SNOMED CT's Ontological Commitment

Stefan Schulz[1], Ronald Cornet[2]
[1] IMBI, University Medical Center, Freiburg, Germany
[2] Academic Medical Center, Amsterdam, The Netherlands

Abstract

SNOMED CT is a clinical terminology that describes the meaning of terms by logical axioms. This requires an ontological commitment, i.e. precise agreements about the ontological nature of the entities referred to. We provide evidence that SNOMED implicitly supports at least three different kinds of commitments, viz. (i) independently existing entities, (ii) representational artifacts, and (iii) clinical situations. Our analysis shows how the truth-value of a sentence changes according to one of these perspectives. We argue that a clear understanding of to what kind of entities SNOMED CT concepts extend is crucial for the proper use and maintenance of SNOMED CT.

Introduction

SNOMED CT[1] is the inheritor of a dynasty of medical nomenclatures and coding systems[2] which had been constructed to provide

1. semantic descriptors to annotate and encode clinical procedures, diagnoses, etc.;
2. standardized medical terms in different languages;
3. guidance for the construction of composed terminological expressions.

SNOMED CT's predecessors made only very basic claims with respect to the domain they represented. The meaning of semantic descriptors was given by the intuitive understanding of the terms they were linked to and it was assumed that they were correctly interpreted by the (human) language users in the communication process. Therefore, none of these systems made any attempt to formally represent any reality beyond a rough mapping of controlled terms to shared concepts with the aim to reduce the high variability of human language through a set of controlled terms or to support the encoding of medical data by means of a coded thesaurus of procedural and administrative terms for the electronic health record.

With the advent of SNOMED RT (and later SNOMED CT), logics entered the scene and added a mathematically rigorous layer to the hitherto informal, close-to-human-language representation of medical terms. However, the use of logic axioms and theorems in a terminology (which imposes the assignment of truth-values) requires an equally precise agreement about the objects and relations being denoted by the terms and concepts, a so-called ontological commitment[3].

In this paper we will substantiate the claim that SNOMED CT's ontological commitment is inconsistent. To this end we will scrutinize three frequent SNOMED CT design features, viz. (i) qualifiers and their values, (ii) context-dependent concepts, and (iii) multiple parenthood.

Furthermore, we will discuss the pros and cons of the inferences they enable and discuss them in the light of competing ontological commitments.

Description Logics

SNOMED CT's backbone is given by a taxonomy of nodes, so-called SNOMED CT concepts. Every concept represents the characteristic properties of all its (concrete) instances. This is done by description logics (DL)[4], which we will briefly introduce. Key notion are concepts (classes) and instances (their extensions). So is the class *Liver* instantiated by every individual liver, just as *Bodily Organ* extends to all individual bodily organs. Putting those two statements together, we get the hierarchy-building principle of taxonomic subsumption: *Liver* is subsumed by *Bodily Organ.* In DL this is expressed by *Liver* $\sqsubseteq$ *Bodily Organ*, which asserts that every *Liver* instance is also an instance of *Bodily Organ*[1].

More complex statements can be obtained by combining representations of classes with operators and quantifiers. In the following example, we employ the $\sqcap$ ("and") operator and add a quantified role, using the existential quantifier $\exists$ ("exists"). For example, the expression *Inflammatory disease* $\sqcap \exists$ *has location.Liver* extends to all instances in which both instantiate *Inflammatory disease* and are further related through the relation *has location* to some *Liver*. This example actually gives us both the necessary and the sufficient conditions needed in order to fully define a class, e.g.: *Hepatitis* $\equiv$ *Inflammatory disease* $\sqcap \exists$ *has-location.Liver*, with the equivalence operator $\equiv$ telling that (i) every particular instance of *Hepatitis* is

[1]Another way of referring to taxonomic subsumption (DL operator $\sqsubseteq$) is the use a relation named "is-a"

also an instance of *Inflammatory disease* that is located at some instance of *Liver*, and (ii) that every instance of *Inflammatory disease* that is located at some *Liver* is an instance of *Hepatitis*. Hence, in any situation, the term on the left can be replaced by the expression on the right without loss of meaning.

Running examples

Having introduced SNOMED CT's formal background we will base our forthcoming deliberations on four examples taken from the January 2009 release of SNOMED CT. All three examples are representative as the phenomena they incorporate occur frequently throughout the terminology.

1. ***Infusion pump*** (430033006), a primitive concept in the *Physical object* branch: *Infusion pump* ⊑ *Pump* ⊑ *Instrument, device* ⊑ … ⊑ *Physical object.*
All SNOMED CT concepts are inserted in this kind of subsumption hierarchies.

2. ***Denied tonsillectomy***. (173422009|: 272125009|= 82975001), a postcoordinated concept, refining tonsillectomy by using the qualifier "priority" with the value "denied", in DL notation: *Tonsillecomy* ⊓ ∃ *Priority.Denied.*
For all SNOMED CT procedure concepts (~50,000) analogous expressions can be created.

3. ***Heart operation planned*** (183983001)[2].
This concept is in SNOMED CT's *Situation with explicit context* branch and is fully defined as
∃ *rg.*(
∃ *Associated procedure.Operation on heart* ⊓
∃ *Procedure context.Planned* ⊓
∃ *Subject rel context. Subject of record* ⊓
∃ *Temporal context.Current or specified time*)
There are currently 17 concepts with the context *Planned*, but numerous others with similar contexts such as *Suspected* or *Known absent.* ,

4. ***Tetralogy of Fallot***. (86299006).
This concept is a child of the four concepts:
Tetralogy of Fallot ⊑ *Pulmonic valve stenosis*
Tetralogy of Fallot ⊑ *Ventricular septal defect*
Tetralogy of Fallot ⊑ *Overriding aorta*
Tetralogy of Fallot ⊑ *Right ventr. hypertrophy*
More precisely, it implies the following expression:
∃ *rg.*(∃ *assoc_morphology.Congenital Anomaly* ⊓
∃ *has_location.Cardiac Ventricular Structure*) ⊓
∃ *rg.* (∃ *assoc_morphology.Defect* ⊓
∃*has_location.Intraventricular Septum Structure*) ⊓
∃ *rg.* (∃ *assoc_morphology.Stenosis* ⊓
∃*has_location.Pulmonary Valve Structure*) ⊓
∃ *rg.* (∃ *assoc_morphology.Overriding Structures* ⊓
has_location.Thoracic Aorta Structure)
Nearly 77,000 SNOMED CT concepts contain relationship groups.

Using these examples we want to demonstrate the different ontological commitments underlying SNOMED CT. In other words, we will ask the question: which entities in the clinical context are instantiated by SNOMED CT concepts? Below we present three possibilities: independently existing entities, representational artifacts, and clinical situations.

SNOMED CT concepts are instantiated by objects that exist independently of the clinical context

We will call this the standard interpretation, as it is the most straightforward one and corresponds to the view commonly defended by the realist approach to ontologies. This stance postulates the existence of objects and processes as independent of the circumstances of their observation[5] .
Under this viewpoint, the concept ***Infusion pump*** would be instantiated by each and every individual infusion pump, independent of its involvement in any clinical process. It would not designate the mental concept or construction plan of an infusion pump. In the same line, the concept ***Tonsillectomy*** would be instantiated by every surgically removal of a tonsil in reality, and ***Aortic Stenosis*** by every morphologically altered state of a real aortic valve in a real patient.

SNOMED CT concepts are instantiated by representational artifacts as contained in an electronic patient record

We will call this the EHR interpretation. Under this view it does not matter whether some thing really exists or not. The only criterion is a mention in a documentation artifact such as an electronic patient record. This can be nicely shown by the postcoordinated SNOMED CT concept ***Denied tonsillectomy***. It is not instantiated by a real tonsillectomy but by an EHR entry on tonsillectomy, an information object which may be further refined by qualifiers such as *Denied, Planned, Scheduled.* Under this point of view, an EHR entry "denied tonsillectomy" is indeed subsumed by the entry "tonsillectomy", so that this sentence holds true. How such an entry is interpreted by the EHR used in terms of what things in reality it denotes is not relevant

[2] *rg* means „role group", cf.[6]

here. But it is obvious that on the level of real objects and processes a denied tonsillectomy can never be a kind of tonsillectomy, so that the sentence is false at the level of real objects.

A similar line of reasoning applies to the example ***Heart operation planned***. Although this concept is not a subconcept of *Heart operation* (which would parallel the above example), its standard interpretation leads to contradictions: According to its definition, *Heart operation planned* implies the sentence

∃ *Associated procedure. Operation on heart*.

Following the description logics semantics this means that for each instance of *Heart operation planned* there is at least one instance of *Operation on heart*. This contention can easily be disproved as planned procedures are not always executed. If *Heart operation planned*, on the contrary, is interpreted as to be instantiated by EHR objects, the sentence becomes true, using the same argument we used to support the subsumption relation between *Tonsillectomy* and *Denied tonsillectomy*.

SNOMED CT concepts are instantiated by patients or clinical situations.

Typical examples that suggest this third flavor of interpreting SNOMED CT concepts is suggested by the way SNOMED CT formalizes composed clinical findings and procedures. The standard interpretation conflicts with the fact that all elements of a combined finding, such as the complex heart malformation called *Tetralogy of Fallot* are introduced as its taxonomic parents. As a result, *Tetralogy of Fallot* is subsumed by the concepts *Septal defect* and *Cardiomegaly*, among others. As this is hardly tenable, SNOMED CT separated the findings from the morphology and separated them using so-called relationship groups. According to[6], role groups are expressed in DL as an anonymous relation called *rg*.

Role groups order the elements of a complex concept definition and prevent it from ambiguous associations. If we re-interpret *rg* as *has_part* as proposed by[7] there is little to criticize from an ontological point of view. However, role groups also appear in definitions where the reason is not obvious, e.g.

Pulmonic Valve Stenosis ≡ ∃ *rg*.
(∃ *assoc_morphology.Stenosis* ⊓
∃ *has_location.Pulmonary Valve Structure*)

Let us rephrase this equivalence, taking description logics semantics seriously: *"Every pulmonic valve stenosis has some part which exhibits at least one stenosis somewhere at a pulmonary valve; and everything having some part which exhibits at least one stenosis somewhere at a pulmonary valve is a pulmonic valve stenosis*".

Whereas the first phrase sounds somewhat circular, the second one expands the concept of pulmonic valve stenosis to an extent that each and every condition which is characterized, among other things, by a stenotic pulmonary valve, is subsumed by the concept *Pulmonic Valve Stenosis.* It is no wonder that in the SNOMED CT hierarchy, this concept does not only subsume *Congenital Stenosis of Pulmonary Valve*, but also *Pulmonic Valve Stenosis With Insufficiency*, *Tetralogy of Fallot* and *Pentalogy of Fallot* .

Coming back to the question of ontological commitment: if we understand by the extension of the concept *Pulmonic Valve Stenosis* the pathological structure as it exists in a patient, then we can't but reject the view that a *Tetralogy of Fallot* is a kind of *Pulmonic Valve Stenosis*. What should be criticized here is that implication is mistaken for subsumption: Of course, for every *Tetralogy of Fallot* there is some *Pulmonic Valve Stenosis*. However, this does not mean that *Tetralogy of Fallot* is related to *Pulmonic Valve Stenosis* by taxonomic subsumption.

The puzzle can be solved if we substitute the standard interpretation by what we will call here the epidemiological interpretation. Under this assumption, disorders and finding concepts do not extend to states or processes but to their participants or bearers, i.e. to patients. Hence, *Pulmonic Valve Stenosis* and *Tetralogy of Fallot* are to be read as "patients with a pulmonic valve stenosis" and "Fallot patients". Then the subsumption statement becomes true: Every person affected by a *Tetralogy of Fallot* is also affected by a *Pulmonic Valve Stenosis.*

The picture is also consistent if we assume that these SNOMED concepts extend to clinical situations[8] rather than to particular disorder or states. Consequently, we can argue that every clinical situation that includes a *Tetralogy of Fallot* also includes a *Pulmonic Valve Stenosis*, paralleling the argument that the set of patients with *Tetralogy of Fallot* forms a subset of the bearers of a *Pulmonic Valve Stenosis.*

Even the EHR interpretation makes sense here, as it is plausible that all records annotated by *Tetralogy of Fallot* should by considered as being annotated by *Pulmonic Valve Stenosis*.

Discussion and Conclusions

As much as we may find good explanations for the discussed types of SNOMED CT modeling decisions we raise our concern in view of the fact that the different ontological commitments are completely

implicit and the choice is up to the user. As long there is no agreement on which SNOMED CT concepts extend to objects in clinical reality, to patients, to situations, or to documentation objects, different users may want to express different things by using the same expressions, and misinterpretations may lead to erroneous conclusions. To give just one example: If the same concept is instantiated to express plans (which always bear the possibility of not being realized) on the one hand and to express actions that have been realized, hospital statistics will become unreliable.

SNOMED CT provides the means to represent situative scenarios that include not only negative contexts but also other contextual "moods". This extends the boundary of what a clinical terminology should represent and therefore overlaps with the realm of information models. The resulting problems with double negations have been intensively discussed in the context of TERMINFO[9].

We therefore defend the position that SNOMED CT should always subscribe to what we named standard interpretation, as it makes no background assumptions and is compatible with the approaches pursued by many other biomedical ontologies, e.g. the ones of the OBO foundry. SNOMED concepts should clearly extend to objects in clinical reality, *viz.* the anatomical structures, the diseases, and the procedures as they occur in patients.

Wherever patients, situations, documentation objects or plans are referred to, this should be made clear in the concept name. For queries that target situations or patients as bearer of disorders but not the disorders themselves, SNOMED CT's postcoordination syntax allows to express this, e.g. by ∃ *bearer-of. Pulmonic Valve Stenosis* or ∃ *bearer-of. Tetralogy of Fallot.* Using a right-identity rule such as *bearer-of * has-part -> bearer-of* would then allow to infer that a Fallot patient has a stenosis of the pulmonic valve even if the problematic assertion *Tetralogy of Fallot* ⊑ *Pulmonic valve stenosis* were removed from SNOMED CT. For the encoding of epistemic aspects of the EHR, such as scheduled or cancelled procedures, the consistent use of an information model (e.g. HL7 or openEHR) should be preferred over the idiosyncratic use of logic-based formalism in a clinical terminology.

Furthermore, SNOMED CT should ensure that any qualifier that can be or is attached to concepts is a pure restriction of the concept it qualifies, and not a modification of this concept, as is the case in "*Priority: Denied*". If SNOMED CT aims to be able to address such information, rather than leaving that to the information model, it should be represented in a consistent way. For example, whereas denied tonsillectomy is not a subclass of tonsillectomy, denial of tonsillectomy is a subclass of denial. Allowing post-coordinating denial with a procedure provides a workable way to specify denied procedures using SNOMED CT.

Acknowledgements

The presented work was executed in the context of the DebugIT project, funded by the EU 7th FP (ICT-2007.5.2-217139).

References

1 International Health Terminology Standards Development Organisation (IHTSDO). Systematised Nomenclature of Medicine – Clinical Terms (SNOMED CT). http://www.ihtsdo.org

2 Cornet R, de Keizer N. Forty years of SNOMED: a literature review. BMC Medical Informatics and Decision Making. 2008 Oct 27;8 Suppl 1:S2.

3 Gruber TR. A translation approach to portable ontology specifications Knowledge Acquisition. 1993; 5(2), 199-220.

4 Bader F, Calvanese D, McGuiness DL, Nardi D, Patel-Schneider PF. The Description Logic Handbook Theory, Implementation, and Applications (2nd Edition). Cambridge: Cambridge University Press, 2007.

5 Ceusters W, Smith B. A realism-based approach to the evolution of biomedical ontologies. AMIA Annual Symposium Proceedings 2006; 121-125.

6 Spackman KA, Dionne R, Mays E, Weis J. Role grouping as an extension to the description logic of ONTYLOG, motivated by entity modeling in SNOMED. AMIA Annual Symposium Proceedings 2002; 712-716.

7 Schulz S, Hanser S, Hahn U, and Rogers J. The semantics of procedures and diseases in SNOMED CT. Methods of Information in Medicine. 2006; 45(4):354–358.

8 Rector AL, Brandt S. Why do it the hard way? The case for an expressive description logic for SNOMED. Journal of the American Medical Informatics Association. 2008 Nov-Dec;15(6):744-51

9 Krog R, Markwell D, Dolin RH, Davera G, Cheetham E, Hamm R, Spackman K, Rector A, Huff S, Ryan S. Using SNOMED CT in HL7 Version 3; Implementation Guide, Release 1.0 http://lists.hl7.org/read/attachment/84028/1/ terminfo_20060307.doc