MEDINFO 2013
C.U. Lehmann et al. (Eds.)

doi:10.3233/978-1-61499-289-9-72

User-Directed Coordination in SNOMED CT

Ronald Cornet[a,b], Mikael Nyström[b], Daniel Karlsson[b]

[a] *Department of Medical Informatics, Academic Medical Center, University of Amsterdam, The Netherlands*
[b] *Department of Biomedical Engineering, Linköping University, Sweden*

Abstract

The possibility of post-coordination of SNOMED CT concepts, especially by clinical users, is both an asset and a challenge for SNOMED CT implementation. To get insight in the applicability of post-coordination, we analyzed scenarios for user-directed coordination that are described in the documentation of SNOMED CT. The analyses were based on experiences from previous and ongoing research and implementation work, including national mapping projects, and investigations on collection of data for multiple uses. These scenarios show various usability and representation problems: high number of relationships for refinement and qualification, improper options for refinement, incorrect formal definitions, and lack of support for applying editorial rules. Improved user-directed coordination in SNOMED CT in real practice requires advanced sanctioning, increased consistency of definitions of concepts in SNOMED CT, and real-time analysis of the post-coordinate expression.

Keywords:

SNOMED CT, Nomenclature as Topic, Clinical Coding.

Introduction

SNOMED CT [1] is considered a comprehensive clinical healthcare terminological system. It addresses most of Cimino's desiderata [2] among which "Content", "Concept Orientation", and "Formal definitions". The "Content" aspect is addressed firstly by formally defining approximately 300,000 active concepts, and secondly by allowing compositional extensibility. Compositional extensibility is also referred to as post-coordination. SNOMED CT provides various ways for performing post-coordination, which are explained in the background section.

In [3], a distinction is made between pre-coordination, post-coordination, and user-directed coordination of concepts.

Pre-coordinated concepts are those that have a single concept identifier, i.e., all concepts in the International Release of SNOMED CT are pre-coordinated concepts.

Post-coordination of concepts is the construction of a concept expression by combining concept identifiers. For example, an emergency CABG procedure can be represented as: CABG: priority = emergency, or, using concept identifiers:

232717009 : 260870009 = 25876001.

It needs to be noted that when such a post-coordinate expression is incorporated into an extension of SNOMED CT (e.g., in a local or national release), it gets a single concept identifier, and has then become a pre-coordinated concept.

User-directed coordination is described in [3] as: "*The User chooses this set of concepts, usually via a Graphical User Interface, and usually we envision that this would occur at the point-of-care. This is the attempt of a User to represent a clinical concept using a set of concepts …*"

Whereas there is a clear need for being able to capture clinically relevant concepts [4], the encoding is challenging. There is a high inter-coder variability [5], and the usability of post-coordination depends on the extent to which it is implemented [6] [7].

As described in section 4.2.2.1.4.4 of [8], user directed coordination raises user interface design issues: "*Application designers will struggle to create sensible and consistent interfaces unless advice on sanctioning is provided. Different issues will apply according to the nature of the interface. For example this may include:*

- *What options to offer users to allow refinement of specific concepts;*
- *How to represent the meaning that results from selecting options on a structured data entry form as a SNOMED CT expression;*
- *How to encode meaning derived from natural language processing*."

In this paper, we aim at studying these questions and determine requirements for answers.

Background

As described in section 8.2.1.2 onwards of [8], post-coordination can be realized by three approaches: refinement, qualification, and combination.

Post-coordination by refinement

In Release Format 2 (RF2) of SNOMED CT, the new format for distribution of the International Release, the relationship table contains defining relationships. These relationships may be further refined, implying that a post-coordinated concept can be created by refinement, i.e., selecting a subtype of a concept used in the definition.

For example, a pneumococcal pneumonia of the left upper lobe of the lung can be defined by further specifying the *finding site* of *pneumococcal lobar pneumonia*, i.e., *structure of lobe of lung*:

pneumococcal lobar pneumonia : *finding site* = *structure of left upper lobe of lung*

Post-coordination by qualification

SNOMED CT is based on a concept model, described in chapter 6 of [8]. This concept model specifies domain and range restrictions for relationships, hence providing a generic way of expressing allowed post-coordination. For example, both the domain and range of the *associated with* relationship are *clinical finding*. This enables specification of an associated

disorder for any finding, such as when the above-mentioned pneumonia is associated with AIDS. This can be specified as:

pneumococcal lobar pneumonia : associated with = AIDS

In addition, SNOMED CT provides the possibility of specifying qualifiers, which represent relationships that do not constitute the formal definition of a concept, but that provide further specification thereof. Qualification is primarily a mechanism to explicitly represent relevant parts of the concept model. For example, clinical course can be represented as a qualifier for the above mentioned concept:

pneumococcal lobar pneumonia : clinical course << courses

This enables specification of chronic pneumococcal lobar pneumonia as:

pneumococcal lobar pneumonia : clinical course = chronic

Post-coordination by combination

The third way of post-coordination is logical conjunction of two or more concepts. In this case, a concept that specifies certain properties is added, instead of specifying the property itself. Applying combination, the previous example of pneumococcal pneumonia of the left upper lobe of the lung can be represented as follows:

pneumococcal lobar pneumonia + left upper zone pneumonia

Materials and Methods

SNOMED CT

The July 2012 version of SNOMED CT was used, of which the snapshot (i.e., most recent status for all concepts, descriptions and relationships) was imported into a MySQL database.

In addition, a transitive closure table was generated, relating concepts to all of their supertypes, including the length of the path from the concept to each supertype. Only concepts, relationships and descriptions that were active as of July 2012 were used.

Additionally, CliniClue[1], and the Dutch Release Center's web-based SNOMED CT browser[2] were used.

Analyses

To address the questions raised in [8] and mentioned in the introduction, we studied real cases that have emerged from previous and ongoing research and implementation work, including national mapping projects, and investigations on collection of data for multiple uses.

Results

Options to allow refinement of specific concepts

Earlier research showed that allowing refinement of concepts can work [6], but only if the suggested refinements are sensible. Allowing refinement based on all characteristics and qualifiers that may be refined was judged inappropriate [7], as this presents a user with too many relationships and too many possible values for these relationships.

Table 107 and Table 124 of [8] provide insight in the amount of possible relationships when applying the concept model for qualification. For clinical findings and procedures, the concept model allows for 16 and 23 relationship types respectively. Providing all these for qualification in user-directed coordination will result in too many options provided to users.

[1] Available from http://www.cliniclue.com/
[2] http://terminologie.nictiz.nl/

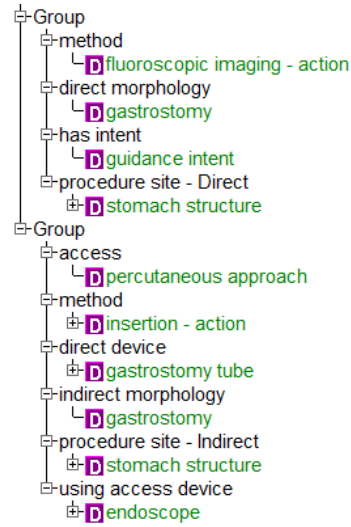

Figure 1- Defining characteristics of percutaneous endoscopic insertion of gastrostomy tube using fluoroscopic guidance

Instead of applying qualification, only allowing refinement will provide a number of choices that is smaller but still significant. Figure 1 shows an example of a procedure with 9 different types of relationships, of which 5 values (indicated by the '+' sign) could be refined. This example shows that allowing refinement of all specified relationships provides many options.

Apart from the number of choices, the relevance of choices is a problem. This is shown by inspection of the subtypes of insertion (Figure 2), which reveals that many of the subtypes are no sensible methods for an insertion of a gastrostomy tube.

Enabling qualification based on the concept model will further increase the number of choices with implausible, irrelevant or nonsensical values.

A second problem of post-coordination is the risk of not specifying required characteristics.

In the process of performing a mapping of a Dutch interface terminology to SNOMED CT, *Acute infective gastroenteritis due to Norwalk virus* was mapped. As this concept isn't pre-coordinated in SNOMED CT, post-coordination is needed. Intuitively this is done as shown in Figure 3a), i.e., by specifying the *sudden onset AND/OR short duration* as *clinical course* of *viral gastroenteritis due to Norwalk-like agents* or as in 3b), i.e., by specifying the *Norovirus* as *causative agent* of *acute infective gastroenteritis*.

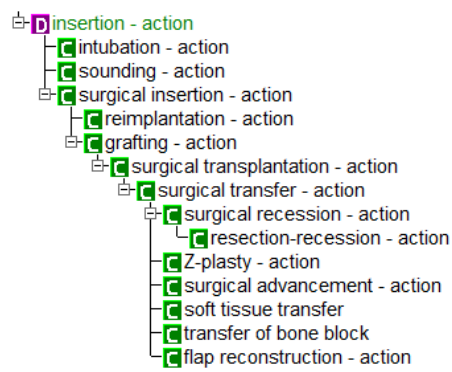

Figure 2- Subtypes of insertion – action

```
a) | viral gastroenteritis due to Norwalk-like agents | :
        | clinical course | = | sudden onset AND/OR short du-
ration | ==
| infective food poisoning | :
| causative agent | = | Norovirus | ,
| clinical course | = | acute onset | ,
| pathological process | = | infectious process | ,
        { | associated morphology | = | inflammation | ,
                | finding site | = | intestinal structure |
        },
        { | associated morphology | = | inflammation | ,
                | finding site | = | stomach structure |
        }

b) | acute infective gastroenteritis | : | causative agent | = | No-
rovirus | ==
| disease | :
| causative agent | = | Norovirus |
| clinical course | = | sudden onset AND/OR short duration | ,
| pathological process | = | infectious process | ,
        { | associated morphology | = | acute inflammation | ,
                | finding site | = | intestinal structure |
        },
        { | associated morphology | = | acute inflammation | ,
                | finding site | = | stomach structure |
        }

c) | viral gastroenteritis due to Norwalk-like agents | + | acute
infective gastroenteritis | ==
| infective food poisoning | :
| causative agent | = | Norovirus |
, | clinical course | = | sudden onset AND/OR short duration |
, | pathological process | = | infectious process |
, { | associated morphology | = | acute inflammation |
, | finding site | = | intestinal structure |
}
, { | associated morphology | = | acute inflammation |
, | finding site | = | stomach structure |
}
```

Figure 3- Three alternative definitions for Acute infective gastroenteritis due to Norwalk virus, together with normal forms

When following the first approach we miss the adequate refinement of the *associated morphology*, which should be an *acute inflammation*, as is defined in *acute infective gastroenteritis*. Conversely, with the second approach we miss out on the fact that SNOMED CT considers this disorder as a kind of *infective food poisoning*, as is defined in *viral gastroenteritis due to Norwalk-like agents*. Consequently, to ensure all inferences are taken into account, combination of concepts would be the preferred approach, as shown in Figure 3c.

Representing selected options on a structured data entry form

Data collection in an ophthalmology department provided a prototypical example of the intricacy of representing information entered in the user interface using SNOMED CT. We illustrate this in Figure 4 using the diagnosis section of the American Optometric Association (AOA) Diabetes Eye Examination Report form[3].

This form poses various challenges, of which we will focus on post-coordination, not discussing issues pertaining to recording of presence or absence, or mutually exclusive options.

For each diagnosis, the form provides two check boxes, OD (oculus dexter) for the right eye, and OS (oculus sinister) for the left eye.

[3] http://www.aoa.org/x8533.xml

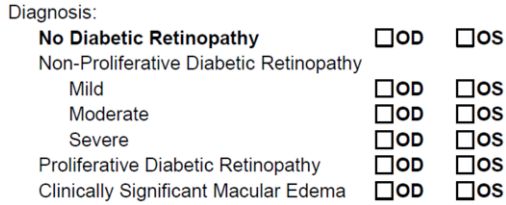

Figure 4- Diagnosis section of the American Optometric Association (AOA) Diabetes Eye Examination Report form

The eye involved can be represented by specification of the *finding site* as *retina of right eye* or *retina of left eye*. Although involvement of both eyes could be represented using *retina of both eyes*, this is discouraged, as this lacks explicit reference to the retina of the left or the right eye.

A further challenge is introduced by the existence of the pre-coordinated concept *severe nonproliferative diabetic retinopathy with clinically significant macular edema*. To record this concept it is required that when the boxes for "severe" and for "Clinically Significant Macular Edema" are selected for the same eye, this shouldn't result in two diagnoses, but in the use of the pre-coordinated concept.

Finally, when selecting "Clinically Significant Macular Edema" it needs to be taken into account that this is in the context of diabetes. Hence, it needs to be represented as the combination of *clinically significant macular edema* and *diabetic macular edema*.

Encoding meaning from natural language processing

Use of natural language processing (NLP) to capture either pre- or post-coordinated SNOMED CT expressions provides a means of overcoming the restrictions that are introduced by encoding using pre-defined hierarchical structures or forms. Performing NLP in pseudo-real-time becomes possible [9] [10], but various challenges remain, even when words in a fragment of natural language exactly match descriptions in SNOMED CT.

The first challenge is disambiguation. In SNOMED CT, homonymy is a frequently occurring phenomenon. For many body structures, both the "structure" and the "entire" share a term, e.g., "heart" is a description for both *heart structure* and *entire heart*. Also between different categories of concepts homonymy occurs, as is demonstrated by the terms "burn" and "bleeding", which are related to both a morphologic abnormality and a clinical finding. Likewise, many concepts in the substance hierarchy have descriptions that are homonymous with descriptions of concepts in the product hierarchy of SNOMED CT.

The second challenge is to determine the relationship between concepts forming a post-coordinate expression. Whereas the concept model provides allowed relationships, this is not sufficient. Additional editorial rules need to be taken into account. For example, according to editorial rules, in the International Release of SNOMED CT, pharmaceutical / biologic product concepts are not used as values for a *direct substance* relationship, although this is allowed by the concept model.

Whereas there are projects in which the use of NLP in post-coordination is applied, we are unaware of evaluations of the quality of the created expressions, other than in [10].

Discussion

The results of our analyses show that there are various improvements needed to adequately enable post-coordination.

First, post-coordination provides too many possibilities, both in the number of relationships that can be specified, and in the concepts that can be selected as refined values. This severely hampers usability of user interfaces enabling post-coordination, and allows for non-probable, non-sensible and nonsensical post-coordinate expressions.

Second, harmonizing the presentation of items on a user interface with the SNOMED CT concepts they represent requires dynamic mapping.

Third, the knowledge represented in a post-coordinate expression need validation regarding editorial rules and modeling patterns.

Tailored post-coordination

To increase applicability of relationships and their allowed values for supporting user-directed coordination of concepts and natural language processing, a more fine-grained specification of refinability is needed. Lessons can be learned from the GALEN project, a predecessor of much of terminology- and ontology-related work today, which introduced a method and syntax for representation of a meta-model of the ontology, corresponding to the SNOMED CT concept model. This facility of GALEN was called the sanctioning mechanism [11]. This mechanism is used to constrain which GALEN expressions are valid and which are invalid in order to avoid nonsensical concepts as exemplified above. There were two levels of sanctioning: a generic or grammatical level and more specific or sensible level. Grammatical sanctions constrain the use of high-level classes comparable to statements of the SNOMED CT concept model while sensible sanctions are used for more fine-grained constraints such as constraining fractures to be located in bones thereby disallowing fractured lips. There are currently no constraints corresponding to the sensible level in the SNOMED CT concept model. The 2010 release of the GALEN meta-model consists of about 280 grammatical sanctions and about 23,000 sensible sanctions [4]. Thus, a significant amount of content would need to be defined to allow more specific constraints. A piecemeal development of this content could be realized by specification of so-called "relationship refinability reference sets", as explained in section 7.4.3.1 of [8].

User Interface and reference terminology binding

To enable automated coding of filled-in structured data entry forms, context needs to be taken into account, and rules must be implemented to enable dynamic mapping.

Taking into account context is relatively straightforward, as this can be determined from the information on the form.

The implementation of rules is more complex, and the extent to which this can be successful depends strongly on the quality of definitions in SNOMED CT. In the ophthalmology example provided above, the consistency of definitions is insufficient.

For example, one of the descriptions of the concept *clinically significant macular edema* is "clinically significant diabetic macular edema". However, the concept is not defined as a subtype of *diabetic macular edema*. So either the description referring to diabetes is incorrect, or the association with diabetes needs to be explicitly added. Adding this association would however not render the concept as a type of *diabetic macular edema*, as the latter concept is primitively defined.

As an alternative, considering that the diagnosis is determined by examination of the patient, the check boxes could be mapped to concepts *on examination - clinically significant macular edema of right eye* and *of left eye* respectively. However, these concepts are defined inconsistently, as only the finding for the right eye is a subtype of *diabetic macular edema*, and neither is a subtype of *clinically significant macular edema*.

Many more examples of inconsistencies are described in [12], stressing the need for consistent modeling. If and only if the consistency of modeling increases, reasoning can be adequately used to implement the rules for dynamic mapping, as will be discussed below.

Analysis of post-coordinate expressions

Post-coordinate expressions cannot be taken "as is". Various steps need to be taken to assess the validity, and if possible correctness and completeness of the expressions.

The first step is to apply editorial rules which are currently in place for the modeling of the International Release of SNOMED CT, as modeling is not only dependent on the classification of expressions. The process also involves so-called Generic Concept Inclusion axioms (GCIs). These axioms provide the possibility of adding relationships to concepts, or specifying values for relationships. An example of the first is the rule: if *causative agent* is a *microorganism* and *associated morphology* is an *inflammation* then add relationship *pathological process = infectious process*.

An example of the second is the rule: if *clinical course* is a kind of *sudden onset AND/OR short duration* and *associated morphology* is an *inflammation* then refine *associated morphology* as *acute inflammation*.

The second step is to validate whether the expression adheres to the concept model, and any other sanctioning, e.g., the relationship refinability reference sets described above.

The third step is to classify the expression, and determine equivalent concepts as well as sub-concepts. This can reduce the number of post-coordinate expressions that are equivalent to pre-coordinated concepts, and can point the user at other relevant concepts.

Two important things need to be taken into account. Some concepts in SNOMED CT lack certain relationships (e.g., the association with diabetes mentioned above), and any post-coordinate expression will never be classified as a subtype of a primitive concept. To maximize the likelihood of determining relevant concepts, performing classification with partial definitions (i.e., ignoring one or more of the relationships) can be considered. More research is needed on this, as the amount and the potentially relevant concepts suggested to a user need to be adequate. Fortunately, software to classify SNOMED CT in pseudo-real-time becomes available [13]. Additionally, post-coordination by combination could be more structurally applied. Whereas this method may not be as intuitive to users as qualification or refinement, the example in Figure 3 demonstrates that this provides the most complete, hence likely most correct, concept expression. This combined concept can then be transformed into an expression that applies post-coordination by qualification.

Related research

The findings are in line with previous findings from an extended literature review [14], in which 3 "consequences of post-coordination in a clinical terminology" are provided.

Clinically nonsensical concepts are still not prevented with the current approach.

Concept duplication, i.e., having multiple ways to represent a concept, is in principle taken care of by the description logic underlying SNOMED CT, which enables detection of logically equivalent definitions. However, we demonstrated in Figure 3 that in practice seemingly equivalent representations

[4] http://opengalen.org/download/opengalen_2-8_metrics.zip

may in fact be logically different, in which case the logic is to no avail.

Inefficiency of Concept Composition is not yet overcome, as refinement and qualification provide too many rarely relevant concepts.

Future work

To realize adequate user-directed coordination, two aspects need further research.

The first aspect is the user experience. Sanctioning is needed for user-directed coordination, both to prevent nonsensical combinations and to present only relevant concepts for qualification or refinement. However, sanctioning is in part context-dependent. Concepts which may be relevant in one context, such as the close to 200 serotypes of *Escherichia coli*, may be too detailed in another context. On the other hand, there are cases in which sanctioning can be reused, e.g., to prevent specification of *structure of deep medullary artery of the brain* as *finding site* for *headache*. Research is needed to distinguish specific sanctioning from context-dependent sanctioning, and to devise methods that prevent the need to specify sanctions for each individual concept.

The second aspect is the concept definition. Whereas there are currently ways to determine whether different concept expressions are logically equivalent, these methods are neither fully automated nor integrated in tools that support user-directed coordination. Further research is needed to automate these methods, provide explanation as well as alternative representations, including those which apply post-coordination by combination, to users.

Whereas the focus in this paper has been on user-directed coordination, the findings are also relevant for other settings in which post-coordinate expressions are developed. In many countries national release centers for SNOMED CT develop and maintain national extensions, and are involved in creating mapping between various healthcare terminologies and SNOMED CT. Also in these activities, support for and quality assessment of post-coordination are essential.

Conclusion

Post-coordination enables clinicians and national release centers to construct highly specific expressions that maximally capture the intended meaning in a formalized manner. Current mechanisms for post-coordination, i.e., the concept model and refinement, are insufficient for providing guidance in construction of complete and correct expressions. Implementing fine-grained sanctioning can overcome this. The resulting post-coordinate expressions need to be continuously scrutinized to prevent missed inferences, including classification and automated application of editorial rules. SNOMED CT content requires further improvement to ensure consistent modeling.

Acknowledgments

Authors are member of the Technical Committee (RC), Implementation & Innovation Committee (MN), and Quality Committee (DK) of the International Health Terminology Standards Development Organization (IHTSDO), which publishes SNOMED CT. Their positions at the IHTSDO, however, had no bearing on the research study or results.

References

[1] Cornet R, de Keizer N. Forty years of SNOMED: a literature review. BMC Med Inform Decis Mak. 2008; 8(Suppl 1): S2.

[2] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med. 1998; 37(4-5): 394-403.

[3] Elkin PL, Tuttle M, Keck K, Campbell K, Atkin G, Chute CG. The role of compositionality in standardized problem list generation. Stud Health Technol Inform. 1998; 52(pt 1): 660-4.

[4] Richesson RL, Andrews JE, Krischer JP. Use of SNOMED CT to represent clinical research data: a semantic characterization of data items on case report forms in vasculitis research. J Am Med Inform Assoc. 2006; 13(5): 536-46.

[5] Andrews JE, Patrick TB, Richesson RL, Brown H, Krischer JP. Comparing heterogeneous SNOMED CT coding of clinical research concepts by examining normalized expressions. J Biomed Inform. 2008; 41(6): 1062-9.

[6] de Keizer NF, Bakhshi-Raiez F, de Jonge E, Cornet R. Post-coordination in practice: evaluating compositional terminological system-based registration of ICU reasons for admission. Int J Med Inform. 2008; 77(12): 828-35.

[7] Bakhshi-Raiez F, de Keizer NF, Cornet R, Dorrepaal M, Dongelmans D, Jaspers MW. A usability evaluation of a SNOMED CT based compositional interface terminology for intensive care. Int J Med Inform. 2012; 81(5): 351-62.

[8] IHTSDO. SNOMED CT Technical Implementation Guide; July 2012. 536.

[9] Patrick J, Budd P. Automatic Conversion of Clinical Notes into Snomed CT at Point Of Care. In Proceedings of HIC2006 and HINZ2006; 2006; Brunswick East, Vic.: 5.

[10] Patrick J, Nguyen D, Wang T, Paoloni R. Computational recognition of SNOMED CT codes from ED case notes. Stud Health Technol Inform. 2012; 178: 175-9.

[11] Rector AL, Nowlan WA, Glowinski A. Goals for Concept Representation in the GALEN project. In Proc Annu Symp Comput Appl Med Care 1993; Washington, DC. 414-8.

[12] Rector AL, Brandt S, Schneider T. Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications. J Am Med Inform Assoc. 2011; 18(4): 432-40.

[13] Dentler K, Cornet R, ten Teije A, de Keizer NF. Comparison of Reasoners for large Ontologies in the OWL 2 EL Profile. Semantic Web. 2011; 2(2): 71-87.

[14] Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc. 2006; 13(3): 277-88.

Address for correspondence

Ronald Cornet, PhD r.cornet@amc.uva.nl
Academic Medical Center, Dept of Medical Informatics
P.O. Box 22700 1100 DE Amsterdam, The Netherlands