical screening procedures, it would be more appropriate to reclassify these terms as a type of *<therapeutic or preventative procedure>* which we recognize as a kind of problem label.

Exclusion of physiologic states such as pregnancy and menopause do illustrate a deficiency in our operational definition of clinical problem. Physiologic states clearly exist that are actively managed, but inclusion of the semantic type *<organism function>* in the definition of clinical problem would admit a significant number of false positive terms. An interesting general solution would be to augment the definition of clinical problem with a sufficiency test to be applied to terms falling outside recognized problem types. This test would permit a computational determination of problem status in addition to the structural determination (i.e., place in the type hierarchy). Such a test would require the addition of attributes or relations to the candidate concepts within UMLS, and remains an area for future exploration.

As others have found, UMLS has been a useful starting point for our investigation into creating domain-specific vocabulary services.[7] Our study reveals reasonable representation of a broad spectrum of terms that are used day-to-day within a working computer-based medical record. These terms appear to be classified in a manner that will be useful to the broad array of applications that support the functionality of a CPR. We have benefited from the broad coverage of clinical concepts and the utilities that accompany the datasets. The immutability of concept identifiers and scheme for concept merger will facilitate consistency of our problem-based services over time. The mappings of multiple source vocabularies will facilitate augmenting our services to support coding and literature search applications.

We will further explore the use of UMLS as a controlled vocabulary as we investigate the management of partially matched problem labels. Recent work enumerates limitations in using UMLS as a basis for a compositional vocabulary including partial coverage of clinical terminology, differing granularity of source vocabularies, differing compositional semantics of source vocabularies, and absence of guidelines for composing complex elements.[8] Ideally, the set of partially-matched problem labels could be composed from atomic elements in UMLS; it is least desirable to add significant numbers of local terms to BI96. The ability to constrain compositional problem labels based on semantic information available in Meta remains an open question.

Conclusion

As an initial task in the creation of the Problem List Toolkit, we have investigated the coverage in UMLS of problem labels used in our ambulatory computer-based patient record. We subsequently examined the compliance of matched problem labels with an operational definition derived to filter findings and processes that are typically the subject of diagnostic evaluation or therapeutic management in clinical practice. 79% of 1262 unique candidate problem labels were found to have lexical matches to UMLS concept strings, of which only 5% did not map to the UMLS concept of their lexical match or had semantic types not in compliance with our operational definition. The UMLS Metathesaurus provides a broad repertoire of pre-coordinated terms that are employed as problem labels in our ambulatory setting. Whether Metathesaurus concepts can be used to compose complex terms remains a topic for further investigation. Although our operational definition provides a structural basis for enumerating clinical concepts typically considered clinical problems, it is likely that the definition will have to be enhanced to completely cover the universe of clinical problems. However, the semantic type hierarchy can be used to create a consistent framework for a large segment of problem labels representing human pathology such that problem labels can be guaranteed to represent "good" clinical problems.

Acknowledgments

Supported in part by a cooperative agreement award from the Agency for Health Care Policy and Research (HS08749) and the National Library of Medicine
(U01 LM 05877)

References

[1] U.S. Department of Health and Human Services. *UMLS Knowledge Sources*. 8th Edition. January 1997.

[2] Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. *Meth Inform Med.* 1993; 32:281-291.

[3] McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. *Proc Ann Symp Comput App Med Care.* 1994;235-239.

[4] Safran C, Rury C, Rind DM, Taylor WC. A computer-based outpatient Medical record for a teaching hospital. *MD Comput* 1992;8:291-299.

[5] Zelingher J, Rind DM, Caraballo E, Tuttle MS, Olson NE, Safran C. Categorization of free-text problem lists: an effective method of capturing clinical data. *Proc Ann Symp Comput App Med Care.* 1995;416-420.

[6] McCray AT, Cheh ML, Bangalore AK, Rafei K, Razi AM, Divita G, Stavri PZ. Conducting the NLM/AHCPR Large Scale Vocabulary Test: A distributed Internet-based experiment. *JAMIA.* 1997;Suppl :560-564.

[7] Payne TH, Martin DR. How useful is the UMLS metathesaurus in developing a controlled vocabulary for an automated problem list? *Proc Ann Symp Comput App Med Care.* 1993;705-9.

[8] Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG Warren J. Phase II Evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. *JAMIA.* 1997;4:238-251.

Address for correspondence

Howard S. Goldberg, MD
Division Of Clinical Computing
Beth Israel Deaconess Medical Center
Harvard Medical School
350 Longwood Ave., GL-One
Boston, MA 02215
email: hgoldber@bidmc.harvard.edu