SNOMED RT: A Reference Terminology for Health Care

Kent A. Spackman, M.D., Ph.D.[1,2], Keith E. Campbell, M.D., Ph.D.[3],
Roger A. Côté, M.D., D.Sc.(Hon)[1,4]
[1]College of American Pathologists, Northfield, IL;
[2]Oregon Health Sciences University, Portland, OR;
[3]Stanford University, Stanford, CA;
[4]CRDMI, Faculty of Medicine, Université de Sherbrooke, P.Q., Canada

We describe the framework for SNOMED RT (Reference Terminology), designed to complement the broad coverage of medical concepts in SNOMED with a set of enhanced features that significantly increases its value as a reference terminology for representing clinical data. We describe what is meant by a reference terminology, and differentiate SNOMED RT from specialized terminologies that enable user interfaces, electronic messaging, or natural language processing, as well as from other specialized reference terminologies whose primary purpose is for representing data that is not primarily clinical in nature. We then describe how SNOMED RT represents multiple hierarchies and incorporates description logic. We believe that such a comprehensive set of concepts at multiple levels of granularity, with multiple logic-based subsumption hierarchies can meet the requirements of a reference terminology for health care.

INTRODUCTION

In recent years there has been a growing awareness of the need for a solution to what might be generically called the "terminology problem" in health care.[1] Different facets of this problem are recognized in many settings: the outcomes researcher who finds a lack of comorbidity information in existing claims data; the information system vendor who needs a terminology for implementing a coded problem list in an electronic medical record product; the managed care organization that needs more detailed data on how they take care of patients, in order manage outcomes and implement guidelines and best practices.

The terminology problem has many aspects, and as awareness of the problem has grown, there has also developed an increasingly sophisticated understanding of the wide variety of different aspects of this problem. Given this more detailed understanding, it is no longer appropriate to simply refer to the problem as a lack of a "common data dictionary" for all applications and institutions. In fact, such an all-encompassing universal solution is unlikely and impractical. Instead, we propose that there are needs for several different kinds of terminology, divided not along boundaries of professional group or concept type, but divided along boundaries separating different uses of terminology; for example, one set may make possible a user-friendly structured data entry interface, another may optimize natural language processing, and another may enable storage, retrieval and analysis of clinical data. This latter terminology set, which we describe further below, may be called a "reference terminology" for clinical data. We propose that SNOMED's natural purpose is to fill the need for such a reference terminology.

SNOMED International, the Systematized Nomenclature of Human and Veterinary Medicine, has been in development for over 20 years.[2] It is a comprehensive set of over 150,000 records in twelve different axes or chapters. The concepts include anatomy (topography), morphology (pathologic structure), normal and abnormal functions, symptoms and signs of disease, chemicals, drugs, enzymes and other body proteins, living organisms, physical agents, spatial relationships, occupations, social contexts, diseases/diagnoses and procedures. Within the disease/diagnosis axis, many disease concepts have cross-references to other concepts in the terminology that are essential characteristics of the disease; these form a useful basis for further formalization. Thus SNOMED forms a tremendous starting point for development of a reference terminology.

REFERENCE TERMINOLOGY

A reference terminology for clinical data is a set of concepts and relationships that provides a common reference point for comparison and aggregation of data about the entire health care process, recorded by multiple different individuals, systems, or

1091—8280/97/$5.00

institutions. The main purpose of a reference terminology for clinical data is the retrieval and analysis of data relating to the causes of disease, the treatment of patients, and the outcomes of the overall health care process. Other reference terminologies may be optimized for other aspects of health care information (e.g. literature retrieval).

The need for a reference terminology can be illustrated by the situation facing many managed care organizations. They may have several different hospitals and clinics, each with an existing set of information systems. These organizations need to aggregate data from several systems in order to manage the quality and cost of care across the entire organization. Rather than totally replacing their existing information systems with one common system, they have a need to record data from each system using or referring to a common reference terminology. Aggregate data can then be grouped and analyzed using the various hierarchies of the reference terminology. A prominent example of this approach is the joint project between Kaiser Permanente and the Mayo Clinic to develop a Convergent Medical Terminology (CMT) based on SNOMED.[3,4]

Vendors of medical information systems and electronic medical records systems also have a need for a common reference terminology; it would enhance their ability to create useful tools that can help systematize best practices and guidelines. It would be highly inefficient and costly to support completely different terminology systems for each client, and would be preferable to begin with a common base and make local modifications only when necessary. With a common reference terminology, rules and reminders could be ported between sites and refined, rather than largely re-written from scratch. As health care changes with new drugs, new procedures, and new tests, it would be difficult if not impossible to maintain sophisticated information systems across multiple sites without a common reference terminology for clinical data.

Thus a reference terminology can provide a foundation upon which health care organizations, information system vendors, the insurance industry, government and others can aggregate and analyze data for the improvement of health care.

Development of usable information systems for health care involves many uses for terminology. A reference terminology does not necessarily meet the need for terminologies optimized for user interface or natural language applications. For example, a set of "pick lists" and abbreviation lists may be useful for menu-based or direct input of clinical data; these lists may comprise a user interface terminology. Once stored, however, the data would be recorded in, or point to, the reference terminology. For the purposes of data retrieval and analysis, it is not necessary that the data have been recorded using a single user interface terminology. A terminology for natural language applications may require additional features such as stemmed forms, lexical variants, synonyms, and multilingual translations. We believe it is useful to separate the development of the reference terminology from the development of terminologies for natural language and user interface applications. This separation may allow different terminological efforts to focus on separate parts of the problem and to cooperate in solving the overall problem.

An important part of the overall picture of health care terminology is the Unified Medical Language System (UMLS).[5] From the beginning of the UMLS project, SNOMED has been, and fully expects to continue to be, one of the largest and most important contributors of terms. The UMLS project has always endorsed the notion of going to the source terminology to determine intensional "meaning."[6] By enhancing SNOMED as a reference terminology, we are improving the availability of precise semantics to the developer, while continuing to support the role of UMLS as a thesaurus. The term relationships in the UMLS do not exist independent of a source terminology – thus it is very important for SNOMED to make explicit and give precise logical semantics to the inter-relationships of its terms.

Many UMLS source terminologies have "overloaded" semantics. For example, "gastrointestinal transit" in the Medical Subject Headings (MeSH) is used to denote both the physiologic function and the diagnostic measure. SNOMED would use separate codes for these two meanings; by supporting its terms with description logic, SNOMED is embracing a formal framework that discourages overloaded semantics.

The UMLS makes available many of the semantics of the source terminologies in a standardized format, but is not itself a reference terminology. When a developer uses the UMLS as a "synonym rich entry point into a reference terminology" with additional value added by providing links to the literature, the need for diversity is provided by the UMLS and the

need for precision and pathophysiologically sound semantics is provided by SNOMED.

DESCRIPTION LOGIC

Description logic, or terminological logic, refers to a group of similar approaches to the logic-based description of terminology.[7] Other medical terminology efforts are employing description logic, notably the GALEN project.[8] Growing out of efforts in knowledge representation and artificial intelligence, description logics have been developed to allow formal representation of the meanings of concepts and their inter-relationships. These formal descriptions can be used by inference algorithms to determine equivalence of different concept descriptions, as well as hierarchical relationships between descriptions. In other words, they provide for algorithmic determination of synonymy and hierarchies.

Dealing with Equivalency and Redundancy

Any terminology system that allows composition (combining existing concepts to represent new meanings) will of necessity encounter multiple ways to express the same concepts. For example, the concept "ruptured ovarian cyst" can be composed of the morphologic concepts "rupture" and "cyst", together with the topographic concept "ovary". There are also the concepts "ruptured ovary", "ovarian cyst", and "ruptured cyst" that can be composed in various combinations. Thus we can have "ruptured ovary" combined with "cyst", or "ruptured cyst" combined with "ovary", or "cyst of ovary" combined with "rupture". It becomes clear that multiple representations are inevitable in any compositional system. The need for a formal solution to this problem has been recognized by the Canon group.[9] Description logic makes it possible to automatically determine the equivalence of concepts, to identify intentional equivalency of composite concepts and their constituent parts, and to discover unintended redundancy in which the same concept is represented by two different terms.

A Description Logic Syntax for SNOMED

In selecting a syntax for representing SNOMED concepts, the Editorial Board considered several different possibilities. In order to maintain future compatibility with a variety of knowledge representation systems, while still making progress, we chose to adopt a syntax specifically designed to permit knowledge sharing. The Knowledge Representation System Specification (KRSS) was selected as the basis for SNOMED concept representations. This specification was designed for description logics by the DARPA knowledge sharing effort.[10,11] In order to improve readability, we change the prefix notation of KRSS to an infix notation, use "&" instead of "and", and make existential quantification of roles implicit. The following example illustrates how concepts are represented in standard KRSS format and in the SNOMED RT syntax for the concept "Postoperative esophagitis":

SNOMED III termcode and English nomenclature:
D5-30150 Postoperative esophagitis

SNOMED III components of the concept::
T-56000 Esophagus
M-40000 Inflammation
F-06030 Post-operative state

Cross-reference field in SNOMED III:
(T-56000)(M-40000)(F-06030)

Parent term in the SNOMED III hierarchy:
D5-30100 Esophagitis, NOS

Essential characteristics, in KRSS syntax:

```
(defconcept D5-30150
    (and D5-30100
        (some assoc-topography T-56000)
        (some assoc-morphology M-40000)
        (some assoc-etiology F-06030)))
```

Essential characteristics, in SNOMED RT syntax:

```
D5-30150:
    D5-30100 &
        (assoc-topography T-56000) &
        (assoc-morphology M-40000) &
        (assoc-etiology F-06030)
```

In order to ensure inter-operability, we have a simple parser that automatically inter-converts expressions in KRSS and SNOMED RT.

HIERARCHIES

Terminological hierarchies are useful for many purposes. They assist in classification of concepts, in correct placement of new concepts, in preventing duplication, and in providing lists of similar concepts (the children of a particular parent). For example, urinary 17-ketosteroids is a parent concept, and retrieving the children of the concept provides a list of particular known 17-ketosteroids. Some of these steroids would also be classified under

Termcode	Class	Enomen	Cross - Reference
DD-00000		CHAPTER D INJURIES AND POISONINGS	
DD-20000	-	SECTION D-2 INJURIES OF THE INTERNAL ORGANS	
DD-22000	0	D-22 INTERNAL INJURIES OF THE CHEST	
DD-22100	01	Internal injury of chest, NOS	
DD-22111	01	**Traumatic pneumothorax, NOS**	**D2-80300**
D2-00000	-	CHAPTER 2 DISEASES OF THE RESPIRATORY SYSTEM	
D2-80000	-	SECTION 2-8 DISEASES OF THE PLEURA, MEDIASTINUM AND DIAPHRAGM	
D2-80000	0	2-80 DISEASES OF THE PLEURA	
D2-80300	01	Pneumothorax, NOS	

Table 1: Multiple hierarchies for DD-22111 "Traumatic pneumothorax", (SNOMED III)

androgenic metabolites, while others may be classified under estrogen metabolites. It is readily apparent that multiple classifications are both useful and necessary.

SNOMED III Hierarchies

The main SNOMED III hierarchy is implicit, but can usually be inferred from the termcodes and "class" field. In addition, multiple hierarchies are supported. An example is given in Table 1. Notice that the cross-reference field allows a link to another hierarchy, thus supporting multiple hierarchical relationships. In this example, traumatic pneumothorax is linked to hierarchies for both injuries and respiratory diseases. However, this use of the cross-reference field is somewhat indirect, and a more explicit direct representation would be desirable. Another reason to make the hierarchies explicit is that concepts may not always have a hierarchical relationship even when the termcode suggests one. For example, termcode DD-22110 represents "traumatic pneumothorax without open wound into thorax" and DD-22111 represents "traumatic pneumothorax". DD-22111 would be expected to be a child of DD-22110, but in this case it is not (semantically). Thus the termcodes themselves cannot always be relied upon to correctly represent the hierarchical relationships among concepts in SNOMED III.

SNOMED RT Hierarchies

In order to provide a single coherent and consistent representation of multiple hierarchies, SNOMED RT contains explicit hierarchy tables. Table 2 shows the new explicit representation of the implied hierarchies from Table 1.

This change allows SNOMED RT to represent multiple hierarchies and to make the type of hierarchy explicit. It also frees SNOMED from the need to attempt to encode hierarchical information in the termcodes.

The RT hierarchies have the following features:

- relational table representation
- explicit, termcode independent representation
- multiple hierarchies per term
- multiple types of hierarchies (Is-A, Part-Of, etc.)
- many levels of granularity
- supported and determined by description logic

Parent Term	Child Term	Relationship
DD-00000	DD-20000	Is-A
DD-20000	DD-22100	Is-A
DD-22100	DD-22111	Is-A
D2-00000	D2-80000	Is-A
D2-80000	D2-80300	Is-A
D2-80300	DD-22111	Is-A

Table 2: SNOMED RT Hierarchy Representation

STATUS AND FUTURE DIRECTIONS

The two new features of SNOMED described in this paper (the hierarchy table and the description logic) have been in active development for over a year. The SNOMED Editorial Board is reviewing the existing (implicit) hierarchies in each axis. In conjunction with the CMT project, several Kaiser Permanente physicians, nurses, and other health care professionals are creating and refining logic-based descriptions for input to the K-Rep environment,[12] a description logic engine. This tool is used to elaborate the hierarchical relationships among concepts. The SNOMED Editor (RAC) and Scientific

Director (KAS) are reviewing the work in preparation for publication as part of SNOMED RT.

A complete logical model for a reference terminology must include more than a syntax, but also a complete and coherent set of roles (binary predicates). In the early stages of development, we have adopted a set of very general roles, such as "associated-topography" and "associated-morphology." As the CMT project progresses, we expect to discover a need to specialize some of these roles. However, we believe that this approach, to establish an initial framework for the terminology and subsequently refine it, will be most cost-effective in the long run.

The description logic syntax described in this paper is not intended to specify a model for forming compositional concepts when encoding data; rather, it is intended to specify the logical description of essential characteristics of concepts in the reference terminology itself. This leaves as yet unanswered the need for a more formal mechanism for combining SNOMED termcodes to express actual assertions about patients and their health care.

CONCLUSION

We have described the basic framework for incorporating explicit multiple hierarchies and description logic into SNOMED RT, with the intention of creating the necessary knowledge to support its use as a reference terminology. We believe that the distinction between reference terminology and other specialized forms of terminology will help to diffuse inaccurate perceptions of competition between various developers of terminological systems in health care, and may promote a path to more rapid development and deployment of applications that utilize sophisticated terminological systems.

Acknowledgments

The authors acknowledge the assistance of the SNOMED Editorial Board in developing the scientific direction for SNOMED RT.

References

1. Board of Directors of the AMIA. Standards for medical identifiers, codes, and messages needed to create an efficient computer-stored medical record. JAMIA 1994;1,1-7.

2. Cote RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L, eds. The Systematized Nomenclature of Human and Veterinary Medicine: SNOMED International. Northfield, IL: College of American Pathologists, 1993.

3. Campbell KE, Cohn SP, Chute CG, Rennels G, Shortliffe EH. Computer-based support for evolution of a convergent medical terminology. In: AMIA Annual Fall Symposium.: 1996:269-273.

4. Campbell KE, Cohn SP, Chute CG, Shortliffe EF, Rennels G. Scalable methodologies for distributed development of logic-based convergent medical terminology. IMIA WG6, January 1997.

5. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Methods of Information in Medicine 1993; 32:281-91.

6. McCray AT, Hole WT. The scope and structure of the first version of the UMLS semantic network. In: Miller RA, ed. Fourteenth Annual Symposium on Computer Applications in Medical Care. Washington DC: IEEE Computer Society Press, 1990..

7. Woods WA, Schmolze JG. The KL-ONE family. Computers and Mathematics with Applications 74:2-5, 1992.

8. Rector AL, Glowinski AJ, et al. Medical-concept models and medical records: An approach based on GALEN and PEN&PAD. JAMIA 2(1):19-35, 1995.

9. Evans DA, Cimino JJ, et al. Toward a medical-concept representation language. JAMIA 1(3):207-217, 1994.

10. KRSS working group of the DARPA Knowledge Sharing Effort. Draft of the specification for description logics. http://www-ksl.stanford.edu/ knowledge-sharing/ papers/ index.html#dl-spec 1993; July 1993.

11. Patil RS, et al. The DARPA knowledge sharing effort: progress report. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference, Cambridge, MA, Morgan Kaufmann, 1992.

12. Mays E, Weida R, Dionne R, et al. Scalable and expressive medical terminologies. AMIA Fall Symposium. Washington DC: Hanley & Belfus, Inc., 1996:259-263.