ONTOLOGY-BASED CONVERGENCE OF MEDICAL TERMINOLOGIES: SNOMED CT AND ICD-11

Schulz S[1,2], Rector A[3], Rodrigues JM[4,5], Chute C[6], Üstün B[7], Spackman K[8]

Abstract
In order to support semantic interoperability in eHealth systems, domain terminologies need to be carefully designed. SNOMED CT and the upcoming ICD-11represent a new generation of ontology-based terminologies. The proposed alignment of these two systems and in consequence the validity of their cross-mappings requires a thorough analysis of the intended meaning of their representational units. We juxtapose and formally dissect two competing interpretations, viz. Condition *vs.* Situation*, and hypothesize that the latter is better suited for both terminologies in their respective use context.*

Keywords *–* ***ICD, SNOMED CT, Ontology***

1. Introduction

eHealth systems are as useful as they enable the exchange of data across system boundaries. System boundaries tend to constitute barriers to syntactic and semantic interoperability. In the last two decades, standardized messaging protocols like HL7 version 2 have contributed to cross the barrier of syntactic interoperability. The challenge is now to exchange not only data but meaning. Semantic interoperability relies on structured data which are annotated by some kind of vocabulary or terminology system. As multiple of such systems coexist, the preservation of the meaning between patient-related data annotated with a term from a vocabulary V_1 with corresponding terms from a vocabulary V_2 requires semantic mappings between these two systems.

1. 1. Applied Ontology

The practice of Applied Ontology [1] is gradually being established as a foundation for building a new generation of biomedical terminology systems. In the nineties, the GALEN approach had paved the way towards a formal description of the meaning of medical terms [2]. Since 2000, tools, techniques, and standards have evolved in the Semantic Web community, with the OWL language

[1] Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria
[2] Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany
[3] University of Manchester, UK
[4] University of Saint Etienne, CHU, Department of Public Health and Medical Informatics, Saint Etienne, France
[5] INSERM UMR 872 Eq 20, Paris, France
[6] Mayo Clinic, Minnesota, US
[7] World Health Organization, Geneva
[8] International Health Terminology Standard Development Organisation, Copenhagen, Denmark

[3] as a prominent description formalism, which draws on Description Logics [4]. These foundations have been readily taken up by the bio-ontology community with Gene Ontology [5] as their flagship. Backbones of ontology artefacts are taxonomic orders of classes. In contrast to thesauri, such as MeSH [06], the key question for justifying a hierarchical link between A and B is not "*does the term/concept B have a broader sense in natural language than the term/concept A*?". Instead, the question must be posed as follows: "*are all members of the class A members of the class B at all time?*", in analogy to the subset relation in set theory (hence we use the element operator $\in$ for class membership). Only if the answer is positive, *A* qualifies as a taxonomic descendent of *B*, or, in other words, a subclass relation holds between *A* and *B*:

$$subClassOf\,(\mathrm{X}, \mathrm{Y}) =_{\mathrm{def}} \forall i\colon i \in X \rightarrow i \in Y$$

(We refrain from time-indexing the class membership relation, taking the quantification over all time instants for granted.)

1. 2. ICD – SNOMED Harmonization

The World Health Organization (WHO) has initiated the revision process of the International Classification of Diseases (ICD-11) in 2007 [7]. Different from past revisions done by WHO FIC collaborating centres, the ICD-11 authoring process, which involves a large community of clinical experts, is supported by ontology-driven tools [8]. Another difference is the distinction between a multi-hierarchical ICD *foundation component* (FC) as the basis for target specific *linearization* products. Also in 2007the International Health Terminology Standard Development Organisation (IHTSDO) adopted the development, maintenance, and distribution of SNOMED CT [9] as an international clinical terminological standard. Ontological principles increasingly guide SNOMED CT's ongoing development [10].

In 2010, following an institutional agreement between WHO and IHTSDO, a joint advisory group (JAG) for the harmonisation between ICD-10, ICD-11 and SNOMED CT was established. A major goal of this group is the creation of a common ontological basis of both ICD11 and SNOMED CT. As intended outcome of this process the following goals have been identified (*Figure 1*):

1. Each class in the ICD-11 foundation component will correspond to exactly one class in SNOMED CT.
2. The equivalence in meaning between these class pairs will be assured by common text definitions.
3. The transitive closure of taxonomic (subclassOf) relations in ICD-11-FC is included in the transitive closure of subClassOf relations in SNOMED CT.

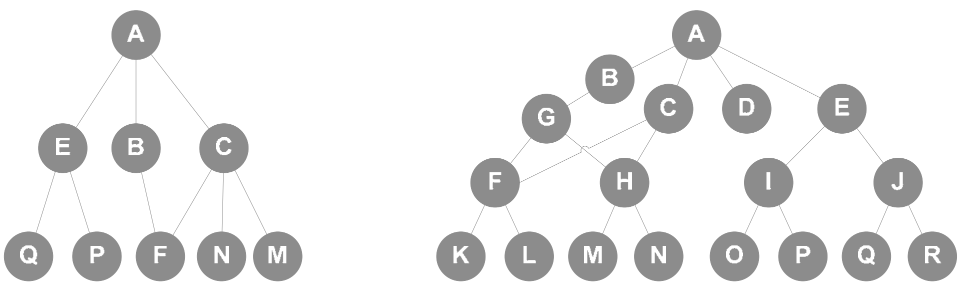

Fig. 1: Extract from ICD-11-FC (left) and SNOMED CT (right).
Each ICD class corresponds to exactly one SNOMED class (symbolized by the same letter).
SubClassOf - links contained in the left but not in right graph can be obtained by transitive closure.

In the following we demonstrate, by the analysis of two exemplar cases, the challenge of this approach. We highlight that without ontological scrutiny, i.e. without getting to the very bottom of what terms or concepts really signify in their application context, consensus can hardly be reached.

2. Case study

The following case study is centred on the adequacy of taxonomic links in both ICD and SNOMED CT. Taxonomic links, i.e. relations between classes and their superclasses, are of utmost importance for data aggregation and retrieval, as a search for a superclass will return all entities which are members of their subclasses. For instance, if I search for "cardiovascular disorder" I want to retrieve all individual instances of, e.g. "arterial hypertension", "congestive heart failure", "myocardial infarction" and all other classes which are placed under "cardiovascular disorder".

2. 1. Examples

We will base the subsequent deliberation on the following two examples, *viz. Tetralogy of Fallot* and *Hypertensive retinopathy*. These examples and the related representational problems are typical for thousands of SNOMED CT and ICD classes.

- *Tetralogy of Fallot* is a multiple malformation of the heart aggregating the four conditions *Pulmonic valve stenosis, Overriding aorta, Ventricular septal defect* and *Right ventricular hypertrophy*. *Pulmonic Stenosis*, which is a narrowing of the Pulmonic artery can occur as a result of a variety of congenital and acquired causes.
- *Hypertensive Retinopathy* is a disease of the retina, caused by elevated arterial blood pressure (*Hypertension*).

A crucial question for ontology based terminology building is the following: Is it correct to place *Tetralogy of Fallot* under *Pulmonic valve stenosis*, and is it appropriate to place *Hypertensive Retinopathy* under *Arterial Hypertensive Disorder*?

If we look into the current versions of ICD-10 and SNOMED CT, the picture is heterogeneous. In SNOMED CT, *Pulmonic valve stenosis* (ID:56786000) is a superclass of *Tetralogy of Fallot* (ID:86299006), but *Hypertensive Retinopathy* (ID:6962006) is not under *Hypertensive disorder, systemic arterial* (ID:38341003).

In ICD-10, *Tetralogy of Fallot* (Q21.3) is under *Congenital malformations of cardiac septa* (Q21), but not under Pulmonic valve stenosis (I37.0). *Hypertensive Retinopathy* (H35.0) is under *Other retinal disorders* (H35) but not under *Hypertensive diseases* (I10-I15).

2. 2. Conditions vs. Situations

We have identified the following diverging interpretation of clinical terms, especially related to diseases:

(Clinical) *Conditions* are patient-related body processes, states, dispositions, or (patho-)anatomical structures, which are reportable in the context of medical records [11], such as my myopia, John's pulmonic stenosis, Mary's retinopathy. In contrast,

Situations are phases of the life of a patient, during which he/she is bearer of a clinical condition. Clinical situations may exist inside and outside episodes of care. They can be short (seizure) or life-long (congenital malformation). The pathological conditions borne by a patient in a given situation

can be single or multiple pathological dispositions, pathological processes, and pathological body structures.
As an example, the class *Pulmonic valve stenosis* could be interpreted as

- *Pulmonic valve stenosis* [Condition]
  is a class that has as members all narrowings of some Pulmonic artery
- *Pulmonic valve stenosis* [Situation]
  is a class that has as members all phases of a person's life in which he/she is bearer of some *Pulmonic valve stenosis* [Condition]

There are two different possible criteria for one classification of terminology code subsuming (being a taxonomic ancestor of) another. This applies equally to ICD and SNOMED. Under the situations interpretation, we look on a complete phase of a person's life. E.g., if it contains the condition *C1*, then this phase is of the type C1[Situation]. If all instances of *C1* imply the existence of some instances of *C2* then all *C1* situations are also *C2* situations. In our examples: every condition of the type *Tetralogy of Fallot* implies some *Pulmonic valve stenosis*, and every condition of the type *Hypertensive Retinopathy* implies some *Arterial Hypertensive Disorder.*

For which application cases does this distinction matter?

- Expanding queries. When I retrieve all cases of *Pulmonic valve stenosis* what cases do I get? Do the cases retrieved include *Tetralogy of Fallot*? When I retrieve cases of *Arterial Hypertensive Disorder* does it include cases in which *Hypertensive Retinopathy* has been coded but not *Arterial Hypertensive Disorder* per se?
- Inheriting facts. When I say something about all cases of *Pulmonic stenosis* - e.g. that it causes strain to the right ventricle - what does it apply to? Similarly when I say something about *Arterial Hypertensive Disorder*, e.g. that it can cause renal disease, does this apply to *Hypertensive Retinopathy*? But if a make a universal statement about treatment episodes with *Hypertensive Retinopathy, can I infer that these patients are at risk of having renal diseases?*

2. 3. Formal reconstruction

In the following we give a formal account of how situations and conditions are related, using first order logics. All formula can be easily translated into description logics.

2.3. 1. Subclass relations between conditions → Subclass relations between situations

Generally we can infer: A situation C_{Sit} is defined as including some condition C_{Cond}, and every situation including some D_{Cond} is a D_{Sit}. If every C_{Cond} is a D_{Cond} then every C_{Sit} is a D_{Sit}.

$$\forall c \in C_{Sit} \leftrightarrow c \in Situation \land \exists c_1 \in C_{Cond} \land \textbf{includes}(c, c_1)$$
$$\forall d \in D_{Sit} \leftrightarrow d \in Situation \land \exists d_1 \in D_{Cond} \land \textbf{includes}(d, d_1)$$
$$\forall x \in C_{Cond} \rightarrow x \in D_{Cond}$$

$$\forall x \in C_{Sit} \rightarrow x \in D_{Sit}$$

In this case, the interpretation of classes as situations or conditions has no impact in the validity of the subclass relations between them: If two condition classes stand in a subclass relations, then this relation holds between the derived situation classes, too. Example: If *Pulmonic valve stenosis* is a subclass of *Heart valve Stenosis*, then a *situation with Pulmonic valve stenosis* is also a *situation with Heart valve stenosis*.
2.3. 2. Has-part relations between conditions → Subclass relations between situations

Generally we can infer: A situation C_{Sit} is defined as including some condition C_{Cond}, and every situation including some D_{Cond} is a D_{Sit}. Every C_{Cond} has some D_{Cond} as part, i.e. it includes it. As every C_{Sit} includes some C_{Cond} , and every C_{Cond} includes some D_{Cond}, every C_{Sit} includes some D_{Cond}, due to the transitivity of the **includes** relation. Situations that include some D_{Cond} are of the type D_{Sit}. Therefore, every C_{Sit} is a D_{Sit}.

$$\forall c \in C_{Sit} \quad \leftrightarrow \quad c \in Situation \wedge \exists c_1 \in C_{Cond} \wedge \textbf{includes}\ (c, c_1)$$
$$\forall d \in D_{Sit} \quad \leftrightarrow \quad d \in Situation \wedge \exists d_1 \in D_{Cond} \wedge \textbf{includes}\ (d, d_1)$$
$$\forall x \in C_{Cond} \quad \rightarrow \quad \exists y \in D_{Cond} \wedge \textbf{has-part}\ (x, y)$$
$$\forall x,y\text{: } \textbf{has-part}\ (x, y) \quad \rightarrow \quad \textbf{includes}\ (x, y)$$
$$\forall x,y,z\text{: } \textbf{includes}\ (x, y) \wedge \textbf{includes}\ (y, z) \quad \rightarrow \quad \textbf{includes}\ (x, z)$$

$$\forall x \in C_{Sit} \quad \rightarrow \quad x \in D_{Sit}$$

Example: The condition *Tetralogy of Fallot* has some *Pulmonic Valve Stenos*is as part. Therefore every situation including *Tetralogy of Fallot* is also a situation including *Pulmonic Valve Stenosis.* This example shows that partitive relations between condition classes feature as taxonomic relations between the corresponding situation classes. Under the Situation interpretation we therefore expect much more taxonomic links.

2.3. 3. Ontological dependency between conditions → Subclass relations between situations

It is tempting to postulate, according to what we discussed in the previous paragraph, that if every instance of C_{Cond} entails some D_{Cond} (e.g. by causation), C_{Sit} is a D_{Sit}. For instance, every instance of *Hypertensive Retinopathy* is caused by some instance of *Arterial Hypertensive Disorder*. Then every situation with *Hypertensive Retinopathy* would be a situation with *Arterial Hypertensive Disorder*. In a taxonomy that subscribes to the situation paradigm, *Arterial Hypertensive Disorder* would surface as a parent of *Hypertensive Retinopathy*.

This pattern is only generalizable under the conditions that for all time instants C_{Cond} is present D_{Cond} is present, too. This is the case with the present example, but we can find many counterexamples. For instance, it depends on the framing of what we understand by *Situation*, whether e.g. *Conductive hearing loss due to disorder of middle ear* (interpreted as situation) can legitimately considered a subclass of *Disorder of middle ear*.

3. Conclusion

The ICD-SNOMED CT harmonization activities have brought to surface that the consensus process crucially depends on a common understanding what the classification / terminology codes under scrutiny actually denote. Under the hypothesis that SNOMED CT concepts or ICD codes denote the clinical conditions proper (as a literal interpretation of their wording suggests), much less taxonomic links can be expected than under the hypothesis that they are used to denote patients (having the condition) or clinical situations (including that condition). It is obvious that this has a direct impact on the use of these systems for data retrieval or aggregation.

The current structure of most formal disease definitions in SNOMED CT (using description logics) suggests that the situation reading can be taken as a default assumption. The main use cases for ICD indicate the same. One could argue that the best solution would be to leave open of whether a code should be interpreted as a condition or a situation, depending on its context of use. Ontologically, this would correspond to disjunctive expressions such as 'C_{Sit} or C_{Cond}'. However, in this case only those taxonomic links would hold in all cases which are valid under the condition expression: Otherwise, given *Tetralogy of Fallot*$_{Sit\ or\ Cond}$ subClassOf *Pulmonic valve stenosis*$_{Sit\ or\ Cond}$, an instance of the former that happens to be a condition would be also an instance of the latter, and would therefore be, by disambiguation, an instance of *Pulmonic valve stenosis*$_{Cond}$, which contradicts our above assumptions.

The JAG has, by now, a preference to the situation interpretation as a common ontological commitment for both ICD-11-FC and SNOMED CT. More evidence is expected from the planned analysis of a representative sample of SNOMED CT concepts.

4. References

[1] Smith B. Applied Ontology: A New Discipline is Born, Philosophy Today, vol. 12, number 29 (1998), 5–6.

[2] Rector A, Rossi A, Consorti MF, Zanstra P. Practical development of re-usable terminologies: GALEN-IN-USE and the GALEN Organisation. International Journal of Medical Informatics 1998 Feb; 48(1–3):71–84.

[3] W3C OWL Working Group: OWL 2 Web Ontology Language Document Overview 2009, [http://www.w3.org/TR/owl2-overview/].

[4] Baader F et al. The Description Logic Handbook: Theory, Implementation, and Applications, 2nd Edition. Cambridge University Press, 2nd ed. 2007.

[5] Ashburner M et al. Gene Ontology: tool for the unification of biology. Nature Genetics 2000, 25:25

[6] Medical Subject Headings. National Library of Medicine [http://www.nlm.nih.gov/mesh/]

[7] Tudorache T, Falconer S, Nyulas C, Storey MA, Ustün TB, Musen MA. Supporting the Collaborative Authoring of ICD-11 with WebProtégé. Proceedings of the AMIA Annual Symposium 2010, 802–806.

[8] Rodrigues JM, Kumar A, Bousquet C, Trombert B. Using the CEN/ISO standard for categorial structure to harmonise the development of WHO international terminologies. Studies of Health Technology and Informatics 2009; 150:255–259.

[9] International Health Terminology Standards Development Organisation (IHTSDO): SNOMED CT: [http://www.ihtsdo.org/snomed-ct/]

[10] Schulz S, Cornet R; Spackman K. Consolidating SNOMED CT's ontological commitment. Applied Ontology 2011; 6: 1–11.

[11] Schulz S, Spackman K, James A, Cocos C, Boeker M. Scalable representations of diseases in biomedical ontologies. Journal of Biomedical Semantics 2011 May 17;2 Suppl 2:S6.

Corresponding Author
Stefan Schulz
Institut für Medizinische Informatik, Statistik und Dokumentation
Medizinische Universität Graz, Auenbruggerplatz 2/V
A-8036 Graz, Austria
Email: stefan.schulz@medunigraz.at