Implementing rules to improve the quality of concept post-coordination with SNOMED CT

Navas H.[a], Lopez Osornio A.[b], Gambarte L.[a], Elías Leguizamón G.[a], Wasserman S.[a], Orrego N.[a], Luna D.[a], Gonzalez B. de Quirós F.[a]

[a] *Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina*
[b] *TermMed, Buenos Aires, Argentina*

Abstract

The use of SNOMED CT as a standard reference terminology enables interoperability between clinical systems. This reference tool provides a method for creating post-coordinated terms by users according to local needs. While the creation of these terms is free, there are a number of rules, as defined in the user manual of SNOMED CT that must be followed. The Hospital Italiano of Buenos Aires has a Terminology Server that encodes medical terms, using SNOMED CT as the reference vocabulary. An interoperability analysis performed with the Nebraska Medical Center in 2006 found a high error rate (26%) in post-coordinated terms. Therefore, we implemented an automatic system of rules within the Terminology Server as defined in the user manual.
Following rules implementation, the error rate decreased from 26% to 2%.

Keywords:

Medical records systems, Vocabulary controlled, Systematized nomenclature of medicine, Terminology, Information systems, SNOMED CT.

Introduction

SNOMED CT has been proposed as a standard terminology in many countries, with an appropriate coverage of medical vocabulary [1]. Development standards are currently maintained by a multinational organization but local implementations often require content extension to address specific requirements. For this reason SNOMED International identifies and promotes mechanisms for post-coordination.

The Electronic Medical Record of the Hospital Italiano in Buenos Aires initially allowed physicians free text entry of problems and procedures. After a period of four years, we analyzed data and migrated this information into a specialized server. Then, a local interface terminology for physicians was developed, where they could choose a more accurate description for a given problem or a procedure [2]. This online terminology interface was created from items originally entered by individual physicians as free text.

The online interface consisted in 21,000 unique concepts, which were related to SNOMED CT as equivalent concepts or as post-coordinated expressions (in those cases where a direct equivalence could not be found). The creation of a post-coordinated expression was necessary for 16,000 of the 21,000 concepts (76.19%). This is a very high proportion of concepts requiring post-coordination, which is related to an implementation decision, by which all concepts recorded for at least 10 times in the problem list repository were included. These included highly specific concepts having a large number of modifiers, such as severity, laterality, clinical course, etc.

There are no specific tools for modeling SNOMED CT logics-dependant vocabularies. Programs such as Apelon or Protégé have been used in the past for modeling ontological standards. However, the specific rules of concept modeling and creation of logical definitions according to SNOMED International and published in SNOMED CT guides had not been tested or reinforced by these programs. The "in house" modeling tool designed and developed in the Hospital Italiano initially did not include rules for controlling modeling logics.

In 2006, an interoperability test was jointly developed and conducted between the Hospital Italiano of Buenos Aires and the Nebraska Medical Center in the United States. It evaluated SNOMED CT features by merging and comparing SNOMED-encoded problem lists from primary care sites in Nebraska and Argentina. Both problem lists showed small differences in semantic content, but differed substantially in the percentage of post-coordinated content. A classification using SNOMED Normal Forms effectively identified semantic equivalence in 65.2% of the reviewed cases. The most common reason for post-coordination failure was non-observance of SNOMED's guidelines (28.8% of remaining 34.8%) [3].

There were also approach differences among users who model post-coordinated expressions [4-6].

After these tests were completed, a new quality assurance tool was added in order to improve the quality of local post-coordinated expressions. This gave users the ability to review, evaluate and correct concept errors online. The system shows a red alert when a rule is broken.

Objective

Our purpose was to test the effectiveness of control modeling rules for concept post-coordination.

Background

The definition of post-coordinated expressions follows the same structure as any concept in SNOMED CT. Concepts included in the current distribution of SNOMED CT and defined by its authority are called pre-coordinated. New concepts defined by local users are called post-coordinated. The User Guide specifies the SNOMED CT post-coordination rules for new concepts. This is the Semantic Model of SNOMED CT [7].

The purpose of post-coordination is to incorporate new concepts into SNOMED CT, building these from existing ones and following their semantic model. The creation of post-coordinated expressions to represent new concepts must follow 2 steps:

- The first one is to define one or more supertypes for the new concept. The supertypes are in this case, more general concepts of SNOMED CT related to the current distribution. The proper assignation of supertypes defines the hierarchy of the concept[1].

- The second step is to incorporate attributes and differentiate the new and more specific concept from its SNOMED CT's supertype included in SNOMED CT. Each attribute is defined as a relationship with other SNOMED concept.

A concept that is fully explained and completely defined by its attributes and by those inherited from its supertypes is called "Fully defined". On the other hand, concepts, which meaning cannot be fully explained or with some different attributes, are "Primitive".

Hospital Italiano's SNOMED CT modeling tool, provides the users utilities for modeling new SNOMED CT concepts created in the local extension. The modeling process includes adding new relationships for describing supertypes and attributes. The creation of a new defining relationship in SNOMED CT consists in the selection of 3 values:

- Source concept: the concept that is being modeled with the relationship

- Relationship type: it can be a "is a" relationship for supertype definition (parent concepts) or any other attribute, like "finding site", "severity", "laterality", etc.

- Target concept: points to the parent concept in "is a" relationships, or to the attribute value, for example in the case of the relationship type "finding site" a usable value can be "lung structure".

According to the semantic model, different types of relationships are valid in different hierarchies, but only a pre-defined proportion of SNOMED CT concepts is a valid target for these relationships. For example, "Severity" attribute is a valid concept of the hierarchy "Clinical Finding", but not for concepts located in the "Procedure" hierarchy. The opposite occurs with the attribute "Direct Device", which is valid for "Procedure" but not for "Clinical Finding".

---

[1]   Relationships: page 10, SNOMED CT User Guide, July 2008

Materials and Methods

Semantic model rules are encoded in a relational database model using tables of valid relationship types. This hierarchical table identifies high-level concepts and a target chart, defined once again by the top level concept of the group of valid target concepts.

The Hospital Italiano vocabulary is also stored in a relational database, including its concepts, descriptions and relationships [8]. A procedure developed using PL / SQL checks if each post-coordinated expression is in accordance with SNOMED's semantic model restrictions.

The following rules were entered to the Terminology server using SNOMED's user guide:

- The concept must belong to only one hierarchy.

- The hierarchy of a given concept cannot be different from the domain hierarchy

- If one relationship depends on the existence of another one, this one must be present.

- The relationship type must be appropriate for that hierarchy.

- Relationship target concepts must be valid.

"In house" Created:

- Only one "IS A (mapping)" relationship can exist

- If the relationship used is "IS A (mapping)" the concept cannot be "Primitive".

- There must be at least one "is a" or "IS A (mapping)".relationship

- There is also a warning if coders use an "entire body structure".

In order to verify the effectiveness of the modeling control system, two samples of post-coordinated terms were selected from the Hospital Italiano vocabulary.

The first one took place in 2006, after Nebraska Medical Center testing and before the implementation of the rules system. The sample included 34,253 post-coordinated terms.

The second sample was assessed in 2008, after the implementation of the rules system (August 2007) and included 9,015 terms.

Post-coordinated concepts included concepts in all domains from the clinical information systems, as well as those included in the list of problems, procedures, drugs or devices.

Each list of concepts was reviewed by an expert that rated the concepts into "right" or "wrong" categories according to the standards of SNOMED post-coordination detailed in the User Guide. A more detailed error description was given for concepts rated as "wrong".

Errors were classified as follows:

- Error in hierarchy definition: The concept is not assigned to the correct hierarchy, e. g. a procedure in the hierarchy of problems (appendectomy as a problem), or a disposable in procedures (as a stent entered as a procedure, etc.).

- Error in the relationship type: a not allowed relationship type used for a given hierarchy, e. g. finding site used in procedure hierarchy.

- Error in relationship target: the target concept is out of the valid range for the relationship type. The relationship is correct, but the target is wrong, e. g. relationship "laterality" with "right knee" as a value, when the proper value should be "right" (otherwise, a "finding site" relationship should have been used).

- Inappropriate use of an anatomical structure as an entire structure: The target concept is a subtype of the concept Entire anatomical structure (body structure), as when the concept "entire leg" is used instead of "leg structure".

- Other errors: those not included in the previously mentioned categories.

The expert also assessed the accuracy of representation; post-coordinated expressions may fulfill the SNOMED CT standards, but not accurately describe the meaning of the concept. Using a 5-point Likert scale, the expert rated meaning representation from the best possible representation to a completely wrong representation.. This measurement will be useful to identify quality control components for the terminology interface, i.e. which parts should be automated and which ones would always require manual review.

Statistical analysis of both samples was performed using a null hypothesis test and power was analyzed. In order to identify a statistically significant difference (0.050), 50 randomly selected terms from each sample had to be analyzed to attain a power of 80.9% in order to obtain a statistically significant result. This estimate assumes that the difference in proportions is 0.20 (specifically, 0.25 versus 0.05).

The sample size also allowed us to report the difference between both samples with an accuracy of approximately 0.13 points (confidence level 95%).

Specifically, a difference of 0.20 would have a confidence interval of 0.07 to 0.33.

Table 1- Examples of post-coordination errors

| Physician text	Post-coordination coder error	System Alert	Proper post-coordination
Lobectomy of right lung	is a: Lobectomy of lung Side: Right	The relationship type must be appropriate for that hierarchy	is a: Lobectomy of lung Procedure site - Direct: Right lung structure
Pain in right buttock	is a: Pain in buttock Side: Right	If one relationship depends on the existence of another one, this one must be present	is a: Pain Finding site: Buttock Side: Right
Implantation of cardiac pacemaker	IS A (mapping): Implantation of cardiac pacemaker (Primitive)	If the relationship used is "IS A (mapping)" the concept cannot be "Primitive"	IS A (mapping): Implantation of cardiac pacemaker (Fully defined)
Fracture of tibia and fibula	IS A (mapping): Fracture of tibia IS A (mapping): Fracture of fibula	Only one "IS A (mapping)" relationship can exist	is a: Fracture of tibia is a: Fracture of fibula
Thoracic aorta atheromatosis	is a: Atherosclerosis of aorta Finding site: Entire thoracic aorta (body structure)	Warning of use an "entire body structure"	is a: Atherosclerosis of aorta Finding site: Thoracic aorta structure
Stent (in Precedure Domain)	is a: Stent (in Procedure Domain)	The hierarchy of a given concept cannot be different from the domain hierarchy	is a: Stent (in Device Domain)
Hematoma of left leg	Finding site: Leg Side: Left	There must be at least one "is a" or "IS A (mapping)".relationship	is a: Hematoma of leg Finding site: Leg Side: Left
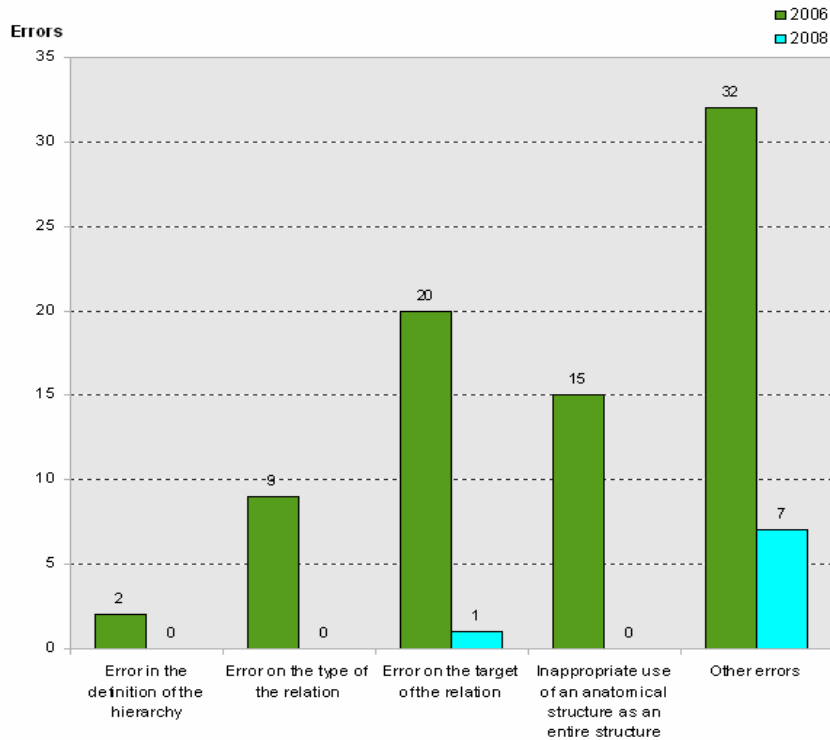
Figure 1- Number of errors grouped by category

Results

In the first sample, we analyzed 300 out of a total of 34,253 concepts; 74% showed no errors. The remaining 26% was distributed as follows: 0.67%, error in hierarchy definition; 3%, error in relationship type; 6.67%, error in relationship target; 5%, improper use of "entire part", and 10.67%, other errors.

In the second sample we analyzed 300 out of a total of 9,015 concepts. No errors were found in 97.3%. The remaining 2.7% consisted in: 0.3% error in relationship type and 2.4%, other errors (Figure 1).

The first sample in the accuracy performance analysis showed that 77% of the terms were adequately represented (points 1 and 2 on the Likert scale), 10.67% were poorly represented (points 4 and 5) and 12.3% could not be determined.

The second sample completed in 2008 showed that 96.33% of terms were properly represented, 2% were poorly represented and 1.67% could not be determined (Figure 2).

Figure 2- Quality performance

Conclusion

The implementation of an automatic system of rules for concept post-coordination, improves their representation, by enabling the proper use of SNOMED CT relationships, as well as the adequate representation of medical concepts.

Clearly, a system of rules will improve interoperability with other health centers, allowing better results than those achieved in 2006 with the Nebraska Medical Center.

It also has a positive impact in educational settings, by improving the training of users (coders) in charge of concept modeling, who should follow consistent rules to enhance concept representation.

This system of rules may reduce inter-user (coders) variability at the time of interpreting meaning and generating post-coordinated expressions, increasing consensus.

References

[1] Wasserman H, Wang J. An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. AMIA Annual Symposium proceedings / AMIA Symposium. 2003:699-703.

[2] Osornio AL, Luna D, Gambarte ML, Gomez A, Reynoso G, de Quiros FG. Creation of a local interface terminology to SNOMED CT. Studies in health technology and informatics. 2007;129(Pt 1):765-9.

[3] Campbell James R; Lopez Osornio, Alejandro; de Quiros, Fernan; Luna, Daniel and Reynoso, Guillermo. Semantic Interoperability and SNOMED CT: A Case Study in Clinical Problem Lists. In: Kuhn KAW, James R; Leong, Tze-Yun, ed. Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems 2007: [2401]-[2].

[4] Andrews JE, Patrick TB, Richesson RL, Brown H, Krischer JP. Comparing heterogeneous SNOMED CT coding of clinical research concepts by examining normalized expressions. Journal of biomedical informatics. 2008 Feb 5.

[5] Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. J Am Med Inform Assoc. 2007 Jul-Aug;14(4):497-506.

[6] Chiang MF, Hwang JC, Yu AC, Casper DS, Cimino JJ, Starren JB. Reliability of SNOMED-CT coding by three physicians using two terminology browsers. AMIA Annual Symposium proceedings / AMIA Symposium. 2006:131-5.

[7] Organization IHTSDO. SNOMED Clinical Terms User Guide. In: Organisation IHTSDO, ed. 2008.

[8] Gambarte ML, Osornio AL, Martinez M, Reynoso G, Luna D, de Quiros FG. A practical approach to advanced terminology services in health information systems. Studies in health technology and informatics. 2007;129(Pt 1):621-5.

Address for correspondence

Hernán Navas
Department of Health Informatics
Hospital Italiano de Buenos Aires
Gascón 450 (C1181ACH) Buenos Aires – Argentina
e-mail: hernan.navas@hospitalitaliano.org.ar