Managing Information with SNOMED: Understanding the Model

David J. Rothwell[a], M.D., Roger A. Côté[b], M.D.
[a] Medical College of Wisconsin, 14515 West Woodridge Circle, Brookfield, WI 53005
[b]C.R.D.M.I., Faculty of Medicine, Université de Sherbrooke, Sherbrooke, Qc, J1H 5N4

SNOMED is an inventory of medical terms and concepts. Each concept has a unique representation, it's primary termcode. The primary termcode carries hierarchical and inheritable information about each concept it represents. Additional hierarchical and inheritable information is carried by cross references linked to each concept. The cross references provide the defining characteristics for each concept thereby creating a computer readable definition for each SNOMED entry. These are some of the attributes that define the SNOMED model.

INTRODUCTION

The Systematized Nomenclature of Human and Veterinary Medicine - SNOMED International [1] is an inventory of medical terms and concepts that encompasses virtually the entire domain of human and veterinary medicine. It is a robust controlled medical vocabulary. Both elemental (granular) and complex concepts are present; the complex concepts are defined by their elementary concepts where each element is placed in a cross reference file that provides the defining characteristics for that term [2]. Rather than placing each term or concept along with its synonyms, related terms and definitions in an alphabetic order as in the standard dictionary format, a grouping or classification of concepts that share important characteristics is created. Thus all anatomic terms are placed together in one class, all biochemical and physiologic terms in another, all living organisms and all neoplasms are each placed in their respective classes. Within each class there is an additional ordering of terms that are more closely related to one another. For example anatomic terms naturally fall into groups or classes defined by body organ systems (e.g. respiratory, cardiovascular, gastrointestinal, etc.), the biochemical terms are grouped by their structure and function (e.g. chemical elements, compounds, amino acids, hormones, etc.).

The number of groups or classes assigned or developed for a complete set of medical terms is in some ways arbitrary. It is, however, largely determined by customary usage of terms, an understanding of the disciplines of medicine, common sense and what works (heuristics) [3]. The subgroups for each class are developed following similar guidelines. In doing this the terms are placed into taxonomic hierarchies, expressing their natural relationships to one another. Adding synonyms and alternate expressions for each term completes this portion of the task.

PRIMARY TERMCODES

In its present form there are eleven major classes of terms in SNOMED, each major class forming a separate module (Figure 1).

SNOMED INTERNATIONAL VERSION 3.3

MODULES	RECORDS
T = Topography	12,803
M = Morphology	5,672
F = Function	18,027
L = Living Organisms	24,480
C = Chemicals, Drugs & Biological Products	14,275
A = Physical Agents, Forces and Activities	1,410
J = Occupations	1,947
S = Social Context	845
D = Disease/Diagnosis	34,377
P = Procedures	28,685
G = General Linkage/Modifiers	1,373
TOTAL RECORDS	**144,257**

Figure 1 - SNOMED modules and record count as of July 1996

The hierarchical information contained in the class and subclass assignments referred to above, is found within each term's primary coded representation, the termcode. That is, each termcode contains implicit information embedded in it's alphanumeric representation. Illustrations of this are shown in Figure 2.

TOPOGRAPHY		MORPHOLOGY	
a. T-30000	Cardiovascular system	c. M-50000	Degeneration
T-32000	Heart	M-54700	Infarct
T-32020	Myocardium	M-54720	Acute infarct

TOPOGRAPHY		FUNCTION	
b. T-50000	Digestive system	d. F-20000	Respiratory function
T-58000	Small intestine	F-23000	Respiratory sounds
T-58200	Duodenum	F-23200	Rales

Figure 2 - Examples of alphanumeric hierarchies

As can be seen from the illustration the implicit hierarchical information contained within the termcode is as follows; the myocardium "is part of" the heart and the heart "is part of" the cardiovascular system; alternatively it may be implied that the cardiovascular system "has part" myocardium. Similar statements can be derived from each of the other examples shown in Figure 2. The duodenum "is part of" the small intestine which "is part of" the digestive system; an acute infarct "is a" type of infarct which in turn "is a" type of degeneration; a rale "is a" respiratory sound which "is a" type of respiratory function. These are hierarchical relationships implied within the termcode.

These implied relationships, however, need to be made explicit in a formal way so that the information contained within the termcode is retrievable. It's precise position in the hierarchy and the nature of it's relationship (link) to other terms must be captured explicitly and made available as data elements for use in a database [4,5].

One mechanism to accomplish this task is to create type hierarchies for each term making explicit the implicit information in each termcode. Each member of the hierarchy can be specifically and explicitly linked to each other. An illustration of type hierarchies is shown in Figure 3.

Parent/Child	Child/Parent	Link
Myocardium	Heart	is part of
Heart	C.V.	is part of
Digestive System	Small intestine	has part
Small intestine	Duodenum	has part
Acute infarct	Infarct	is a
Infarct	Degeneration	is a
Respiratory sound	Respiratory function	is member of
Rales	Respiratory sound	is member of

Figure 3 - Examples of type hierarchies

As shown in the table, myocardium "is part of" the heart and heart "is part of" the cardiovascular system, the digestive system "has part" small intestine which in turn "has part" duodenum. In each of these cases inheritance rules can be invoked to establish that the myocardium is part of the cardiovascular system, that the duodenum is part of the digestive system and that rales are not only a member of the family of respiratory sounds, but also a member of the family of respiratory functions [6].

CROSS REFERENCES

In the printed edition of SNOMED there are approximately 34,000 cross references. Each cross reference is a SNOMED termcode taken from an existing SNOMED module. For complex concepts the cross references provide the elemental terms that make up that concept, for other terms they provide the terms' defining characteristics, i.e. their essential pathophysiology. When completed, cross references will provide a mechanism for a computerized definition for each SNOMED term.

In the July 1996 release of SNOMED, version 3.3, approximately 25% of the relevant clinical terms include cross-references. A major initiative to complete this phase of the work is underway.

Illustration of cross references and their properties are shown in Figure 4.

DE- Infectious Diseases

DE-13620	**Topography**
Streptobacillary fever	T-01000 Skin
Haverhill fever	**Morphology**
Epidemic arthritic erythema	M-14360 Bite
	Living Organisms
	L-81400 Rat
	L-25001 Streptobacillus moniliformis

D0- Diseases of Skin and Subcutaneous tissues

D0-10100	**Topography**
Eczema, NOS	T-01000 Skin
Eczematous dermatitis	**Morphology**
	M-40000 Inflammation
	M-01735 Papulovesicular rash
	Function
	F-C3000 Hypersensitivity reaction
	F-A2300 Pruritus

Figure 4 - Examples of cross-references in Disease module

In the first example the preferred term Streptobacillary fever is followed by it's two synonyms. The primary termcode itself carries with it a significant amount of information i.e. Streptobacillary fever is one of a set of rat bite fevers (DE-136..) which are in the class of Infectious diseases (DE-.....).

The cross references provide further defining characteristics of the term namely that the organism involved is Streptobacillus moniliformis transmitted by a vector, the rat, by means of a bite. The cross references in the second example provide even more in-depth information about the term Eczema. Eczema involves the skin, there is inflammation with a papulovesicular rash with itching due to a hypersensitivity reaction. When terms defining characteristics are isolated in this way the information, i.e. each data element, is easily retrievable from a database for use in clinical alerts, decision analysis, clinical guidelines and assessment of outcomes [7]. Only when the full extent of knowledge implied in a particular term is made explicit and retrievable can it be used to fulfill these purposes.

DISCUSSION

This is the essence of the SNOMED data model. Each medical term is placed into its own family of related terms and assigned a primary termcode which embeds generic and partitive hierarchical information about that term. Creation of type hierarchies for each term makes all of this embedded information explicit and available. Hierarchies, if properly constructed, are probably the most efficient way of providing for inheritance - an essential property of any representation language. Type hierarchies illustrated above can not only define characteristics of a term, but also can provide constraints for their use, a necessary condition in any database application.

Cross references provide a powerful extension of the basic SNOMED model. The defining characteristics of each term are explicitly provided giving what is in effect definitions for each term in a computer readable and computer processable form i.e. a computerized dictionary. From this it can be seen that SNOMED codes are not simply a substitute of codes for words but are in fact a compact, computable method for encapsulating a significant amount of relevant information about a concept into a data structure - the SNOMED termcode and its cross-references.

Other representational techniques could equally well be applied to capture the definitional information that surrounds each medical term. Conceptual graphs are a representational language based on a first order logic with a graph notation [8]. They integrate several features present in semantic and frame representations. Campbell et al. [9] has recently shown the feasibility of this approach when applied to SNOMED. An alternative technique would be to consider each SNOMED concept as an Object with its hierarchical and definitional components embedded in them. Each object carries within it the rules by which it can be accessed and used providing the necessary constraints. Whatever the representational technique used, the embedded information in the SNOMED primary termcode and in it's cross references contains the necessary medical knowledge for construction of these representations [10].

The requirements for a medical vocabulary outlined in recent papers by the Canon Group [11,12] are that it be a coherent conceptual representation across applications and subject domains. It was suggested as a primary tenet that terms should correspond to concepts that have their meaning made explicit through a deep representation structure suitable for a variety of uses. Important aspects also suggested were that it be a controlled vocabulary expressed within a semantic network or hierarchy, with an explicit concept model and notation, and incorporate granular as well as complex terms. Upon completion of the type hierarchies and cross references proposed for SNOMED and in progress, essentially all of these requirements will have been met. In representation languages as well as in natural languages, multiple ways to express the same concept remain an issue. A remaining challenge for informaticists is to create methods to recognize and to prevent or minimize these redundant expressions from occurring. Logical definitions of each term or concept may be a solution to this problem.

References

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

2. Rothwell DJ, Cote RA, Cordeau JP, Boisvert MA. Developing a Standard Data Structure for Medical Language - The SNOMED Proposal. In Proc SCAMC 94. Ozbolt JG, Ed. Hanley & Belfus, Inc 1994:695-699.

3. Wingert F. Medical Linguistics: Automated indexing into SNOMED. Crit Rev Med Inform. 1988, 1(4):333-403.

4. Satomura Y, do Amaral MB. Automated diagnostic indexing by natural language processing. Med Inform (London). 1992 Jul-Sep, 17(3):149-163.

5. Henry SB, Campbell KE, Holzemer WL. Representation of nursing terms for the description of patient problems using SNOMED III. In Proc SCAMC 93. Safran C, Ed. McGraw-Hill, Inc 1993:700-704.

6. Campbell KE, Musen MA. Representation of clinical data using SNOMED III and Conceptual Graphs. In Proc SCAMC 92. Frisse ME, Ed. McGraw-Hill, Inc 1992: 354-358.

7. Huff SM, Rocha RA, Bray BE et al. An Event Model of Medical Information Representation. J Am Med Informatics Assoc. 1995,2: 116-134.

8. Sowa JF. Conceptual Structures. Reading, MA: Addison-Wesley. 1984.

9. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. J Am Med Informatics Assoc. 1994, 1:218-232.

10. Sager N, Lyman M, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. J Am Med Informatics Assoc. 1994, 1:42-60.

11. Evans DA, Cinimo JJ, Hersh WR, et al. Toward a medical-concept representation language. J Am Med Informatics Assoc. 1994, 1:207-217.

12. Friedman C, Huff SM, Hersh WR, et al. The Canon Group's Effort: Working Toward a Merged Model. J Am Med Informatics Assoc. 1995, 2:4-18.