Representation of Part-Whole Relationships in SNOMED CT
A. Patrice Seyed1, Alan Rector2, Uli Sattler2, Bijan Parsia2, and Robert Stevens2
1 Department of Computer Science and Engineering, University at Buffalo, USA
2 Department of Computer Science, University of Manchester, UK

ABSTRACT
In this paper we investigate representation of the part-whole relationship in SNOMED CT. We discuss the current approach, based on “SEP” triples, and several translations of it, which involve DLs at different levels of expressivity. We intend that our analysis will concretely inform the SNOMED community about the important tradeoffs of expressivity for their ontology, and help with future decisions about the representation of the SNOMED CT’s anatomical taxonomy.

1 INTRODUCTION
A common pattern in knowledge representation is that a fault of a part is considered a fault of the whole. For example, a fault in the battery is a fault in the ignition system, and a fault in the car. This pattern pervades common medical terminology: “Heart disease” includes diseases of any of the parts of the heart - muscle, valves, walls, etc. Gastrointestinal disease includes any disease of the stomach (gastrum) or any of the parts of the intestine. The same is true of procedures: fixing a heart valve is a kind of heart operation; repair of the retina is a kind of eye operation, etc.

However, the pattern does not always hold. “Amputation of the hand” means amputation of the entire hand. “Amputation of a finger” is not a kind of “Amputation of the hand” (although it is a kind of “Operation on hand”). Similarly, there are diseases that affect an entire organ, for example “pancarditis” means literally, “inflammation throughout (pan) the heart”.

In general, therefore, there is a requirement to represent two cases:
1. “Disorder/Procedure of A and/or any of its parts” and
2. “Disorder/Procedure of the entire A”
where A is any anatomical structure.

In common medical language, the distinction is usually implicit. The distinction between the meaning of “Operation on hand” and “Amputation of hand” is left to the medical knowledge of the reader. It is only in unusual cases such as “pancarditis” (“inflammation throughout the heart”) that the distinction is made explicit in the language. However, when representing diseases and procedures formally, the distinction must be made explicitly and systematically.

Over the past twenty years, there have been at least three mechanisms used to represent this pattern and the associated distinctions:
1. Propagation across transitive properties - the property used for “of”, usually “has_locus”, is said to be inherited across the property “part_of”. In modern description logics this is achieved by using property paths in subproperty axioms (Horrocks and Sattler, 2004). In earlier languages it was achieved by equivalent mechanisms known as “right identities” (Stearns et al., 2001)

2. Definition of diseases in terms of disjunctions - e.g., “Heart disease” is defined explicitly as “Disorder that has_locus some Heart”.

3. The use of Structure-Entity-Part (SEP) triples - separate classes for the whole or its parts (Structure), just the whole (Entity), or just the parts (Part). In this case “Heart disease” is defined as a “Disorder that has_locus some Heart Structure”.

Note that these three methods require different expressiveness in the description logic:

1. Propagation across transitive properties requires property-paths, which were not supported in early description logics and are not part of the basic specification of the standard starting description logic, ALC. They were originally thought to be intractable, but have since been shown not only to be tractable (Horrocks and Sattler, 2004) but to be even available in EL++, a maximal description logic with polynomial complexity (Baader et al., 2005).

2. Definition of diseases in terms of disjunctions requires a disjunction operator, which falls within ALC but outside EL++.

It also requires transitive properties but not property paths.

3. SEP triples can be implemented within the simplest possible description logic, and does not require transitive properties, disjunction or properties paths (Hahn et al., 1999).

The history of the use of these three methods and their variants is intertwined with the development of description logics for use with medical terminologies. The large description logic based terminology, SNOMED CT (Steams et al., 2001) was originally developed using a variant of propagation along transitive properties (Method 1) as was GALEN, the other large description logic based terminology developed in the mid 1990s (Rector et al., 1997), (Rogers and Rector, 2000). SNOMED converted to Method 3, and is now being re-examined in the light of experience, one format being considered being a variant of Method 2 (Personal communication, Kent Spackman, 2011). Re-examination of these approaches is therefore particularly timely.

The purpose of this paper is to explore variants on the three methods in the light of modern description logics. Although we comment briefly on the apparent cognitive complexity for the user of the different representations, any of the three techniques might be “hidden” from users by syntactic and user interface mechanisms. Our primary concern has been, therefore, with their formal, rather than cognitive aspects.

*To whom correspondence should be addressed: apseyed@buffalo.edu
2 THE CURRENT APPROACH (SEP TRIPLES)
We view SNOMED’s set of class names \( C \) to be partitioned into:
\[
C_S \cup C_E \cup C_P
\]
where \( C_S \cup C_E \cup C_P \) are specific to (human) anatomy. We use \( X_S \) for class names in \( C_S \), \( X_E \) for class names in \( C_E \), and \( X_P \) for class names in \( C_P \). We assume that in any occurrence of \( X_S, X_E, \) or \( X_P \) in an axiom, ‘X’ refers to the same term, e.g., Heart.

The SEP “triple” approach represents parthood implicitly within a class hierarchy (Hahn et al., 1999). For an anatomical entity of a certain kind, \( X_S \) represents its Structure class, and refers to any part of the anatomical entity, including the entire entity. For instance, Hearts refers to any part of a heart or an entire heart. \( X_E \) represents its Entire class, and refers to an entire anatomical entity, and \( X_P \) represents its Part class, and refers to a certain part of an entity. For instance, Heart refers to an entire heart, and Heart refers to any part of a heart but not an entire heart. \( X_E \) and \( X_P \) classes are immediate subclasses of \( X_S \); hence, \( Heart_E \) and \( Heart_P \) are immediate subclasses of \( Heart_S \). In the OWL version of the SNOMED CT ontology, the SEP notation is part of the class label, for example ‘Heart Structure’, ‘Entire Heart’, and ‘Part of Heart’, but in this paper we apply subscripts for notational convenience.

Ideally, a SEP triple is given for each anatomical entity, and every \( X_S \) class (except that for the top anatomical class) is a subclass of some \( Y_P \) class.\(^2\)

\(^2\) In SNOMED-CT, however, the SEP triples are thus far incompletely populated.
\(^3\) http://texasheart.org/HIC/Topics/Cond/myocard.cfm

whole heart is a part of some body, and furthermore, a specific part of a myocardium or a whole myocardium is a part of some body. These axioms are also illustrated in Figure 2??, and given formally below:

\[
\text{Myocardium} \sqsubseteq \text{Myocardium}_S \sqsubseteq \text{Heart}_P \sqsubseteq \text{Heart}_S \sqsubseteq \text{Body}_P \sqsubseteq \text{Body}_S
\]

\( Heart_E \sqsubseteq \text{Heart}_S \sqsubseteq \text{Body}_P \sqsubseteq \text{Body}_S \)

Note that, in SNOMED-CT, we neither find disjointness axioms for classes \( X_E \) and \( X_P \) nor covering axioms for \( X_S \), \( X_E \), and \( X_P \), although both are assumed to be true under the SEP triple theory.

The SEP triples approach is iteratively applied along what is considered a partonomic hierarchy, for example for the anterior myocardium under the SEP triple for myocardium. The subsumption relationships are explicit, as given, but their reading is implicit; in particular, there is no ‘part of’ property that link \( X_E \) and \( X_P \). However, transitivity of the subsumption relation implies the transitivity of this implicit part of reading, and so transitive parthood entailments are determined by subsumption reasoning. We refer to the SEP triple approach from SNOMED-CT described so far and sketched in Figure 2 as the Current SEP Triple Approach (A). In the following sections we discuss several alternative approaches to representing part-whole relations and discuss their relative expressivity.

On how approach A applies to subsumption reasoning for disorders, take for example a disorder specified in some anatomical location that is given as some class \( X_S \), Carditis is inflammation that is located in some specific part of a heart, or a whole heart, therefore

---

2 The Current Approach (SEP Triples)

We view SNOMED’s set of class names \( C \) to be partitioned into:
\[
C_S \cup C_E \cup C_P
\]
where \( C_S \cup C_E \cup C_P \) are specific to (human) anatomy. We use \( X_S \) for class names in \( C_S \), \( X_E \) for class names in \( C_E \), and \( X_P \) for class names in \( C_P \). We assume that in any occurrence of \( X_S, X_E, \) or \( X_P \) in an axiom, ‘X’ refers to the same term, e.g., Heart.

The SEP “triple” approach represents parthood implicitly within a class hierarchy (Hahn et al., 1999). For an anatomical entity of a certain kind, \( X_S \) represents its Structure class, and refers to any part of the anatomical entity, including the entire entity. For instance, Hearts refers to any part of a heart or an entire heart. \( X_E \) represents its Entire class, and refers to an entire anatomical entity, and \( X_P \) represents its Part class, and refers to a certain part of an entity. For instance, Heart refers to an entire heart, and Heart refers to any part of a heart but not an entire heart. \( X_E \) and \( X_P \) classes are immediate subclasses of \( X_S \); hence, \( Heart_E \) and \( Heart_P \) are immediate subclasses of \( Heart_S \). In the OWL version of the SNOMED CT ontology, the SEP notation is part of the class label, for example ‘Heart Structure’, ‘Entire Heart’, and ‘Part of Heart’, but in this paper we apply subscripts for notational convenience.

Ideally, a SEP triple is given for each anatomical entity, and every \( X_S \) class (except that for the top anatomical class) is a subclass of some \( Y_P \) class.\(^2\)

\(^2\) In SNOMED-CT, however, the SEP triples are thus far incompletely populated.
\(^3\) http://texasheart.org/HIC/Topics/Cond/myocard.cfm

whole heart is a part of some body, and furthermore, a specific part of a myocardium or a whole myocardium is a part of some body. These axioms are also illustrated in Figure 2??, and given formally below:

\[
\text{Myocardium} \sqsubseteq \text{Myocardium}_S \sqsubseteq \text{Heart}_P \sqsubseteq \text{Heart}_S \sqsubseteq \text{Body}_P \sqsubseteq \text{Body}_S
\]

\( Heart_E \sqsubseteq \text{Heart}_S \sqsubseteq \text{Body}_P \sqsubseteq \text{Body}_S \)

Note that, in SNOMED-CT, we neither find disjointness axioms for classes \( X_E \) and \( X_P \) nor covering axioms for \( X_S \), \( X_E \), and \( X_P \), although both are assumed to be true under the SEP triple theory.

The SEP triples approach is iteratively applied along what is considered a partonomic hierarchy, for example for the anterior myocardium under the SEP triple for myocardium. The subsumption relationships are explicit, as given, but their reading is implicit; in particular, there is no ‘part of’ property that link \( X_E \) and \( X_P \). However, transitivity of the subsumption relation implies the transitivity of this implicit part of reading, and so transitive parthood entailments are determined by subsumption reasoning. We refer to the SEP triple approach from SNOMED-CT described so far and sketched in Figure 2 as the Current SEP Triple Approach (A). In the following sections we discuss several alternative approaches to representing part-whole relations and discuss their relative expressivity.

On how approach A applies to subsumption reasoning for disorders, take for example a disorder specified in some anatomical location that is given as some class \( X_S \), Carditis is inflammation that is located in some specific part of a heart, or a whole heart, therefore
These axioms and entailments are illustrated in Figure 3.\(^4\)

In SNOMED CT, there are numerous disorders defined in terms of their location. For instance, Myocarditis is inflammation that is located in some specific part of a myocardium or a whole myocardium, therefore, Myocardium\(_S\). As illustrated in Figure 3, because Myocardium\(_S\) is a subclass of Hearts, the location for Myocarditis is also Hearts, and further, Myocarditis is a subclass of Carditis. We provide the DL representation for these findings and the corresponding inferences:

\[
\text{Carditis} \equiv \text{Inflammation} \sqcap \exists \text{has locus}.\text{Hearts}
\]

\[
\text{Myocarditis} \equiv \text{Inflammation} \sqcap \exists \text{has locus}.\text{Myocardium}\_S
\]

\[\vdash \text{Myocarditis} \sqsubseteq \text{Inflammation} \sqcap \exists \text{has locus}.\text{Hearts}\_S\]

\[\vdash \text{Myocarditis} \sqsubseteq \text{Carditis}\]

A disorder that occurs at some location that is specified as a class \(X_E\), however, does not have such inferred subclasses. For example, Pancarditis is a disorder that is characterized by inflammation and is specified as being located in the entire heart and not just some part of the heart, therefore Heart\(_E\). Recall that Myocarditis is located in some specific part of the myocardium or the entire myocardium, therefore Myocardium\(_S\). As illustrated in Figure 4, it is accurately not entailed that Myocarditis is a subclass of Pancarditis:

\[
\text{Pancarditis} \equiv \text{Inflammation} \sqcap \exists \text{has locus}.\text{Heart}\_P
\]

\[\not\vdash \text{Myocarditis} \sqsubseteq \text{Pancarditis}\]

\[\text{Pancarditis} \equiv \text{Inflammation} \sqcap \exists \text{has locus}.\text{Heart}\_E\]

\[\text{Myocarditis} \equiv \text{Inflammation} \sqcap \exists \text{has locus}.\text{Myocarditis}\_S\]

\[\not\equiv \text{Myocarditis} \sqsubseteq \text{Pancarditis}\]

3 ALTERNATIVE APPROACHES FOR REPRESENTING PART-WHOLE RELATIONSHIPS

We discuss five alternative approaches for representing part-whole relationships in SNOMED CT, the first of which is a reformulation of approach A.

3.1 Alternative Approach 1

We define Alternative Approach 1 (A1) such that \(X_S\) and \(X_P\) are fully defined based on \(X_E\) by introducing a transitive and reflexive part\(_P\) of property, as described by Seidenberg and Rector (2006). SNOMED is the set-theoretic difference of the original anatomy-specific SNOMED CT axioms from all SNOMED CT axioms. We define A1 as follows:

\[
\text{SNOMED} \cup \{X_S \equiv X_E \sqcup \exists \text{part of}_E. X_E \mid X_E \in C_S, X_E \in C_E\} \cup \{X_P \equiv \exists \text{part of}_E. X_E \mid X_E \in C_P\}
\]

\(\text{Hearts}\) and \(\text{Heart}_P\) are therefore defined as follows:

\[
\text{Hearts} \equiv \text{Heart}_E \sqcup \exists \text{part of}_E. \text{Heart}_E
\]

\[
\text{Heart}_P \equiv \exists \text{part of}_E. \text{Heart}_E
\]

Myocardium\(_S\) and Myocardium\(_P\) are also defined in this manner, and the following axiom connects the two triples:

\[
\text{Myocardium}_S \sqsubseteq \text{Heart}_P
\]

Therefore Myocardium\(_S\) and Myocardium\(_P\) are subclasses of the expression \(\exists \text{part of}_E. \text{Hearts}\). Because Myocarditis is an inflammation located in Myocardium\(_S\), and by inference Hearts\(_S\), it appropriately follows that Myocarditis is a subclass of Carditis.

---

\(^4\) When there is any question, SNOMED CT uses the Structure class.

\(^5\) Inferred relationships are given as dotted arcs.
3.2 Alternative Approach 2

Alternative Approach 2 (A2) is based on modifications to A1 which is obtained by the following steps:

1. Remove all axioms of the form \( X_r \subseteq X_s \) and \( X_r \subseteq X_s \) (where \( X \) and \( Y \) are different) with \( X \subseteq \exists \text{ has part of } Y \).
2. Replace all connecting axioms of the form \( X_r \subseteq Y_r \) (where \( X \) and \( Y \) are different) with \( X \subseteq \exists \text{ proper part of } Y \).
3. Replace every occurrence of \( X_s \) of a class name in \( C_s \) with \( X \subseteq \exists \text{ proper part of } X \) and every occurrence of \( X_s \) of a class name in \( C_s \) with \( X \).

Applying step (2) in A2, the connecting axiom for our running example classes is:

Myocardium \( \sqsubseteq \exists \text{ part of } \text{ Heart} \)

Applying step (3) the example disorders are defined as:

\[ \text{Carditis} \equiv \text{Inflammation} \sqcap \exists \text{ has locus (~Heart \sqcup \exists \text{ part of } \text{ Heart})} \]

Myocarditis \( \equiv \text{Inflammation} \sqcap \exists \text{ has locus (~Myocardium \sqcup \exists \text{ part of } \text{ Myocardium})} \)

And by applying (3) to an inflammation disorder that is located in the entire heart, we apply the X class, Heart:

\[ \text{Pancarditis} \equiv \text{Inflammation} \sqcap \exists \text{ has locus } \text{ Heart} \]

By the connecting axiom, every myocardium is a part of some heart, and because part is transitive, every part of some myocardium is a part of some heart. Because Myocarditis is an inflammation of the myocardium or some part, both of which are parts of the heart, as in the prior two approaches, Myocarditis is a subclass of Carditis.

3.3 Alternative Approach 3

As a variant of this replacement approach, A2, Alternative Approach 3 (A3) introduces a property \( \exists \text{ proper part of } \), which is a subproperty of part, is transitive and assumed irreflexive. As in A2, part remains transitive and reflexive. A3 is obtained by the following steps:

1. Remove all axioms of the form \( X_E \subseteq X_S \) and \( X_E \subseteq X_S \) (where \( X \) and \( Y \) are different) with \( X \subseteq \exists \text{ proper part of } Y \).
2. Replace all connecting axioms of the form \( X_S \subseteq Y_r \) (where \( X \) and \( Y \) are different) with \( X \subseteq \exists \text{ proper part of } Y \).
3. Replace every occurrence of \( X_S \) of a class name in \( C_S \) with \( X \subseteq \exists \text{ proper part of } X \) and every occurrence of \( X_E \) of a class name in \( C_E \) with \( X \).
4. Add \( \exists \text{ proper part of } \subseteq \text{ part of } \).

A3 differs from A2 in that for (2) \( \exists \text{ proper part of } \) replaces part in the connecting axiom, and also (4) is an additional step in A1.

These steps are required in this approach because parthood between anatomical entities is defined with the \( \exists \text{ proper part of } \) property.

Applying (2) the connecting axiom for Myocardium and Heart is then:

Myocardium \( \sqsubseteq \exists \text{ proper part of } \text{ Heart} \)

The application of (3) results in analogous definitions for Carditis, Myocarditis, and Pancarditis as are given for A2. By the connecting axiom, along with (4) and the transitivity of part, as was the case for A1, A2, and A2, Myocarditis is appropriately an inferred subclass of Carditis.

3.4 Alternative Approach 4

Alternative Approach 4 (A4) applies the \( \exists \text{ proper part of } \) property, repeats Step (1) and (2) from A1, and includes the following steps:

3. Replace every occurrence of \( X_S \) of a class name in \( C_S \) with \( \exists \text{ part of } X \) and every occurrence of \( X_E \) of a class name in \( C_E \) with \( X \).
4. Add \( \exists \text{ proper part of } \subseteq \text{ part of } \).
5. Add \( \exists \text{ part of } \circ \exists \text{ proper part of } \subseteq \exists \text{ proper part of } \).

A4 differs from A3 in two respects. First, for (3) \( \exists \text{ part of } \) replaces \( X \) \( \sqcup \exists \text{ part of } X \); second, step (5) introduces a left identity axiom which is necessary because it allows us to infer:

\[ \models \exists \text{ part of } \text{ Myocardium} \subseteq \exists \text{ proper part of } \text{ Heart} \]

and subsequently:

\[ \models \text{ Myocarditis} \subseteq \exists \text{ has locus } \exists \text{ proper part of } \text{ Heart} \]

Applying (2) the connecting axiom for Myocardium and Heart is:

Myocardium \( \subseteq \exists \text{ proper part of } \text{ Heart} \)

But, different from A2 and A3, applying (3) for our example disorders results in:

Carditis \( \equiv \text{Inflammation} \sqcap \exists \text{ has locus } \exists \text{ part of } \text{ Heart} \)

Myocarditis \( \equiv \text{Inflammation} \sqcap \exists \text{ has locus } \exists \text{ part of } \text{ Myocardium} \)

The definition for Pancarditis remains the same as A2 and A3.

By the connecting axiom, along with (4) and the transitivity of part, as was the case for A, A1, A2, and A3, Myocarditis is an inferred subclass of Carditis. A major issue with this approach is that (5) in connection with (4) leads to cycles, which is not allowed in the DL language that underlies OWL 2 (Baader et al., 2009).

3.5 Alternative Approach 5

Alternative Approach 5 (A5) introduces the \( \exists \text{ has locus entire } \) property, a subproperty of has locus, which expresses when a finding is located in some Xe class. A5 repeats Step (1) and (2) from A1 also, and includes the following steps:

3. Replace every occurrence of \( X_S \) of a class name in \( C_S \) with \( \exists \text{ part of } X \) and every occurrence of \( \exists \text{ has locus } X_E \) with \( \exists \text{ has locus entire } X \).
4. Add \( \exists \text{ proper part of } \subseteq \text{ part of } \).
5. Add \( \exists \text{ has locus } \circ \text{ part of } \subseteq \exists \text{ has locus } \).

A5 differs from A4 in two respects. First, in (3) A5 employs the \( \exists \text{ has locus entire } \) property. Second, for A5 in (5) a right identity axiom is applied, where the has locus property is “transitive over” the part of relation.

Applying (2) the connecting axiom for Myocardium and Heart is the same as for A4. Different from all other alternative approaches, applying (3) for our example disorders results in:

\[ \text{Carditis} \equiv \text{Inflammation} \sqcap \exists \text{ has locus } \exists \text{ part of } \text{ Heart} \]

\[ \text{Myocarditis} \equiv \text{Inflammation} \sqcap \exists \text{ has locus } \exists \text{ part of } \text{ Myocardium} \]

\[ \text{Pancarditis} \equiv \text{Inflammation} \sqcap \exists \text{ has locus } \exists \text{ part of } \text{ Heart} \]

6 Composite properties cannot be defined as irreflexive in OWL.

7 A left identity axiom can be formalized in OWL2 as a property chain axiom.
Carditis ≡ Inflammation ⊓∃ has_locus Heart

Myocarditis ≡ Inflammation ⊓∃ has_locus Myocardium

Also applying (3) to an inflammation disorder that is located in the entire heart yields:

Pancarditis ≡ Inflammation ⊓∃ has_locus entire Heart

By the connecting axiom, along with (4) and (5), the same inferences hold for our example disorders, primarily that Myocarditis is a subclass of Carditis.

4 DISCUSSION

In Section 1 we introduced three major methods for representing part-whole relationships, by applying: (1) transitive properties (2) disjunctions and (3) SEP triples. In Section 2 we introduced the logic underlying the current approach in SNOMED CT, and in Section 3 the logic underlying five alternative approaches. The approach used in SNOMED CT currently, A, applies (3), which is within ALC expressivity. A1 applies both (2) and (3), and A2 and A1 correspond to just (2), because SEP triple classes are removed for both. Due to disjunctions the first three alternative approaches are beyond EL++ expressivity. A4 and A5 apply (1) only, therefore fall within EL++.

5 CONCLUSION

A major difference between the current approach, A, and the alternative approaches, A1 - A5, is that the former offers only a propositional representation and the latter offer a relational representation of part-whole. A does not model partonomic structure, but rather partonomic “level”.

It is reported by users of SNOMED-specific browsers that SEP triples are cumbersome to browse and search through. We suggest that this problem can be addressed by providing more intuitive labels. In the context of user navigation, it is simply a rendering issue. It is for this reason we do not necessarily recommend one of A1 - A5 in its place. Nevertheless, A1 - A5 do provide the benefit of allowing a user to explicitly query parts, for A queries require knowledge of the SEP class hierarchy.

In preliminary performance testing, A1 performed the worst for classification across all the DL reasoners we tested. This is no doubt attributable to the inclusion of disjuncts in the class definitions, and corresponding unfolding performed by the reasoner. Despite this, A1 has utility as a representation used for mapping between ontologies that use the propositional approach and those that use the relational approach. Clearly, formulations that include the part_of property facilitate ontology modularity, merging, and enrichment where A1 can serve as a bridge.

In future work we will empirically measure classification and query performance for these different SNOMED ontology formulations approaches across several DL reasoners. Furthermore, we will apply an evaluation framework across the formulations for various types of information requests. In that work we will address what kinds of information requests are expressible as OWL class expressions, and which require a more expressive query language.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation, NSF Grant IIS-1107011.

REFERENCES


