How should parthood relations be expressed in SNOMED CT?

Franz Baader,[1] Stefan Schulz,[2] Kent Spackman,[3] and Boontawee Suntisrivaraporn[4]

[1]TU Dresden, Germany, baader@tcs.inf.tu-dresden.de
[2]Freiburg University Hospital, Germany, stschulz@uni-freiburg.de
[3]International Health Terminology Standards Development Organisation, USA, ksp@ihtsdo.org
[4]SIIT, Thammasat University, Thailand, sun@siit.tu.ac.th

The Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT)[1] is a clinical terminology with a broad coverage of health care, which has been developed with the help of a rather inexpressive description logic dialect known as $\mathcal{EL}$ [1]. The advantage of using a description logic (DL) for defining a medical ontology is that, instead of error-prone "hierarchy engineering," where each newly introduced concept needs to be manually positioned at the right place in the concept hierarchy, one adds a definition of the new concept to the knowledge base and the DL reasoner then automatically finds the right position of this concept in the concept hierarchy. The advantage of using an *inexpressive* DL is that classification (i.e., the computation of the concept hierarchy) is fast even for a very large ontology like SNOMED CT. Efficient reasoners for $\mathcal{EL}$, like Snorocket™,[2] which is based on the classification algorithm introduced in [2], can classify SNOMED CT in less than a minute.

The disadvantage of using an inexpressive DL is that not all relevant properties can be explicitly expressed. In particular, $\mathcal{EL}$ does not allow to state that relations such as $\mathsf{part\text{-}of}$ are transitive, and consequently the reasoner cannot take transitivity into account during classification. In order to overcome such limitations in DLs without transitive relations, the SEP-triplet encoding was proposed in [3]. An SEP-triple for the concept A is actually composed of three concepts: the *s*tructure A_S, the *e*ntity A, and the *p*art A_P. Intuitively, the E-concept is supposed to be instantiated by entire anatomical objects (such as my hand), the P-concept by the proper parts of the referred objects (such as any part of my hand), and the S-concept by both entire objects and their parts. Fig. 1 gives an example of how a correct use of the SEP-triplet encoding should look like. It is easy to see that transitivity of the part-of relation can be simulated through the intra-triple part-of relationships and the intrinsic transitivity of (both intra- and inter-triple) subsumption relationships. In fact, in the example of Fig. 1, the DL reasoner is able to infer that the finger is part of the upper limb since we have $\mathsf{Finger} \sqsubseteq \mathsf{Finger}_S \sqsubseteq \mathsf{Hand}_P \sqsubseteq \mathsf{Hand}_S \sqsubseteq \mathsf{UpperLimb}_P \sqsubseteq \exists \mathsf{part\text{-}of.UpperLimb}$. Since characteristics are inherited along the $\mathsf{is\text{-}a}$ hierarchy, the SEP-triplet encoding also allows us to simulate inheritance of characteristics along the $\mathsf{part\text{-}of}$ hierarchy. In our example, by connecting an injury via a location link to the *S-concept*, we can ensure that 'injury to finger' is classified as 'injury to hand' and 'injury to upper limb'. To suppress such inheritance along the $\mathsf{part\text{-}of}$ hierarchy (e.g., 'amputation of finger' should not be classified as 'amputation of hand' or 'amputation of upper limb'), one needs to connect via location to the *E-concept*. There are, however, several problems with the SEP-triplet encoding. On the one hand, the SEP-triplet approach is error prone since it works correctly only if it is employed with a very strict modelling discipline. For instance, incorrect links to the S-concept rather than the E-concept may result in unintended consequences like the classification of 'amputation of finger' as a subconcept of 'amputation of upper limb'. On the other hand, the approach introduces for every proper concept in the ontology two auxiliary concepts, which results in a drastic increase in the ontology size, and thus in the time needed for classification.

[1] http://www.ihtsdo.org/snomed-ct/
[2] http://aehrc.com/hie/snorocket.html

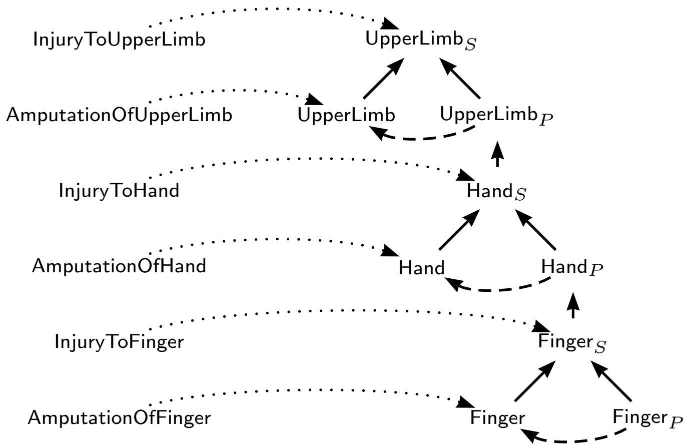

Fig. 1. Example of a correct use of the SEP-triplet encoding. The solid edges denote subsumption (IS-A), the dashed edges part-of, and the dotted edges has-location relationships.

To avoid these problems, we have proposed in [4] to use the more expressive DL $\mathcal{EL}^{++}$ [5, 6], for which classification can still be done in polynomial time. The complex role inclusion axioms available in $\mathcal{EL}^{++}$ can be used to state reflexivity and transitivity of roles like part-of, subrole relationships (e.g., between proper-part-of and part-of), and right-identity rules (which can, e.g., be used to express the inheritance of characteristics along the part-of relation). To avoid unintended inheritance of characteristics (e.g., in the case of amputation), we use two distinct relations: has-location, which is inherited from a part to its whole, and has-exact-location, a subrelation of has-location, which is not inherited that way. Fig. 2 shows the re-engineered ontology obtained this way from the knowledge base of Fig. 1.

This new modelling approach avoids the introduction of the two additional auxiliary concepts (the S-concept and the P-concept) for every anatomical concept. The experiments reported in [4] show that this actually speeds up the time needed for classification. However, for backward compatibility, it would be nice to be able to define the S-concept and/or the P-concept in case it is needed (e.g., since it is used directly in other parts of the ontology). According to the underlying intuition, this should be easy: these concepts can be pre-coordinated as fully defined concepts, as illustrated here for the concept hand: $\mathsf{Hand}_P \equiv \exists\mathsf{proper\text{-}part\text{-}of.Hand}$ and $\mathsf{Hand}_S \equiv \exists\mathsf{part\text{-}of.Hand}$.

$$\mathsf{Finger} \sqsubseteq \mathsf{BodyPart} \sqcap \exists\mathsf{proper\text{-}part\text{-}of.Hand} \tag{1}$$
$$\mathsf{Hand} \sqsubseteq \mathsf{BodyPart} \sqcap \exists\mathsf{proper\text{-}part\text{-}of.UpperLimb} \tag{2}$$
$$\mathsf{UpperLimb} \sqsubseteq \mathsf{BodyPart} \tag{3}$$
$$\mathsf{AmputationOfFinger} \equiv \mathsf{Amputation} \sqcap \exists\mathsf{has\text{-}exact\text{-}location.Finger} \tag{4}$$
$$\mathsf{AmputationOfHand} \equiv \mathsf{Amputation} \sqcap \exists\mathsf{has\text{-}exact\text{-}location.Hand} \tag{5}$$
$$\mathsf{AmputationOfUpperLimb} \equiv \mathsf{Amputation} \sqcap \exists\mathsf{has\text{-}exact\text{-}location.UpperLimb} \tag{6}$$
$$\mathsf{InjuryToFinger} \equiv \mathsf{Injury} \sqcap \exists\mathsf{has\text{-}location.Finger} \tag{7}$$
$$\mathsf{InjuryToHand} \equiv \mathsf{Injury} \sqcap \exists\mathsf{has\text{-}location.Hand} \tag{8}$$
$$\mathsf{InjuryToUpperLimb} \equiv \mathsf{Injury} \sqcap \exists\mathsf{has\text{-}location.UpperLimb} \tag{9}$$
$$\mathsf{proper\text{-}part\text{-}of} \circ \mathsf{proper\text{-}part\text{-}of} \sqsubseteq \mathsf{proper\text{-}part\text{-}of} \tag{10}$$
$$\mathsf{proper\text{-}part\text{-}of} \sqsubseteq \mathsf{part\text{-}of} \tag{11}$$
$$\mathsf{part\text{-}of} \circ \mathsf{part\text{-}of} \sqsubseteq \mathsf{part\text{-}of} \tag{12}$$
$$\epsilon \sqsubseteq \mathsf{part\text{-}of} \tag{13}$$
$$\mathsf{has\text{-}exact\text{-}location} \sqsubseteq \mathsf{has\text{-}location} \tag{14}$$
$$\mathsf{has\text{-}location} \circ \mathsf{proper\text{-}part\text{-}of} \sqsubseteq \mathsf{has\text{-}location} \tag{15}$$

Fig. 2. The re-engineered version of the knowledge base in Fig. 1, now without SEP-triplets.

Unfortunately, this solution (which was already proposed in [4]) is not completely satisfactory since not all subsumption relationships for the auxiliary concepts that follow from the SEP-encoded version of the knowledge base (Fig. 1) follow from the re-engineered version (Fig. 2) extended by the definitions for the S- and P-concepts for Finger, Hand, and UpperLimb. For example, in Fig. 1 we have the (stated) subsumption relationship $\mathsf{Finger}_S \sqsubseteq \mathsf{Hand}_P$. Using the complex role inclusion axioms in Fig. 2 together with the definitions for the auxiliary concepts, we can only conclude $\exists\mathsf{part\text{-}of}.\mathsf{Finger} \sqsubseteq \exists\mathsf{part\text{-}of}.\mathsf{Hand}$ (i.e., $\mathsf{Finger}_S \sqsubseteq \mathsf{Hand}_S$), but *not* $\exists\mathsf{part\text{-}of}.\mathsf{Finger} \sqsubseteq \exists\mathsf{proper\text{-}part\text{-}of}.\mathsf{Hand}$ (i.e., *not* $\mathsf{Finger}_S \sqsubseteq \mathsf{Hand}_P$). In order to obtain the second subsumption, we would need to add the complex role inclusion

$$\mathsf{part\text{-}of} \circ \mathsf{proper\text{-}part\text{-}of} \sqsubseteq \mathsf{proper\text{-}part\text{-}of}.$$

Interestingly, this left-identity rule, together with $\mathsf{proper\text{-}part\text{-}of} \sqsubseteq \mathsf{part\text{-}of}$, creates a so-called cycle over role inclusions, which is not allowed in the DL $\mathcal{SROIQ}$ underlying the new version of the Web Ontology Language, OWL 2. Consequently, OWL 2 compliant reasoners (like FaCT and Pellet) would not accept this extended knowledge base as an input. Fortunately, such a cyclic dependency is allowed in $\mathcal{EL}^{++}$ and can be processed by our reasoner CEL.[3] Recently, Kazakov [7] was able to design a decidable extension of $\mathcal{SROIQ}$ that can also express the extended knowledge base.

To sum up, we have recalled the re-engineering of SNOMED CT as proposed in [4], and have shown that a backward compatible version, which also contains definitions for the auxiliary S- and P-concepts, requires an additional complex role inclusion that destroys the acyclicity property of the set of complex role inclusion. For this reason, the backward compatible re-engineered version of SNOMED CT is not expressible in OWL 2, but it is expressible in $\mathcal{EL}^{++}$ and an appropriate extension of $\mathcal{SROIQ}$.

References

1. F. Baader. Terminological cycles in a Description Logic with existential restrictions. In Georg Gottlob and Toby Walsh, editors, *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 325–330. Morgan Kaufmann, 2003.
2. F. Baader, C. Lutz, and B. Suntisrivaraporn. Is tractable reasoning in extensions of the Description Logic $\mathcal{EL}$ useful in practice? In *Proceedings of the 2005 International Workshop on Methods for Modalities (M4M-05)*, 2005.
3. S. Schulz, M. Romacker, and U. Hahn. Part-whole reasoning in medical ontologies revisited: Introducing SEP triplets into classification-based Description Logics. *Journal of the American Medical Informatics Association (JAMIA)*, pages 830–834, 1998. Section VIII Standards and Policies - Issues in Knowledge Representation.
4. B. Suntisrivaraporn, F. Baader, S. Schulz, and K. Spackman. Replacing SEP-triplets in Snomed ct using tractable Description Logic operators. In Jim Hunter Riccardo Bellazzi, Ameen Abu-Hanna, editor, *Proceedings of the 11th Conference on Artificial Intelligence in Medicine (AIME'07)*, volume 4594 of *Lecture Notes in Computer Science*, pages 287–291. Springer-Verlag, 2007.
5. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proceedings of the 19th International Conference on Artificial Intelligence (IJCAI-05)*, Edinburgh, UK, 2005. Morgan-Kaufmann Publishers.
6. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope further. In Kendall Clark and Peter F. Patel-Schneider, editors, *In Proceedings of the OWLED 2008 DC Workshop on OWL: Experiences and Directions*, 2008.
7. Y. Kazakov. An extension of regularity conditions for complex role inclusion axioms. In *Proceedings of the 2009 International Workshop on Description Logics (DL'09)*, 2009.

[3] http://lat.inf.tu-dresden.de/systems/cel/