

Research Paper ■

Evaluation of a “Lexically Assign, Logically Refine” Strategy for Semi-automated Integration of Overlapping Terminologies

ROBERT H. DOLIN, MD, STANLEY M. HUFF, MD, ROBERTO A. ROCHA, MD, PHD,
KENT A. SPACKMAN, MD, PHD, KEITH E. CAMPBELL, MD, PHD

Abstract **Objective:** To evaluate a “lexically assign, logically refine” (LALR) strategy for merging overlapping healthcare terminologies. This strategy combines description logic classification with lexical techniques that propose initial term definitions. The lexically suggested initial definitions are manually refined by domain experts to yield description logic definitions for each term in the overlapping terminologies of interest. Logic-based techniques are then used to merge defined terms.

Methods: A LALR strategy was applied to 7,763 LOINC and 2,050 SNOMED procedure terms using a common set of defining relationships taken from the LOINC data model. Candidate value restrictions were derived by lexically comparing the procedure’s name with other terms contained in the reference SNOMED topography, living organism, function, and chemical axes. These candidate restrictions were reviewed by a domain expert, transformed into terminologic definitions for each of the terms, and then algorithmically classified.

Results: The authors successfully defined 5,724 (73%) LOINC and 1,151 (56%) SNOMED procedure terms using a LALR strategy. Algorithmic classification of the defined concepts resulted in an organization mirroring that of the reference hierarchies. The classification techniques appropriately placed more detailed LOINC terms underneath the corresponding SNOMED terms, thus forming a complementary relationship between the LOINC and SNOMED terms.

Discussion: LALR is a successful strategy for merging overlapping terminologies in a test case where both terminologies can be defined using the same defining relationships, and where value restrictions can be drawn from a single reference hierarchy. Those concepts not having lexically suggested value restrictions frequently indicate gaps in the reference hierarchy.

■ **JAMIA.** 1998;5:203–213.

Affiliations of the authors: Kaiser Permanente, Southern California, La Palma, California (RHD); Intermountain Health Care, Salt Lake City, Utah (SMH); Departamento de Cirurgia, Universidade Federal do Paraná, Paraná, Brazil (RAR); Department of Pathology, Oregon Health Sciences University, Portland, Oregon (KAS); and the Section on Medical Informatics, Stanford University School of Medicine, Stanford, California (KEC).

Supported in part by grant HS/LM08751 from the National Library of Medicine and the Agency for Health Care Policy and Research. Principal funding provided by Kaiser Permanente.

Correspondence and reprint requests to: Robert H. Dolin, MD, Kaiser Permanente, 5 Centerpointe Dr, La Palma, CA 90623. E-mail: (Robert.H.Dolin@kp.org).

Received for publication: 7/24/97; accepted for publication: 11/12/97.

Standardized definitions and standardized use of terminology in medical record systems are prerequisites for robust informatics applications such as automated decision support and outcomes analysis. The proper way to attain such standards and to formally define healthcare terminology is an area of active discussion and research.^{1–6} Often, proposed approaches seem to be polarized into two paradigms: lexically based, which relies on analysis of morphemic patterns within the terms being defined to derive meaning, and logically based, which relies on axiomatic definition of concepts and subsequent classification on those definitions to derive meaning.

We believe that pragmatic solutions to our terminology problems may be found in hybrid solutions that leverage the best characteristics of each approach. Here we describe an evaluation of a “lexically assign, logically refine” (LALR) strategy that combines the efficiency of linguistically based approaches with the precise semantics of description logic. We leverage lexical algorithms and language’s inherent structure to propose relationships for terms in a terminology. These proposed definitions are then reviewed and refined by domain experts, and subsequently loaded into a description logic classifier. The LALR strategy allows us to efficiently and formally define terms to the extent pragmatically possible with the concomitant benefits of decreased ambiguity and increased precision available to represent medical observations, diagnoses, and patient management plans.

Using our hybrid approach, we seek to demonstrate the practicality of a LALR strategy that we believe can scale to answer the terminologic challenges posed by the Computer-Based Patient Record Institute⁷ and the Institute of Medicine.⁸ Although our test case focuses on the semi-automated merging of overlapping terminologies, we believe that the LALR approach can be generally applied to a wide variety of terminologic challenges.

We selected the SNOMED (Systematized Nomenclature of Human and Veterinary Medicine) laboratory procedures⁹ and the LOINC (Logical Observation Identifier Names and Codes) laboratory procedures¹⁰ as our test case because the terminologies are individually important to our respective organizations, and the two terminologies have complementary strengths.

SNOMED III is the foundation of Kaiser Permanente’s Convergent Medical Terminology project,¹¹ and has been shown to have broad content coverage in many clinical domains.^{12,13} Despite SNOMED’s broad coverage of medical concepts, it is missing many of the detailed laboratory terms necessary for standardizing reporting of laboratory information, although it provides a hierarchical classification of laboratory procedures.

The LOINC database contains finely detailed test result identification codes suitable for standardized reporting of laboratory information and has been endorsed by the American Clinical Laboratory Association (ACLA). ACLA recommends that all members report laboratory results using LOINC codes (ACLA members account for approximately 70% of the volume of tests performed in the United States). Additionally, the Health Care Financing Administration is constructing ICD-10-PCS codes using LOINC terms.¹⁴

Despite LOINC’s finely detailed identification codes and consensus regarding its adoption, LOINC lacks a hierarchical organization of those identification codes. We chose to test our LALR strategy using LOINC and SNOMED because we felt that a combination of LOINC and SNOMED could produce a comprehensive strategy for representing detailed laboratory terms as well as appropriately classify those terms.

Background

Systematized Nomenclature of Medicine (SNOMED)

SNOMED III is a multiaxial hierarchical coding scheme.⁹ Terms are assigned to one of 11 independent systematized axes: Topography, Morphology, Function, Living Organisms, Chemicals, Physical Agents, Occupations, Social Context, Diseases, Procedures, and General Linkage/Modifiers. The code assigned a term places it into its proper location in a monohierarchy, thereby conveying contextual information for each term (e.g., Tuberculous pneumonia, DE-14817, is a Respiratory tuberculosis, DE-14810, which is a Tuberculosis, DE-14800, which is a Bacterial infectious disease, DE-10000, which is an infectious disease, DE-00000). SNOMED concepts are further defined through the use of cross-referencing to related concepts in other axes (e.g., Tuberculous pneumonia, DE-14817, affects the topographic region Lung, T-28000, has a morphologic feature of inflammation, M-40000, and an etiologic agent of *Mycobacterium tuberculosis hominis*, L-21801). In the July 1996 release of SNOMED, version 3.3, approximately 25% of the relevant clinical terms included cross references.¹⁵

Logical Observation Identifier Names and Codes (LOINC)

The LOINC database provides a set of universal names and codes for identifying laboratory test results.^{10,16} Initially developed in mid-1995, the August 1996 version contains approximately 8,200 codes. Each result is defined as having a measured component or analyte (e.g., potassium), a property being measured (e.g., mass), whether the measurement is a momentary observation at a point in time or an observation integrated over time (e.g., 24-hour collection), the type of system or sample (e.g., serum), the type of scale (e.g., quantitative), and, where relevant, the method used to produce the result or other observation. In addition, each concept is placed into a category or class (e.g., Chemistry, Microbiology). Thus, each LOINC concept may have up to six defining relationships: Analyte, Property, Time Aspect, System, Scale, and possibly a Method. A non-defining at-

tribute of a LOINC concept is its Class. The LOINC data model is illustrated in Figure 1, using Coad and Yourdon's modeling representation.¹⁷

Merging and Mapping Techniques

Prior efforts at automated mapping and merging of overlapping terminologies have relied primarily on two different classes of techniques: lexical, which base their comparisons on morphemic components of the terms themselves, and logical, which base their comparisons on classification of the axiomatic definition of terms.

Lexical Techniques

Lexical algorithms typically perform an initial normalization process on the terms to be compared in an attempt to abstract away from lexical (syntactic, morphologic, and orthographic) variation in terms.^{2,5,6,18,19} The process may involve breaking each string into its constituent words and morphemes; lower-casing each word; removing punctuation, stop words, and duplicate words; and sorting the words in alphabetical order (e.g., the term "Cell count of synovial fluid with differential count" would become "cell count differential fluid synovial").

More sophisticated normalization techniques attempt to convert each word to its canonical uninflected form (e.g., "treats" or "treated" becomes "treat"),^{5,18} expand abbreviations,⁵ attempt to correct spelling errors,²⁰ and substitute preferred terms for synonyms (e.g., "kidney" becomes "renal").^{2,5,21} Rocha et al.^{18,19} lexically

compared normalized word fragments (digrams) using a similarity function. Digram comparisons break each word into fragments of two letters each for comparison. For example, the word "morphemic" would be broken into the fragments mo, or, rp, ph, he, em, mi, and ic for comparison with the digrams from another word. This technique is less sensitive to minor morphologic variations in words introduced by misspellings or other normal variations in morphology. Digrams are created algorithmically and do not require any knowledge about word formation rules, and do not rely upon the existence of affix dictionaries.

Following the normalization process, terms are compared using various techniques such as exact normalized string matching, the Longest Common Substring algorithm,²² or a similarity scoring approach.¹⁸ Many of these techniques are included in the National Library of Medicine's lexical variant generation (lvg) tools, which are available to anyone who has signed the UMLS license agreement.²³

Logic-based Techniques

Logic-based techniques are used to formally define concepts and to map and classify terms based on similarities of their definitions.²⁴⁻²⁷ So, for example, if two distinct and lexically unrelated terms each define a type of laboratory procedure that measures sodium, the two terms will be appropriately classified together. There has been a general trend in terminology research to make more explicit the definition of medical concepts,^{4,19,27-29} both to disambiguate concepts and to facilitate automated terminology merging efforts. The Canon group has argued for a medical-concept representation language containing semantic types for each concept and a network of general medical concepts in which implicit relations between elements are made explicit.⁴

There has been an evolution over time in the formal representation of medical concepts. An early effort to abstract away from specific terms to term representations for use in modeling general medical language was developed by Evans in the MedSORT project.²⁸ To represent concepts, MedSORT required both a semantic classification scheme and rules that determined how elements in the classification scheme might combine. Masarie et al.²⁵ defined concepts from several vocabularies using a frame-based system, where the definitions were based on an analysis of the terms to be merged. Cimino et al.²⁷ have used a semantic network for defining the concepts in the Columbia-Presbyterian MED vocabulary. Their network is a notation for representing conceptual entities and links between them, allowing the storage of factual knowledge that can be intensional (describing the

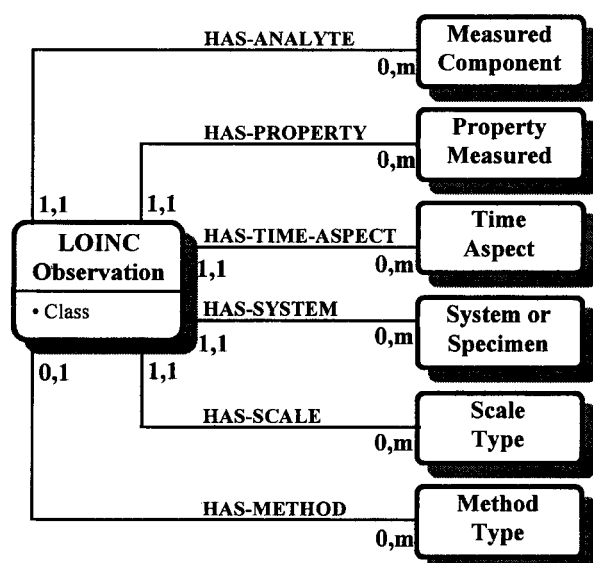


Figure 1 The LOINC (Logical Observation Identifier Names and Codes) data model. (See the Background section for details.)

entities themselves) and extensional (describing how entities are related to other entities), and includes a classification hierarchy. Description logic, which defines concepts based on formal logic theory, was developed to make explicit the semantics of frame-based systems* and semantic networks while retaining an emphasis on taxonomic structure as an organizing principle.³⁰

Campbell et al.³ have previously argued that formal logic can provide a framework for formalizing the representation of medical concepts. It is well recognized that there is a tradeoff in expressiveness and tractability of term classification that has to be considered when using formal logic.^{31,32} Rector et al.³³ have argued for a more expressive formalism, claiming that worst-case computational complexity is an inappropriate guide to the choice of formalism. On the other hand, several description logics have been defined that support less than full first-order logic, and offer complete and sound algorithms for the classification of terms defined by the logic.^{34,35} Such a subset, used in this study, is supported by K-Rep, a description logic classification and terminology authoring system developed by researchers at IBM.³⁶ Expressiveness is deliberately constrained in the K-Rep system so that algorithmic classification is guaranteed to be sound and complete. While this constraint may pose a limitation in the definition of certain concepts requiring, for example, a logical “or” or a logical “not,” this was not an issue in the definition of the concepts in this study, all of which could be fully defined. The explicit representation of medical concepts in a description logic enables the formal determination of the expressiveness and tractability of a particular representational syntax.

Logic-based classification algorithms typically organize a set of concepts into a taxonomy based on axiomatic term definitions. As an initial step, the defining characteristics for the set of concepts to be classified must be determined. A LOINC concept, as shown in Figure 1, is defined by its analyte, measured property, measurement time aspect, specimen, scale, and method. Cimino et al.²⁷ defined laboratory specimens, laboratory tests, and medications based on relationships taken from the UMLS Semantic Network.³⁷ Once the defining relationships have been determined, appropriate value restrictions for each defining relationship must be assigned. These values can be deter-

mined by a combination of manual assignment, lexical suggestion, or logical inheritance using the methods described above.

Once concepts are defined, logic-based classification algorithms are employed.^{27,32,34,35} Cimino et al.^{27,38} have provided a description of their classification algorithms, which are used to deepen a hierarchy and to identify the most appropriate location for a given term in a given classification hierarchy. Description logic classification algorithms are discussed in more detail by Fitting.³⁹

Prior experience with logic-based techniques suggests that these methods may enhance results achievable with lexical techniques alone. Here we present our evaluation of a specific strategy for combining logic-based and linguistic-based techniques: Lexically Assign, Logically Refine (LALR).

Methods

Definitions

This section defines the terms “defining relationship” and “value restriction,” which are used throughout this report to specify how we are defining healthcare terminology concepts. Defining relationships (also known as “defining roles”) are those relationships that link a concept to its defining characteristics. In Figure 1, for instance, the defining relationship HAS-ANALYTE indicates that a LOINC Observation is in part defined by its analyte or measured component. A value restriction (also known as a “role restriction” or “relationship value”) is the target value that restricts the domain of a defining relationship. In Figure 3, for instance, the concept HLA-A-serotyping.P3-68530 has a value restriction of HLA-A-Antigen-NOS_F-C4100 for the HAS-ANALYTE defining relationship. This value restriction in effect states that any concept in the terminology database subsumed by HLA-A-Antigen-NOS_F-C4100 is a valid analyte for those concepts subsumed by HLA-A-serotyping-NOS.P3-68530. The value restrictions on a concept are the logical intersection of the value restrictions stated in the definition of that concept and in the definitions of the transitive closure of its parent concepts.

Test Data Set

Of the 8,182 concepts in the August 1996 version 1.0h LOINC database, 419 (those in LOINC Classes CLIN, MISC, and DRUGDOSE) were excluded due to lack of overlap with any SNOMED Procedure categories, leaving a total of 7,763 LOINC concepts for study. 2,317 concepts are present in the SNOMED version 3.3 Procedure Axis, Chapter P3 Laboratory Procedures.

*Many frame-based systems do not have formal semantics defined by description logic. We prefer to reserve the phrase “frame-based” for systems where the semantics of their classification is not clear, and describe systems that are consistent with formal logic theory as “description-logic-based.”

Of these, 267 (section P3-0 General Laboratory Procedures and Services, subsection P3-53 Microbial Identification Kit Methods, and subsection P3-70 Chemistry Methods) were excluded because they contain concepts not readily adaptable to definition using the LOINC model, for a total of 2,050 SNOMED concepts for study.

Defining Relationships Are Taken from the LOINC Model

As noted above and shown in Figure 1, a LOINC concept is defined by its analyte, property, time aspect, system, scale, and method. From this model, the corresponding defining relationships became HAS-ANALYTE, HAS-PROPERTY, HAS-TIME-ASPECT, HAS-SYSTEM, HAS-SCALE, and HAS-METHOD.

Value Restrictions Are Drawn from SNOMED

All LOINC concepts are assigned an analyte, property, time aspect, specimen, scale type, and method by the LOINC committee. In a prior study, these components of the 7,763 LOINC concepts in our test data set were lexically mapped to SNOMED concepts,¹⁹ using the methods described by Rocha et al.¹⁸ These SNOMED mappings formed the value restrictions for the LOINC concepts in this study.

Lexical techniques were used to suggest value restrictions for the SNOMED test data set. The set of SNOMED concepts was cross-referenced with each of the other SNOMED axes looking for potential lexical matches that might provide an appropriate restriction for any of the defining relationships. Prior to comparison, term strings in the SNOMED test data set and in the target SNOMED axis were normalized (including all SNOMED synonyms) using the norm program available from the National Library of Medicine.²³ A detailed description of this process is provided by McCray et al.,⁵ and includes breaking each string into its constituent words, lower-casing each word, removing punctuation and stop words, sorting the words in alphabetical order, and conversion of each word to its uninflected form. In addition, certain patterns, such as "oscope/oscopy" were identified as high-yield for detecting matches. Comparison of the normalized concepts included partial phrase matching (e.g., "Alkaline phosphatase isoenzymes measurement" would pair with "Alkaline phosphatase isoenzyme").

Potential matches were placed into a report for manual review. A single reviewer (RHD) examined the report to determine whether or not the lexically suggested target SNOMED term would be accepted as a value restriction for the corresponding test SNOMED

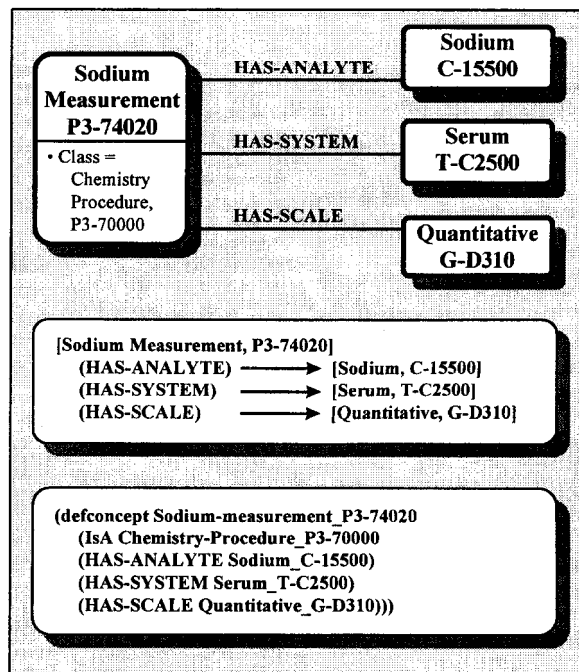


Figure 2 A sample concept defined according to the LOINC model (of Figure 1). The middle section shows the concept defined in conceptual graph notation, while the bottom section shows a similar definition in description logic syntax.

concept, and to indicate the defining relationship for the restriction.

Our primary focus was on defining each concept at least with respect to its measured component (i.e., determining the concept filling the HAS-ANALYTE value restriction) because procedure names frequently contain some lexical variant of the substance being analyzed, thereby greatly enabling a lexical approach, and because a classification based on measured components results in a clinically sensible hierarchical structure.

Terminologic Definitions Are Expressed in a Description Logic Syntax

We used a Knowledge Representation Syntax Specification (KRSS)⁴⁰-based description logic syntax to represent terminologic definitions. Once these definitions were created, they were classified by the K-Rep system.³⁶ The syntax allows us to express a concept's parents (i.e., specify "IsA" relationships), defining relationships and their restrictions, and non-defining attributes of a concept (such as its SNOMED code). Additional expressiveness supported by the syntax, such as cardinality constraints, was not required to be used. Figure 2 shows how a sample LOINC concept might

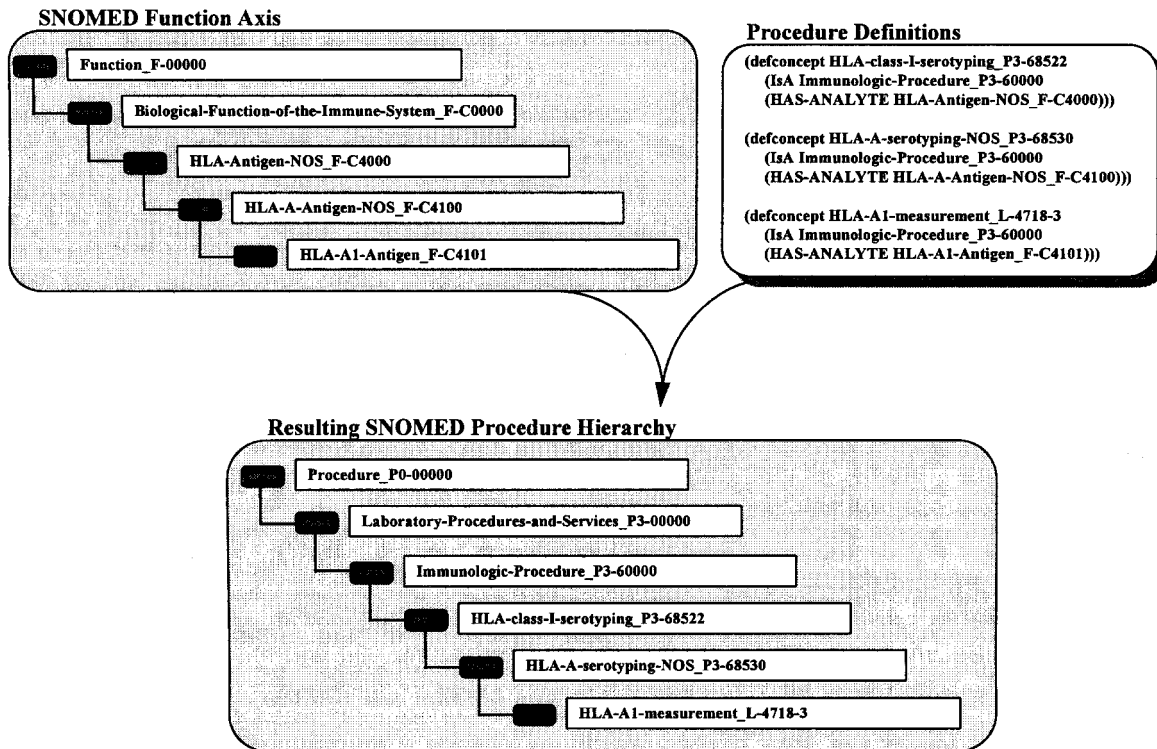


Figure 3 Algorithmic classification based on measured component. Given the analyte hierarchy for the SNOMED Function Axis and the concept definitions shown, the resulting SNOMED Procedure hierarchy is automatically generated. (Concept HLA-A1-Measurement_L-4718-3 is from LOINC. The other two concepts are from SNOMED.)

be represented in conceptual graph notation, and a similar KRSS-like representation.

All target SNOMED axes (i.e., all SNOMED axes that contain value restrictions for concepts in the SNOMED and LOINC test data sets), are also defined in the description logic and included in the terminology database. In general, these axes are comprised of “primitive” concepts—concepts that cannot be fully defined, and are instead placed into a manually constructed hierarchy. The hierarchical structure of these axes exist within the SNOMED database, and was reflected in the description logic via the specification of “IsA” relationships. For example, Figure 3 shows a representation of the SNOMED Function Axis, in which HLA-A-Antigen-NOS_F-C4100 IsA HLA-Antigen-NOS_F-C4000.

Defined Concepts Are Conceptually Merged

The initial merge placed LOINC concepts into corresponding SNOMED Procedure categories (e.g., Chemistry-Procedure_P3-70000, Microbiology-Procedure_P3-50000), based on the LOINC Class. (This choice was arbitrary, and subsequent techniques would have been equally applicable had we instead placed SNO-

MED concepts into the LOINC classes.) The mapping between the 18 LOINC classes and the nine corresponding SNOMED Procedure categories was performed manually. As a result, the IsA relationship for each concept was set equal to the name of its corresponding SNOMED Procedure category. All subsequent classification within a category was algorithmically determined, based on the logical definition of each concept.

A detailed description of classification algorithms similar to those used by K-Rep for this report have previously been published.³² The initial step involves parsing the textual concept definitions for syntactic correctness. Next comes a process of normalization that converts the definitions into a standard form. This is followed by completion, during which value restrictions are inherited from the transitive closure of all parent concepts. These inherited restrictions are logically intersected with local restrictions contained in the concept definition to form a completely explicit concept definition. From there, a concept is automatically integrated into the taxonomy by comparing its definition with definitions of concepts already in the taxonomy. The process of determining a new concept’s location in the taxonomy is called classification.

When one concept is more general than another, the first concept properly subsumes the second. This can arise if the restriction of a defining relationship of one concept subsumes the restriction of the same relationship of another concept, as represented in the target SNOMED axis, or when the second concept's definition expresses additional constraints beyond those of the first concept. When two concepts are equivalent, they subsume one another, but neither properly subsumes the other. For one concept to subsume another, every value restriction of the first concept must also be true of the second concept.

Results

Lexically Suggested Value-restriction Determination

As noted above, all LOINC concepts are assigned an analyte by the LOINC committee. In a prior study, the analytes of 5,724 (74%) of the 7,763 LOINC concepts in our test data set were lexically mapped to SNOMED concepts,¹⁹ using the methods described by Rocha et al.¹⁸ (Table 1). Examining the LOINC concepts in another way, there are a total of 4,191 unique analytes among the 7,763 LOINC concepts studied, of which 2,241 (53%) could be mapped to a SNOMED concept. There are 53 unique properties, 12 (23%) of which could be mapped to SNOMED; eight unique time aspects, none of which mapped; 83 unique system or specimen types, 61 (73%) of which mapped; four unique scale types, two (50%) of which mapped; and 140 unique method types, 73 (52%) of which mapped to a SNOMED concept.

Previously, SNOMED concepts had not been cross-referenced to their corresponding analytes. Lexical algorithms in this study suggested 3,852 restrictions for the HAS-ANALYTE relationship, of which 1,151 (30%) were accepted as valid, resulting in 56% of the SNOMED concepts being assigned a HAS-ANALYTE value restriction (Table 1). These restrictions were found in several SNOMED axes, including Topography (e.g., Eosinophil, T-C1340), Function (e.g., 11-Deoxycortisol, F-B2480), Living Organisms (e.g., *Bordetella pertussis*, L-12801), and Chemicals (e.g., Acetaminophen, C-60130). 472 restrictions for the HAS-SYSTEM role and 45 restrictions for the HAS-SCALE-TYPE role were also determined. No restrictions for the HAS-PROPERTY, HAS-TIME-ASPECT, or HAS-METHOD-TYPE roles were found using our lexical approach.

Overall, of the 9,813 concepts in the combined SNOMED and LOINC test set, 6,875 (70%) had a restriction for the HAS-ANALYTE relationship determined by lexical techniques.

Terminology Merging and Algorithmic Classification

As mentioned above, the initial merge placed LOINC concepts into corresponding SNOMED Procedure categories (e.g., Chemistry-Procedure_P3-70000, Microbiology-Procedure_P3-50000), based on the LOINC Class. All subsequent classification within a category was algorithmically determined, based on the logical definition of each concept. Of the 9,813 concepts in the test data set, the 6,875 whose measured component was able to be lexically determined were then merged and algorithmically classified. Because each

Table 1 ■

Lexical Determination of the Measured Components or Analytes of LOINC and SNOMED Concepts

| SNOMED Section | LOINC | | SNOMED | | | Totals | |
|--------------------|--------------|-------------------------------|--------------|------------------------------|-------------------------------|--------------|----------------------------------|
| | Concepts (n) | SNOMED Analyte Determined (n) | Concepts (n) | SNOMED Analyte Suggested (n) | SNOMED Analyte Determined (n) | Concepts (n) | SNOMED Analyte Determined (n, %) |
| P3-1 Coagulation | 211 | 184 | 118 | 247 | 89 | 329 | 273 (83) |
| P3-2 Blood bank | 540 | 480 | 131 | 370 | 78 | 671 | 558 (83) |
| P3-3 Hematology | 359 | 293 | 95 | 312 | 59 | 454 | 352 (78) |
| P3-4 Anatomic path | 1 | 0 | 232 | 274 | 27 | 233 | 27 (12) |
| P3-5 Microbiology | 636 | 476 | 300 | 251 | 74 | 936 | 550 (59) |
| P3-6 Immunology | 2901 | 1491 | 429 | 877 | 234 | 3330 | 1725 (52) |
| P3-7 Chemistry | 3045 | 2747 | 681 | 1441 | 577 | 3726 | 3324 (89) |
| P3-9 Urine & semen | 70 | 53 | 43 | 76 | 13 | 113 | 66 (58) |
| P3-A Food analysis | 0 | 0 | 21 | 4 | 0 | 21 | 0 (00) |
| Totals | 7763 | 5724 | 2050 | 3852 | 1151 | 9813 | 6875 (70) |

NOTE: LOINC concepts have been merged into SNOMED sections based on the class of each LOINC concept.

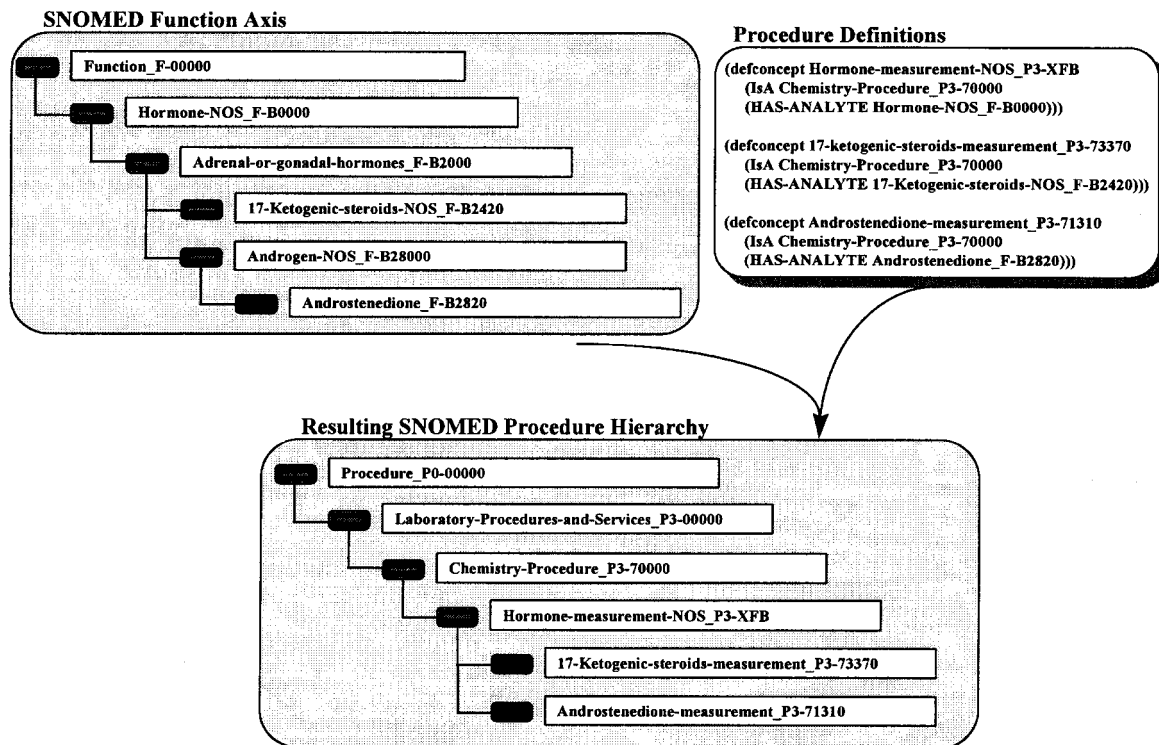


Figure 4 Faulty classification due to reference hierarchy. Classification errors in the reference Function Axis are mirrored in the automatically generated Procedure Axis. In this case, Androstenedione_F-B2820 also IsA 17-Ketogenic-steroids_F-B2420. Making this correction in the Function Axis will automatically correct the Procedure classifications.

concept is defined with respect to its measured component, and because each measured component is a SNOMED concept, the resulting algorithmic classification hierarchy assumed the structure of the corresponding analyte hierarchies. This is illustrated in Figure 3.

Faulty classification can result from errors in the reference hierarchy. This is illustrated in Figure 4. In this example, Androstenedione_F-B2820 is modeled as a child of Androgen-NOS_F-B2800, but should also be modeled as a child of 17-Ketogenic-steroids_F-B2420. Making this manual correction in the Function Axis will automatically correct the classification of concepts in the Procedure Axis.

When two concepts measure the same component (i.e., have the same restriction for the HAS-ANALYTE relationship), the determinant of the parent-child relationship is determined by the other defining relationships and value restrictions. More formally, if one concept is more general than another, the first concept is said to properly subsume the second. In such a case, the second concept's definition expresses additional constraints beyond those of the first concept. In the hierarchy, the second concept will appear as a child of the first concept. This is shown in Figure 5.

Of the SNOMED procedures in the test database, 80% share some measured component with a LOINC concept, and as a result will acquire a LOINC concept as a child. Because the initial merge combines LOINC and SNOMED concepts into SNOMED Procedure categories, all LOINC concepts acquire a SNOMED parent. Of the LOINC concepts, 50% share some measured component with a SNOMED concept, and as a result will also be classified under a more granular SNOMED concept (as in Figure 5). The remaining LOINC concepts potentially can form a flat hierarchy and an algorithm similar to that used by Cimino et al.^{27,38} can be used to deepen these areas. A flat hierarchy can also result where there are a number of specific tests without a more general test defined. This is shown in Figure 4, where the concept Hormone-measurement-NOS_P3-XFB has been created based on the structure of the reference hierarchy to deepen the Procedure hierarchy.

Discussion

We found the LALR strategy works well for automatically merging overlapping terminologies if the terminologies to be merged can be logically defined

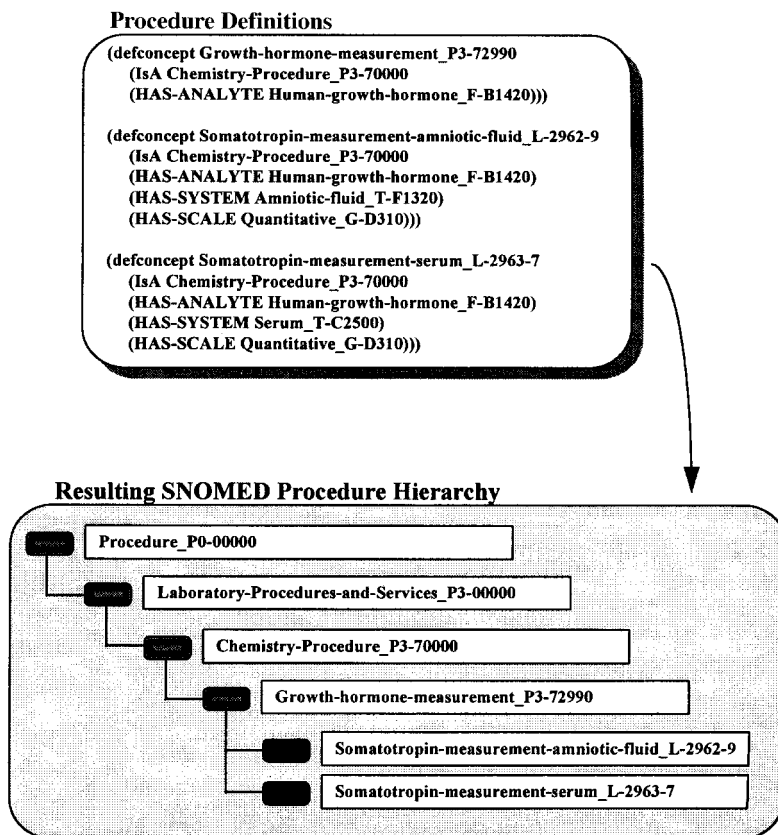
using the same model (i.e., both terminologies can be defined using the same defining relationships), and if their value restrictions can be drawn from the same reference hierarchy. Key points in our approach are that concept definitions are expressed in a formal description logic, defining relationships are taken from a pre-established model (LOINC), lexical techniques are used in the determination of value restrictions, and value restrictions for the terminologies to be merged are drawn from the same reference terminology (SNOMED). Generalizing this approach outside the realm of laboratory procedures will rely on the presence of defining models for concepts in other domains, and a rich reference terminology that is likely to provide lexically suggested value restrictions for concepts in the terminologies to be merged.

Blois⁴¹ has described a hierarchical schema of the sciences, ranging from the low-level sciences (e.g., mathematics, chemistry, and physics), which lend themselves to formalization over a relatively small domain of possible values and can therefore be more precisely defined, to the more abstract sciences (e.g., psychiatry and social science), which in part are built up from the underlying sciences, and are less able to be precisely defined. This would suggest that the application of description logic and classification might in-

herently be easier to implement for certain domains (where concepts can be defined) than for other domains. Consistent with Blois's assertions are the emergence of models defining surgical procedures, human anatomy, and medications,^{1,14,27,42} although models defining mental health, rehabilitation, and human disease states are also being developed.^{14,43} The emergence of these models will support the extension of our methods outside the scope of laboratory procedures.

Within the LOINC database, concepts filling the HAS-PROPERTY relationship are drawn from the International Union of Pure and Applied Chemistry (IUPAC),⁴⁴ and those filling the HAS-SYSTEM relationship are drawn from ASTM E1238-94⁴⁵ and HL7 version 2.2.⁴⁶ Lexical techniques were then used to map these concepts to synonymous SNOMED concepts. If evolving terminologies were to be defined from the outset based on atomic SNOMED building blocks, subsequent merging efforts could be substantially decreased, not to mention the beneficial impact this might have on increasing the overall content coverage of SNOMED. Our test database contained 2,039 LOINC concepts that could not be lexically mapped to a SNOMED analyte. To allow these LOINC concepts to be fully defined, 590 new concepts were

Figure 5 Classification of laboratory procedures measuring the same substance. When one concept is more general than another, the first concept will properly subsume the second. Given the concept definitions shown, the resulting SNOMED Procedure hierarchy is automatically generated. (Concept Growth-hormone-measurement_P3-72990 is from SNOMED. The other two concepts are from LOINC.)



added to SNOMED 3.4 (Kent Spackman, personal communication). Such enhancements to the overall content coverage of SNOMED are likely to benefit subsequent applications of our approach.

Our methodology suggests the need for different manual tasks when modeling concepts that can be logically defined than when modeling those that cannot. The reference SNOMED hierarchies that furnished value restrictions (such as Chemicals, Living Organisms) are generally composed of "primitive" concepts, meaning concepts that are unable to be fully defined with defining relationships. Manual tasks may include hierarchy construction and refinement as well as the addition of new concepts. For those concepts that can be logically defined (such as laboratory procedures), manual tasks may be better directed at refining concept definitions, including the review of lexically suggested restrictions, and adding new reference terms where necessary to allow for expression of value restrictions. Emerging environments such as Gálapagos¹¹ will enable the comparison of manually and automatically generated hierarchies to cross validate both the term definitions and the structure of the reference hierarchies.

Cimino et al.³⁷ have previously described a set of terminology design criteria. While the use of description logic does add formalism to terminology, allowing for the greater fulfillment of these criteria, it also will require a consideration of the optimal balance between expressiveness and tractability (of classification), similar to such considerations in other areas of medical informatics and computer science in general.^{32,47} Criteria for a healthcare terminology include, among others, the ideas of nonvagueness and nonredundancy. Nonvagueness is achieved when concepts in the terminology are complete in meaning. Redundancy exists when multiple terms for the same concept are added to the terminology as unique concepts. If the achievement of nonvagueness requires the expressive power of full first-order logic, then the automatic determination of redundancy can become intractable or undecidable. The emergence of a workable terminology that is applied to real clinical situations and to pooled clinical data should help illustrate these tradeoffs.

The authors thank Simon P. Cohn, MD, MPH (Kaiser Permanente Healthcare), and John E. Mattison, MD (Kaiser Permanente, Southern California Region), for their support of this work; Stephanie Lipow, of Lexical Technology, Incorporated, for her help with the lexical processing; the JAMIA reviewers for their detailed insightful comments; and the members of the Kaiser Permanente Interregional Convergent Medical Terminology group for their constant reminder that nothing is without controversy, and that through controversy comes convergence.

References ■

1. Rector AL, Glowinski AJ, Nowlan WA, Rossi-Mori A. Medical-concept models and medical records: an approach based on GALEN and PEN&PAD. *J Am Med Informatics Assoc.* 1995;2:19–35.
2. Barrows RC Jr, Cimino JJ, Clayton PD. Mapping clinically useful terminology to a controlled medical vocabulary. *Proc SCAMC.* 1994:211–5.
3. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. *J Am Med Informatics Assoc.* 1994;1:218–32.
4. 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.
5. McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. *Proc SCAMC.* 1994;235–9.
6. Sheretz DD, Tuttle MS, Olson NE, Erlbaum MS, Nelson SJ. Lexical mapping in the UMLS Metathesaurus. *Proc SCAMC.* 1989:494–9.
7. Computer-Based Patient Record Institute. Proposal to accelerate standards for computer-based patient record systems. Version 3.0. CPRI, April 29, 1994.
8. Institute of Medicine, Committee on Regional Health Data Networks. Health data in the information age: use, disclosure, and privacy. Washington, DC: National Academy Press, 1994.
9. Cote RA, Rothwell DJ, Beckett RS, Palotay JL (eds). SNO-MED International—The systematized nomenclature of human and veterinary medicine. Northfield, IL: College of American Pathologists, 1993.
10. Forrey AW, McDonald CJ, DeMoore G, et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clin Chem.* 1996;42:81–90.
11. Campbell KE, Cohn SP, Chute CG, Gennels G, Shortliffe EH. Gálapagos: computer-based support for evolution of a convergent medical terminology. *J Am Med Informatics Assoc. Symposium supplement* 1996:269–73.
12. Board of Directors of the American Medical Informatics Association. Standards for medical identifiers, codes, and messages needed to create an efficient computer-stored medical record. *J Am Med Informatics Assoc.* 1994;1:1–7.
13. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. *J Am Med Informatics Assoc.* 1996;3:224–33.
14. Averill RF, Mullin RL, Steinbeck BA, Goldfield NI, Grant T. The development on the ICD-10 procedure coding system (ICD-10-PCS). Draft, June 5, 1996. 3M Health Information Systems, Wallingford, CT.
15. Rothwell DJ, Cote RA. Managing information with SNO-MED: understanding the model. *J Am Med Informatics Assoc. Symposium supplement* 1996:80–3.
16. Logical Observation Identifier Names and Codes (LOINC) World Wide Web Home Page. (<http://www.mcis.duke.edu/standards/HL7/termcode/loinclin/loinclin.html>) (or search for "LOINC" from the HL7 Home Page at (<http://www.mcis.duke.edu/standards/HL7/hl7.htm>)).
17. Coad P, Yourdon E. Object-oriented Analysis. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.
18. Rocha RA, Huff SM. Using digrams to map controlled medical vocabularies. *Proc SCAMC.* 1994:172–6.
19. Rocha RA, Huff SM. Coupling vocabularies and data structures: lessons from LOINC™. *J Am Med Informatics Assoc. Symposium supplement* 1996:90–4.

20. Peterson JL. Computer programs for detecting and correcting spelling errors. *Communications of the ACM*. 1980;23:676–87.
21. Zeng Q, Cimino JJ. Mapping medical vocabularies to the Unified Medical Language System. *J Am Med Informatics Assoc. Symposium supplement* 1996:105–9.
22. Friedman C, Sidelli R. Tolerating spelling errors during patient validation. *Comput Biomed Res*. 1992;25:486–509.
23. National Library of Medicine's Unified Medical Language System (UMLS) World Wide Web Home Page. (<http://www.nlm.nih.gov/research/umls/UMLSDOC.HTML>) (or search for "UMLS" from the National Library of Medicine's Home Page at (<http://www.nlm.nih.gov/>)).
24. Rocha RA, Rocha BHSC, Huff SM. Automated translation between medical vocabularies using a frame-based interlingua. *Proc SCAMC*. 1993:690–4.
25. Masarie FE Jr, 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.
26. Cimino JJ, Barnett GO. Automated translation between medical terminologies using semantic definitions. *MD Comput*. 1990;7:104–9.
27. 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.
28. Evans DA. Pragmatically-structured, lexical-semantic knowledge bases for unified medical language systems. *Proc SCAMC*. 1988:169–73.
29. Rector AL, Nowlan WA, Glowinski A. Goals for concept representation in the GALEN Project. *Proc SCAMC*. 1993:414–8.
30. Russel SJ. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice–Hall, 1995.
31. Doyle J, Patil RS. Two theses of knowledge representation: language restrictions, taxonomic classification, and the utility of representation services. *Artificial Intelligence*. 1991;48:261–97.
32. Levesque HJ, Brachman RJ. Expressiveness and tractability in knowledge representation and reasoning. *Comput Intel*. 1987;3:78–93.
33. Rector AL, Nowlan WA, Kay S. Conceptual knowledge: the core of medical information systems. *MEDINFO*. 1992:1420–6.
34. Borgida A, Brachman RJ, McGuinness DL, Alperin Resnick L. CLASSIC: a structural data model for objects. *SIGMOD Record*. 1989;18:58–67.
35. Brachman RJ, Schmolze JG. An overview of the KL-ONE knowledge representation system. *Cogn Sci*. 1985;9:171–202.
36. Mays E, Weida R, Dionne R, et al. Scaleable and expressive medical terminologies. *J Am Med Informatics Assoc. Symposium supplement* 1996:259–63.
37. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. *Meth Inform Med*. 1993;32:281–91.
38. 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. *Proc SCAMC*. 1990:175–9.
39. Fitting M. *First-order Logic and Automated Theorem Proving*. New York: Springer-Verlag, 1990. (Gries D, ed. *Texts and Monographs in Computer Science*).
40. Patel-Schneider PF, Swartout B. Working version (draft): Description logic specification from the KRSS effort. June 30, 1993. (<http://www-ksl.stanford.edu/knowledge-sharing/papers/dl-specs.ps>).
41. Blois MS. Medicine and the nature of vertical reasoning. *N Engl J Med*. 1988;318:847–51.
42. CEN (European Committee for Standardization) Technical Committee 251: Medical Informatics. Structure for classification and coding of surgical procedures. Interim Draft. CEN/TC251/PT2-002S. November 1994. (URL (http://miginfo.rug.ac.be:8001/centc251/prestand/wg2/pt2_002s/pt2_002s.htm), or search from the CEN/TC 251 homepage at (<http://miginfo.rug.ac.be:8001/index.htm>)).
43. Brown PJB, O'Neil M, Price C. Semantic representation of disorders in Version 3 of the Read Codes. *Proceedings of the International Medical Informatics Association, Working Group 6*. January 1997, Jacksonville, FL: 209–14.
44. International Union of Pure and Applied Chemistry/International Federation of Clinical Chemistry. *The Silver Book: Compendium of Terminology and Nomenclature of Properties in Clinical Laboratory Sciences*. Oxford, England: Blackwell Scientific Publishers, 1995.
45. ASTM E1238-94. *Standard Specification for Transferring Clinical Observations between Independent Computer Systems*. Philadelphia, PA: American Society for Testing Materials, 1994.
46. Health Level Seven. *An application protocol for electronic data exchange in healthcare environments. Version 2.2*. Ann Arbor, MI: Health Level Seven, 1994.
47. Dolin RH. Expressiveness and query complexity in an electronic health record data model. *J Am Med Informatics Assoc. Symposium supplement* 1996:522–6.