Representation of Part-Whole Relationships in SNOMED CT

A. Patrice Seyed [1]*, Alan Rector [2], Uli Sattler [2], Bijan Parsia [2], and Robert Stevens [2]

[1]Department of Computer Science and Engineering, University at Buffalo, USA
[2]School 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 is 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) or "refined_by" (Rogers and Rector, 2000). This amounts to an axiom that the disorder of the part is a disorder of the whole. In this case a mechanism must be provided to cope with the exceptions when the rule does not apply. For example, in this case "Heart disease" is defined simply as "Disorder that has_locus some Heart".
2. Explicit definition of diseases as disjunctions - e.g., "Heart disease" is defined explicitly as "Disease that has_locus some Heart OR some part_of Heart".
3. The use of Structure-Entity-Part (SEP) triples - separate classes for the whole or its parts (**S**tructure), just the whole (**E**ntity), or just the parts (**P**art). 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 (Stearns *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 1 (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, which has also been investigated in (Baader *et al.*, 2009). 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.

---

*To whom correspondence should be addressed: apseyed@buffalo.edu

Our primary concern has been, therefore, with their formal, rather than cognitive aspects.

2 THE CURRENT APPROACH (SEP TRIPLES)

We view SNOMED's set of class names C to be partitioned into:

$$C_n \cup C_S \cup C_E \cup C_P$$

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, $Heart_S$ 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, $Heart_E$ refers to an entire heart, and $Heart_P$ 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,[1] 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]

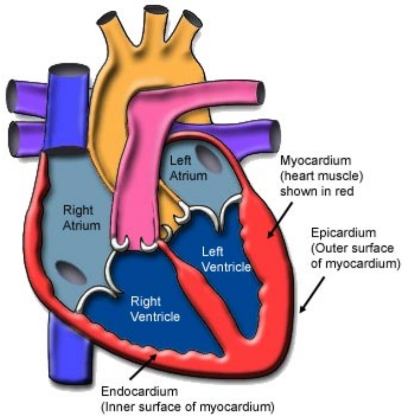

Fig. 1: Illustration of the Human Heart

The heart has as part of it a muscular wall that contracts to pump blood out of the heart, and then relaxes as the heart refills with returning blood. This wall is called the *myocardium*. The heart and myocardium are illustrated in **Figure 1**.[3] Applying SEP triples, $Myocardium_S$ is a subclass of $Heart_P$ and $Heart_S$ is a subclass of $Body_P$. This means that a specific part of a myocardium or a whole myocardium is a part of some heart, a specific part of a heart or a 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:

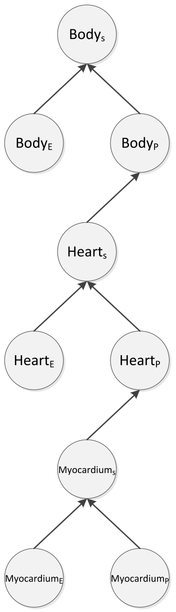

Fig. 2: Taxonomy of SEP Triple classes for Heart, Myocardium, and Body. Unlabeled arcs represent the subclass relationship.

$$Myocardium_E \sqsubseteq Myocardium_S \sqsubseteq Heart_P \sqsubseteq Heart_S \ldots \sqsubseteq Body_P \sqsubseteq Body_S$$

$$Heart_E \sqsubseteq Heart_S \ldots \sqsubseteq Body_P \sqsubseteq Body_S$$

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 links 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 an inflammation that is located in some specific part of a heart, or a whole heart,

---

[1] http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html.

[2] In SNOMED CT, however, the SEP triples are thus far incompletely populated.

[3] http://texasheart.org/HIC/Topics/Cond/myocard.cfm

therefore *Heart*$_S$.[4] These axioms and entailments are illustrated in **Figure 3**.[5]

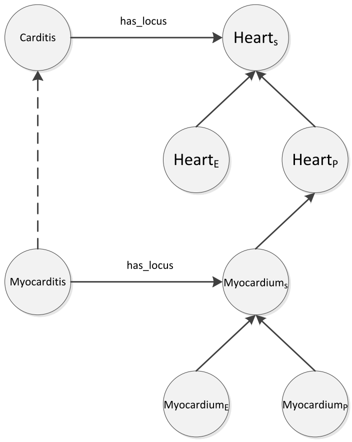

Fig. 3: Entailment given the Part-Whole Relationship. In the OWL representation class definition for *Carditis*, *Inflammation* is the range restriction for the property *Associated morphology*. We exclude this expression from the definition of *Carditis* above in order to simplify our examples.

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 *Heart*$_S$, the location for *Myocarditis* is also *Heart*$_S$, and further, *Myocarditis* is a subclass of *Carditis*. We provide the DL representation for these findings and the corresponding inferences:

$$\mathit{Carditis} \equiv \mathit{Inflammation} \sqcap \exists \mathit{has\_locus}.\mathit{Heart}_S$$

$$\mathit{Myocarditis} \equiv \mathit{Inflammation} \sqcap \exists \mathit{has\_locus}.\mathit{Myocardium}_S$$

$$\models \mathit{Myocarditis} \sqsubseteq \mathit{Inflammation} \sqcap \exists \mathit{has\_locus}.\mathit{Heart}_S$$

$$\models \mathit{Myocarditis} \sqsubseteq \mathit{Carditis}$$

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*:

$$\mathit{Pancarditis} \equiv \mathit{Inflammation} \sqcap \exists \mathit{has\_locus}.\mathit{Heart}_E$$

$$\mathit{Myocarditis} \equiv \mathit{Inflammation} \sqcap \exists \mathit{has\_locus}.\mathit{Myocarditis}_S$$

$$\not\models \mathit{Myocarditis} \sqsubseteq \mathit{Pancarditis}$$

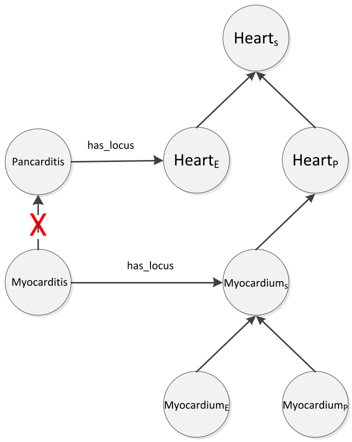

Fig. 4: No Entailment given the Part-Whole Relationship

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** (A_1) such that X_S and X_P are fully defined based on X_E by introducing a transitive *part_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 A_1 as follows:

$$\begin{aligned}&\mathit{SNOMED} \cup \\ &\quad \{X_S \equiv X_E \sqcup \exists \mathit{part\_of}.X_E \mid X_S \in C_S, X_E \in C_E\} \cup \\ &\quad \{X_P \equiv \exists \mathit{part\_of}.X_E \mid X_P \in C_P\}\end{aligned}$$

Heart$_S$ and *Heart*$_P$ are therefore defined as follows:

$$\mathit{Heart}_S \equiv \mathit{Heart}_E \sqcup \exists \mathit{part\_of}.\mathit{Heart}_E$$

$$\mathit{Heart}_P \equiv \exists \mathit{part\_of}.\mathit{Heart}_E$$

Myocardium$_S$ and *Myocardium*$_P$ are also defined in this manner, and the following axiom connects the two triples:

$$\mathit{Myocardium}_S \sqsubseteq \mathit{Heart}_P$$

Therefore *Myocardium*$_E$ and *Myocardium*$_P$ are subclasses of the expression $\exists \mathit{part\_of}.\mathit{Heart}_E$. Because *Myocarditis* is an inflammation located in *Myocardium*$_S$, and by inference *Heart*$_S$, it appropriately follows that *Myocarditis* is a subclass of *Carditis*.

3.2 Alternative Approach 2

Alternative Approach 2 (A_2) is based on modifications to A_1 which is obtained by the following steps:

1.Remove all axioms of the form $X_E \sqsubseteq X_S$ and $X_P \sqsubseteq X_{S.}$

---

[4] When there is any question, SNOMED CT uses the Structure class.

[5] Inferred relationships are given as dotted arcs.

2. Replace all connecting axioms of the form $X_S \sqsubseteq Y_P$ (where X and Y are different) with $X \sqsubseteq \exists part\_of.Y$.
3. Replace every occurrence of X_S of a class name in C_S with $X \sqcup \exists part\_of.X$ and every occurrence of X_E of a class name in C_E with X.

Applying step (2) in A_2, the connecting axiom for our running example classes is:

$$Myocardium \sqsubseteq \exists part\_of.Heart$$

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

$$Carditis \equiv Inflammation \sqcap \exists has\_locus.(Heart \sqcup \exists part\_of.Heart)$$

$$Myocarditis \equiv Inflammation \sqcap \exists has\_locus.(Myocardium \sqcup \exists part\_of.Myocardium)$$

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

$$Pancarditis \equiv Inflammation \sqcap \exists has\_locus.Heart$$

By the connecting axiom, every myocardium is a part of some heart, and because *part_of* 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

Alternative Approach 3 (A_3) repeats Step (1) from A_2, applies the *proper_part_of* property as a subproperty of *part_of*, and includes the following steps for the connecting axiom and treatment of class names in C_S and C_E:

2. Replace all connecting axioms of the form $X_S \sqsubseteq Y_P$ (where X and Y are different) with $X \sqsubseteq \exists proper\_part\_of.Y$.
3. Replace every occurrence of X_S of a class name in C_S with $\exists part\_of.X$, and every occurrence of X_E of a class name in C_E with X.

Additionally, for inferences of parthood:

4. Add $proper\_part\_of \sqsubseteq part\_of$.
5. Add $part\_of \circ proper\_part\_of \sqsubseteq proper\_part\_of$.

A_3 differs from A_2 in three important respects. First, for (3) $part\_of.X$ replaces $X \sqcup part\_of.X$; second, *part_of* here is defined as reflexive, where it is assumed irreflexive in A_2 (and A_1); and third, Step (5) introduces a left identity axiom which is necessary because it allows us to infer:[6]

$$\models \exists part\_of.Myocardium \sqsubseteq \exists proper\_part\_of.Heart$$

and subsequently:

$$\models Myocarditis \sqsubseteq \exists has\_locus.\exists proper\_part\_of.Heart$$

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

$$Myocardium \sqsubseteq \exists proper\_part\_of.Heart$$

But, different from A_2, applying (3) for our example disorders results in:

$$Carditis \equiv Inflammation \sqcap \exists has\_locus.\exists part\_of.Heart$$

$$Myocarditis \equiv Inflammation \sqcap \exists has\_locus.\exists part\_of.Myocardium$$

The definition for *Pancarditis* remains the same as A_2.

By the connecting axiom, along with (4) and the transitivity of *part_of*, as was the case for A, A_1, and A_2, *Myocarditis* is an inferred subclass of *Carditis*. Note that by this approach, that (5) in connection with (4) leads to cycles (as described in (Baader *et al.*, 2009)), which is not allowed in the DL language that underlies OWL 2. Fortunately this does not pose any problems for those reasoners implemented for EL++ expressivity.

3.4 Alternative Approach 4

Alternative Approach 4 (A_4) introduces the *has_locus_entire* property, a subproperty of *has_locus*, which expresses when a finding is located in some X_E class. This approach was first introduced in (Baader *et al.*, 2009)). A_4 repeats Step (1) from A_2, as A_3 did, and repeats Step (2), from A_3, while including the following step for the treatment of class names in C_S and C_E:[7]

3. Replace every occurrence of X_S of a class name in C_S with X and every occurrence of $\exists has\_locus.X_E$ of a class name in C_E with $\exists has\_locus\_entire.X$.

A_4 also repeats (4) and (5) from A_3, while including an additional step:

6. Add $has\_locus \circ part\_of \sqsubseteq has\_locus$.

A_4 differs from A_3 in two respects. First, in (3) A_4 treats X—instead of $\exists part\_of.X$—as a replacement for X_S, and employs the *has_locus_entire* property. Second, for A_4 in (6) 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 A_3. Different from all other alternative approaches, applying (3) for our example disorders results in:

$$Carditis \equiv Inflammation \sqcap \exists has\_locus.Heart$$

$$Myocarditis \equiv Inflammation \sqcap \exists has\_locus.Myocardium$$

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

$$Pancarditis \equiv Inflammation \sqcap \exists has\_locus\_entire.Heart$$

which prevents erroneous propagation via the right identity axiom. 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)

---

[6] A left identity axiom can be formalized in OWL2 as a property chain axiom.

[7] Baader *et al.* (2009) also keep Structure and Part expressions fully defined as $X_S \equiv \exists part\_of.X$ and $X_P \equiv \exists propert\_part\_of.X$, for legacy reasons.

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 four alternative approaches. The approach used in SNOMED CT currently, A, is an application of (3), which is within ALC expressivity. A_1 is an application of both (2) and (3), while A_2 is an application of just (2); both are within ALC but are outside EL++ due to disjunctions. A_3 and A_4 are an application of just (1), and fall within EL++.

In general, there is a modeling choice between treating a generalized 'part of' property as reflexive or irreflexive. In A_1 and A_2 the *part_of* property corresponds to the latter choice, and is assumed irreflexive. It is only assumed because in OWL2 we cannot assert that a transitive property is irreflexive, but we can assert that a transitive property is reflexive. Therefore we can also introduce approaches (as shown for A_3 and A_4) which correspond to the former choice, where 'part of' is reflexive, which can be therefore be applied—directly and without disjunctions—for representing the X_S class expression. In these approaches a subproperty *proper_of*, again assumed irreflexive, is also introduced for representing the X_P class expression; subsequently cyclic role chains are required in order for the respective ontologies to entail correct subclasses of the pattern $\exists$*proper_part.X*.

Also, an important distinction between the approaches A_3 and A_4 is that while A_4 has the same approach as A_3 for translating and thus representing SEP class expressions (via patterns $\exists$*part_of.X* and $\exists$*propert_part_of.X* for Structure and Part expressions, respectively), A_4 has a different approach for inheritance of properties along a partonomy. For A_4 the inheritance is through a right identity axiom, while for A_3 it is through the transitivity of *part_of*.

5 CONCLUSION

A major difference between the current approach, A, and the alternative approaches, A_1 - A_4, is that the former offers only a propositional representation and the latter offer a relational representation of parthood. A does not model partonomic structure, but rather partonomic "level". By modeling partonomic structure explicitly via the *part_of* property we can make explicit statements of how *part_of* interacts with other properties (i.e., laterality):

$\exists\ hasLat.Left \sqsubseteq (\forall\ part\_of.(\exists\ hasLat.\top \Rightarrow \exists\ hasLat.Left))$

says that, if something has a left laterality, then, what it is a part of, if this 'whole' has a laterality at all, it has a left laterality. Modelling this kind of interaction requires an explicit part of - which then can, of course, be used in sub-role and inverse role axioms as well.

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 against the A or A_1 approach. Nevertheless, A_1 - A_4 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, A_1 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, A_1 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 A_1 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) in conjunction with IJCAI 2011. We would like to give thanks to Luigi Iannone for assistance in using the OPPL scripting toolkit and useful advice for using the OWLAPI for the translation work. We would also like to give thanks to Kent Spackman and the reviewers for their helpful feedback.

REFERENCES

Baader, F., Brandt, S., and Lutz, C. (2005). Pushing the el envelope. In *International Joint Conference on Artificial Intelligence*, volume 19, page 364. Citeseer.

Baader, F., Schulz, S., Spackman, K., and Suntisrivaraporn, B. (2009). How Should Parthood Relations be Expressed in SNOMED CT? *Proceedings of the First Workshop des GI-Arbeitskreises Ontologien in Biomedizin und Lebenswissenschaften*.

Hahn, U., Schulz, S., and Romacker, M. (1999). Partonomic reasoning as taxonomic reasoning in medicine. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference*, pages 271–276. American Association for Artificial Intelligence.

Horrocks, I. and Sattler, U. (2004). Decidability of SHIQ with complex role inclusion axioms. *Artificial Intelligence*, **160**(1-2), 79–104.

Rector, A., Bechhofer, S., Goble, C., Horrocks, I., Nowlan, W., and Solomon, W. (1997). The GRAIL concept modelling language for medical terminology. *Artificial intelligence in medicine*, **9**(2), 139–171.

Rogers, J. and Rector, A. (2000). GALEN's model of parts and wholes: experience and comparisons. In *Proceedings of the AMIA symposium*, page 714. American Medical Informatics Association.

Seidenberg, J. and Rector, A. (2006). Representing transitive propagation in OWL. *Conceptual Modeling-ER 2006*, pages 255–266.

Stearns, M., Price, C., Spackman, K., and Wang, A. (2001). SNOMED clinical terms: overview of the development process and project status. In *Proceedings of the AMIA Symposium*, page 662. American Medical Informatics Association.