Applying description logics extended with meta-modelling to SNOMED-CT

Regina Motz¹, Edelweis Rohrer¹, and Paula Severi²

Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Uruguay {rmotz, erohrer}@fing.edu.uy
Department of Computer Science, University of Leicester, England pgs11@leicester.ac.uk

Abstract. SNOMED-CT is a clinical and medical ontology that covers a wide range of concepts in the health domain. It is mostly used as a standard vocabulary to be referenced in electronic health records of patients. The SNOMED-CT ontology has been formalized with a description logic, the Web Ontology language OWL 2 profile $\mathcal{OWL}\text{-}\mathcal{EL}$. To integrate electronic health records to SNOMED-CT in an $\mathcal{OWL}\text{-}\mathcal{EL}$ ontology, references to SNOMED-CT concepts in these records are modelled as instances of the referenced concepts. In this paper, the ontological model of this integration is analyzed. As our main contribution, we propose an approach to give solution to some problems of this integration, by defining a meta-model layer of SNOMED-CT with its concepts represented also as individuals, using an approach of description logics extended with meta-modelling.

1 Introduction

SNOMED-CT is a clinical and medical ontology that covers a wide range of concepts in the health domain, such as diseases, treatments and so on [10, 22]. It has been formalized with the Web Ontology language OWL 2 profile $\mathcal{OWL}\text{-}\mathcal{EL}$, which is a lightweight description logic with some modelling restrictions that improve reasoning efficiency and scalability [13, 23, 24]. The building blocks of all description logics are concepts, roles and individuals [2, 7]. The $\mathcal{OWL}\text{-}\mathcal{EL}$ SNOMED-CT ontology has concepts, such as Disease that subsumes HeartDisease and Endocarditis, and roles, such as findingSite that represents where diseases are located. However, it has not individuals

SNOMED-CT is mostly used as a standard vocabulary to be referenced in electronic health records of patients [4, 12, 18, 20]. In most case studies, an ontology of electronic health records of patients referencing SNOMED-CT terms is modelled with individuals, that represent such references and instantiate the SNOMED-CT concepts [3, 4, 18]. Figure 1 illustrates the electronic health records of two patients Juan and Pedro, who suffer endocarditis. Ovals represent concepts, bullets represent individuals and arrows represent roles. The individuals juanEHRendocarditis, juanEHRendocardium and juanEHRinflamation, instances of the concepts Endocarditis, Endocardium and Inflammation, are just the references to these concepts in the electronic health records

2 Motz, Rohrer, Severi

of Juan. Moreover, the individuals are linked through instances of roles, such as findingSite.

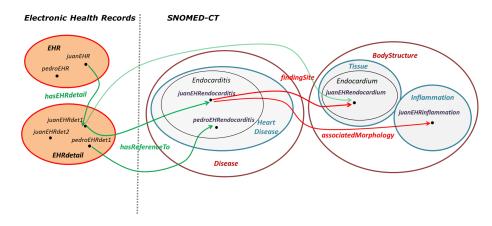


Fig. 1: Integration of Electronic Health Records with SNOMED-CT

In the present work, the ontological model of SNOMED-CT is analyzed. In particular, we study the integration of SNOMED-CT to electronic health records, as sketched in Figure 1. There exist several works which criticize the logical structure of SNOMED-CT and its integration to electronic health records, and propose some changes to the model of SNOMED-CT [19–21]. However, we argue that a more realistic approach is to keep SNOMED-CT unchanged and add an upper view by representing its concepts also as individuals. As a result, we conceptualize the health domain at a higher level of abstraction and introduce a different approach to solve some of the identified problems. To formalize our proposal, we need a logic which allows us to represent the same real object as an individual and as a concept, for example, the term Endocarditis as a concept (as in SNOMED-CT) and also as an individual. Standard description logics, and in particular $\mathcal{OWL-EL}$, are not expressive enough to model this scenario. Hence, we use our approach of description logics extended with meta-modelling [14–16], that allows to express that an individual corresponds to a concept through meta-modelling. We adjust this extension to allow that concept names to be treated also as individuals.

The main contribution of the present work is to enrich the logical model of the SNOMED-CT ontology with meta-modelling, to enhance its the integration with electronic health records. Our solution prevents from possible mistakes in the population of the SNOMED-CT concepts with references of electronic health records.

The remainder of this paper is organized as follows. Section 2 presents some related work about SNOMED-CT. Section 3 outlines the meta-modelling extension to description logics. In Section 4 we explain our proposal of integration of SNOMED-CT to electronic health records using the meta-modelling extension presented in Section 3. Finally, Section 5 presents some conclusions and future work.

2 Related work

In this section, we present some related work about SNOMED-CT.

Despite nowadays SNOMED-CT is broadly used in several health applications, the correctness of its logical structure is questioned in several works. Schulz et al. [21] describe several structural problems. Since SNOMED-CT concepts were thought as a node hierarchy, and not as ontological concepts, they say that nodes could be concepts, meta-concepts, individuals or roles. They found that the *role-group* relation hides roles such as *has-part*, and that the concept hierarchy subsumption is overloaded, instead of defining roles. They propose to represent SNOMED-CT using the description logic \mathcal{EL}^{++} , to add the *part-of* role, symmetric and reflexive roles, and domain and range axioms.

Schulz et al. [20] distinguish three ways to instantiate SNOMED-CT concepts: (i) with real instances in patients (*standard interpretation*) (ii) with references to SNOMED-CT concepts in electronic health records (*EHR interpretation*) or (iii) with patients affected by a disease (*epidemiological interpretation*). They consider as correct the standard interpretation, but the EHR interpretation is the more feasible to be implemented.

Rector et. al. [18] analyze a mechanism for using generic information models, such us HL7 RIM or OpenEHR [3], with different code systems, as SNOMED-CT or ICD [1]. They identify three elements, (i) electronic healthcare records messages, implemented according to an information model (ii) an ontology or "model of meaning", such as SNOMED-CT, where instances of its concepts are the references in health records, for example "John Smith's diabetes" and (iii) the coding system of the ontology, that "should be a meta model of the model of meaning", where each code (referenced in records messages) is associated to a concept. They propose that the three elements to be separated, with interfaces between the application and each code system.

Schulz et al. [19] analyze SNOMED-CT complex concepts such as the following:

```
ExtrOfForeignBodyFromStomachByExcision \equiv \\ \exists hasPart(\exists procedureSite.StomachStructure \sqcap \\ \exists method.IncisionAction)\sqcap \\ \exists hasPart(\exists procedureSite.StomachStructure \sqcap \\ \exists directMorphology.ForeignBody\sqcap \\ \exists method.RemovalAction) 
(1)
```

They observe that instances of the concept Stomach in the first existential can be different from instances of Stomach in the second existential, and that these "stomachs" could belong to different patients, but they do not give solution to this kind of misinterpretations.

The above overview shows that SNOMED-CT have several structural problems, that nowadays have not been fixed. The definition (1) shows that SNOMED-CT also has problems to be integrated into electronic health records, in the scenario where references to SNOMED-CT terms are instances of the SNOMED-CT concepts. The example shows that its concepts does not properly describe such references. In Section 4 we introduce an approach for giving a solution to this problem, among others we have detected.

4

3 A meta-modelling extension to description logics

In this section we introduce a meta-modelling extension to description logics which allows to treat a given concept A as an individual.

Example 1. Consider the ontology

DiseaseObject(Endocarditis) $Endocarditis \sqsubseteq Disease$

In the first axiom, the name Endocarditis plays the role of an individual whereas in the second axiom it plays the role of a concept. This is possible by dropping the syntactic requirement of description logics that the sets of atomic concepts and individuals should be disjoint. From the semantical point of view, the interpretation domain cannot consist of only basic objects, but it can contain sets, sets of sets, etc. The key point of our semantics is that the interpretation of a given name is the same either as concept or as individual.

Example 2. Consider the following ontology with meta-modelling:

DiseaseObject(Endocarditis) $DiseaseObject \subseteq Endocarditis$

By the defined semantics, $Endocarditis^{\mathcal{I}} \in DiseaseObject^{\mathcal{I}} \subseteq Endocarditis^{\mathcal{I}}$

Here the interpretation of *Endocarditis* belongs to itself, which is nonsense for real applications. To formalize the intuition of Example 2, some definitions are recalled [15, 16] before defining the notion of model of an ontology with meta-modelling.

Definition 1 $(S_n \text{ for } n \in \mathbb{N})$. Given a non empty set S_0 of atomic objects, we define S_n by induction on \mathbb{N} as follows: $S_{n+1} = S_n \cup \mathcal{P}(S_n)$

Definition 2 (Well-founded set). A set X is well-founded if X does not have infinite \ni -decreasing sequences, i.e. there is no $\{x_i \mid i \in \mathbb{N}\} \subseteq X$ such that $x_i \ni x_{i+1}$ for all $i \in \mathbb{N}$.

It is important to see that the sets S_n are well-founded [16]. In Example 2, the interpretation of Endocarditis is not a well-founded set.

Given a logic \mathcal{L} without meta-modelling, with the restriction that the sets of concept, individual and role names are pairwise disjoint, we denote \mathcal{LM} the same logic but dropping the requirement that the sets of atomic concepts and individuals be disjoint. We next define the notion of model for \mathcal{LM} .

Definition 3 (Model of an Ontology with meta-modelling). Given an ontology \mathcal{O} in \mathcal{LM} , an interpretation \mathcal{I} is a model of \mathcal{O} in \mathcal{LM} (denoted as $\mathcal{I} \models \mathcal{O}$) if the following holds:

- 1. the domain $\Delta^{\mathcal{I}}$ of the interpretation is a subset of some S_n for some $n \in \mathbb{N}$.
- 2. I is a model of the ontology \mathcal{O}' in \mathcal{L} obtained from \mathcal{O} by replacing each name N playing the role of an individual and a concept by two new names: a_N for the individual occurrence of N and A_N for the concept occurrence of N.
- 3. For each name N in O, the interpretation of a_N and A_N is the same.

The first part of Definition 3 restricts the domain of an interpretation to be a *subset* of S_n , so $\Delta^{\mathcal{I}}$ can now contain sets of objects. The second part of Definition 3 refers to the semantics for the description logic \mathcal{L} , and the third part restricts the interpretation of a concept name to be the same, either as individual or as concept.

In our previous work, we required that the sets of concepts and individuals should be disjoint and added new axioms of the form $a =_{\mathsf{m}} A$ where a is an individual and A is an atomic concept [15, 16]. These two approaches are equivalent in the sense that it is easy to transform one ontology where the same name N is used as an individual and as a concept by introducing the two new names a_N and A_N and the axiom $a_N =_{\mathsf{m}} A_N$. The approach of [15, 16] is suitable for integrating ontologies where the same real object is represented as an individual in one ontology and as a concept in the other one (the URI's will be necessarily different). But it is not the scenario of our case study where the concept names in SNOMED-CT are treated as individuals.

A tableau algorithm for checking consistency of an ontology with meta-modelling is defined in [15, 16], by adding new rules and a new condition to ensure the well-foundedness of the interpretation domain. This algorithm can be used for an ontology with meta-modelling for the approach presented here, by applying the transformation described above.

An exhaustive analysis of different meta-modelling extensions is done in [16]. Our extension differs from others [6, 8, 9, 11, 17, 5] in that we combine a flexible syntax (without fixed meta-modelling layers) and a strong semantics, Henkin semantics, that gives the same interpretation of a given name either as concept or as individual. It allows us to check for inconsistencies such as that of Example 2, not detected with other semantics, as the Hilog semantics.

4 SNOMED-CT with meta-modelling in electronic health records

In this section, we present a different approach for the integration of electronic health records of patients to SNOMED-CT.

Taking as example the simplified description (2) of the concept Endocarditis, first of all, we describe some problems we found in the model illustrated in Figure 1, in which references to SNOMED-CT concepts in electronic health records are visualized as individuals that are instances of these concepts.

$$Endocarditis \sqsubseteq \exists findingSite.Endocardium \sqcap \\ \exists associatedMorphology.Inflammation$$
 (2)

General knowledge of the health domain is not represented at the proper level. In Figure 1, the patients Juan and Pedro have references to the concept Endocarditis, represented by the instances juanEHRendocarditis and pedroEHRendocarditis. The TBox axiom (2) is consistent with the following ABox axioms:

$$findingSite(juanEHRendocarditis, juanEHRendocardium) \\ findingSite(pedroEHRendocarditis, pedroEHRendocardium)$$
 (3)

6 Motz, Rohrer, Severi

Since the disease endocarditis will always be located in the endocardium, having these assertions at the level of each patient does not add any value, since it is general knowledge of the health domain, which does not differ for each patient.

Definitions of SNOMED-CT concepts does not give a real description of references in electronic health records. The TBox axiom (2) also admits extensions, as illustrated in Figure 2, with the assertions:

Even though the knowledge base is consistent, it does not represent a real situation, because Pedro suffers endocarditis located in the endocardium of Juan!! In order to restrict that references to SNOMED-CT concepts to be linked for the same patients, a more expressive description logic with inverse roles and cardinality restrictions is needed. Considering that SNOMED-CT is already a large knowledge base and that of electronic health records is even larger, a more expressive description logic increase the complexity to exponential, becoming no longer tractable.

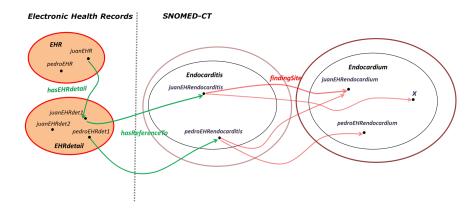


Fig. 2: An extension for the definition of the concept Endiocarditis

Some frequent queries about records of patients can return invalid results. Suppose we want to obtain a chronological report about all clinical situations that affected the Pedro's endocardium, for the scenario of Figure 2. If we formulate the query below, we obtain the instances juanEHRendocardium and pedroEHRendocardium.

$$q(z) = \exists x, y. hasEHR detail(pedroEHR, x) \land hasReferenceTo(x, y) \\ \land findingSite(y, z) \land Endocardium(z)$$
 (5)

From the analysis of the above problems we start elaborating a solution to solve the identified drawbacks. Several works criticize the logical structure of SNOMED-CT, as

well as the interpretation given to its concepts. In particular, Schulz et al. [19] admit that SNOMED-CT concepts does not give a real description of references in electronic health records. However, even though they propose some solutions such as to modify the structure of SNOMED-CT, or to represent SNOMED-CT in a logic more expressive than $\mathcal{OWL}\text{-}\mathcal{EL}$, nowadays SNOMED-CT have the same problems.

As SNOMED-CT is broadly used, we think it is not a realistic approach to change its structure. In the present paper, we introduce a solution that, instead of changing SNOMED-CT, adds a layer that represents the same knowledge at an upper level. Our proposal consists in to treat SNOMED-CT concepts also as individuals, and for the semantics, a given concept name has the same interpretation either as a concept or as an individual, representing the same real object. Moreover, instead of having references to diseases as instances of SNOMED-CT concepts, we propose to link instances of electronic health records directly to the SNOMED-CT terms treated as individuals.

We use the meta-modelling approach described in Section 3, in which the same concept name plays the role of individual or concept depending on its position in the $\mathcal{OWL}\text{-}\mathcal{EL}$ axiom. Then, our proposal is to add to SNOMED-CT a layer of ABox axioms that represent the general health domain knowledge independent from the records of patients. For each TBox axiom containing an existential restriction, such as the description (2), we add an ABox axiom to represent the existential connecting SNOMED-CT concept names treated as individuals through the SNOMED-CT roles. Moreover, records of patients are linked to the SNOMED-CT individuals. Our solution for the scenario of Figure 1, illustrated in Figure 3, is given by the following ABox axioms:

```
has EHR detail(juan EHR, juan EHR det1) \\ has Reference To(juan EHR det1, Endocarditis) \\ finding Site(Endocarditis, Endocardium) \\ associated Morphology(Endocarditis, Inflammation)  (6)
```

With our proposal, to have instances of SNOMED-CT concepts such as juanEHRendocarditis and juanEHRendocardium, connected by SNOMED-CT roles becomes unnecessary because now we have this kind of information at the level of the meta-model layer. The SNOMED-CT terms represented as individuals are now connected through the SNOMED-CT roles. We argue that the SNOMED-CT meta-model has some advantages that are explained below.

Facilitates a modular design and reuse. As SNOMED-CT is now broadly used, we propose a solution that keeps SNOMED-CT as a hierarchy of concepts, favoring the reuse of existing ontologies. We think the integration of SNOMED-CT in medical applications can be enhanced by adding a meta-model of the hierarchy of concepts.

Connects patients to medical terms at the proper level. We represent the health domain through two different layers with different purposes. In the lower level, we have the SNOMED-CT ontology that represent the hierarchy of medical terms, distinguishing more general from more specialized concepts. The upper level is defined to link electronic health records of patients to medical terms. As illustrated in Figure 3, for the patient Juan there is a record represented by the individual juanEHRdet1 that is

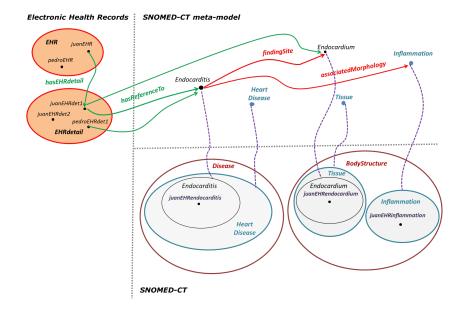


Fig. 3: SNOMED CT with the meta-model layer

linked to the name Endocarditis treated as individual. To connect the upper with the lower layer we apply the meta-modelling approach described in Section 3, that maps a given name to a unique interpretation, either as individual or as concept. In Figure 3, the individual Endocarditis and the concept Endocarditis represent the same real object.

Avoids redundancy of SNOMED-CT role instances. The representation of references to SNOMED-CT concepts as instances of them has the drawback that roles of SNOMED-CT link individuals in a redundant way, as showed in (3). It is more intuitive to represent this kind of knowledge at a more general level, independently of the electronic health records of patients. So, by representing medical terms as individuals, we avoid to having SNOMED-CT role instances at the level of each patient. Hence, we have just ABox axioms such that findingSite(Endocarditis, Endocardium), illustrated in Figure 3 through an arrow representing the role findingSite, that connects the medical terms Endocarditis and Endocardium in the meta-model, but not at the level of patients.

Prevents from invalid extensions of SNOMED-CT concepts. The TBox axiom (2) admits extensions such that the assertions (4). It is avoided in our approach because records of patients are directly connected to SNOMED-CT terms as individuals. These individuals are in a meta-model layer that describe how the medical terms are conceptually related, whereas the lower layer describes the hierarchy of the vocabulary.

Provides a more direct mechanism to query records of patients, avoiding unexpected results. Let's come back to the example of obtaining a report about all situations that

affected the Pedro's endocardium. With our solution, it is sufficient to obtain the records of Pedro that point to SNOMED-CT individuals related to the individual Endocardium. Then, this kind of queries can be solved at the upper level going through the SNOMED-CT individuals. The query for the new approach is:

$$q(x) = \exists y.hasEHRdetail(pedroEHR, x) \land hasReferenceTo(x, y) \land findingSite(y, Endocardium)$$

$$(7)$$

Links upper and lower levels to infer useful information and detect inconsistencies. Finally, we argue why it is important that SNOMED-CT terms both as individuals and concepts to be interpreted as the same real objects. Suppose we have the patient Juan who suffers endocarditis. In order to exploit the SNOMED-CT hierarchy, it is useful to get inferences like "if Juan has endocarditis then Juan has a heart disease", and in this case to obtain all patients that suffer a heart disease. We can formulate the query:

$$q(x) = \exists y, z. hasEHR detail(x, y) \land hasReferenceTo(y, z) \land z \sqsubseteq HeartDisease \tag{8}$$

Here we exploit the fact that SNOMED-CT terms can be treated either as individuals or concepts. The set of solutions is the set of electronic health records of patients that reference SNOMED-CT individuals which, treated as concepts are subsumed by the concept HeartDisease. Moreover, giving the same interpretation to SNOMED-CT terms as individuals and concepts we also prevent from inconsistencies, such as those of Example 2, where the interpretation domain becomes a non-well founded set.

5 Conclusions and future work

In this paper, we have analyzed the logical model of the ontology SNOMED-CT, in a case study of electronic health records of patients referencing SNOMED-CT terms. Several works have proposed to modify the logical structure of SNOMED-CT. However, we propose an approach that, on the one hand, keeps the SNOMED-CT ontology unchanged, extending it with a meta-model layer, and on the other hand, gives a solution to the problem of the population of SNOMED-CT presented by Schulz et al. [19] for the "stomachs" example. Moreover, general knowledge about the health domain is represented at an upper level. In this layer SNOMED-CT concepts are treated as individuals, and the key point is that they are semantically equal in both layers. To formalize our approach, we slightly adapt an existing meta-modelling extension of description logics. As future work we aim to implement our approach in a concrete case study of a medical institution, to integrate electronic health records of patients with SNOMED-CT concepts, as well as for other medical applications.

References

- 1. Icd. http://www.who.int/classifications/icd/en/, Last date accessed December 2016.
- F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, 2003.

- 3. T. Benson and G. Grieve. *Principles of Health Interoperability. SNOMED CT, HL7 and FHIR*. Springer-Verlag, 2016.
- 4. S. El-Sappagh, M. E. Mogy, and A. M. Riad. A standard fragment of EHR relational data model for diabetes mellitus diagnosis. In *Proceedings of Informatics and Systems*, pages 235–243, 2014.
- G. De Giacomo, M. Lenzerini, and R. Rosati. Higher-order description logics for domain metamodeling. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelli*gence, AAAI 2011, pages 183–188. AAAI Press, 2011.
- 6. B. Glimm, S. Rudolph, and J. Völker. Integrated metamodeling and diagnosis in OWL 2. In *International Semantic Web Conference, ISWC 2010*, pages 257–272, 2010.
- P. Hitzler, M. Krötzsch, and S. Rudolph. Foundations of Semantic Web Technologies. Chapman & Hall/CRC, 2009.
- 8. M. Homola, J. Kluka, V. Svátek, and M. Vacura. Towards typed higher-order description logics. In *Proceedings of DL Workshop*, pages 221–233, 2013.
- M. Homola, J. Kluka, V. Svátek, and M. Vacura. Typed higher-order variant of SROIQ why not? In *Proceedings of DL Workshop*, pages 567–578, 2014.
- 10. IHTSDO. SNOMED CT Starter Guide, 2014.
- 11. N. Jekjantuk, G. Gröner, and J. Z. Pan. Modelling and reasoning in metamodelling enabled ontologies. *I. J. Soft. Informatics*, 4(3):277–290, 2010.
- 12. Bevan Koopman, Guido Zuccon, Anthony Nguyen, Deanne Vickers, Luke Butt, and Peter D. Bruza. Exploiting snomed ct concepts and relationships for clinical information retrieval: Australian e-health research centre and queensland university of technology at the trec 2012 medical track. In 21st Text REtrieval Conference (TREC 2012), pages 1–8, 2012.
- 13. M. Krötzsch. OWL 2 profiles: An introduction to lightweight ontology languages. In *Reasoning Web. Tutorial*, pages 112–183, 2012.
- 14. M. Martinez, E. Rohrer, and P. Severi. Complexity of the description logic ALCM. In *Proceedings of KR 2016*, pages 585–588, 2016.
- R. Motz, E. Rohrer, and P. Severi. Reasoning for ALCQ extended with a flexible metamodelling hierarchy. In Proceedings of JIST, Lecture Notes in Computer Science, pages 47–62, 2014.
- 16. R. Motz, E. Rohrer, and P. Severi. The description logic *SHIQ* with a flexible meta-modelling hierarchy. *J. Web Sem.*, 35(4):214–234, 2015.
- 17. J. Z. Pan, I. Horrocks, and G. Schreiber. OWL FA: A metamodeling extension of OWL DL. In *OWLED*, CEUR Workshop Proceedings, 2005.
- A. L. Rector, R. Qamar, and T. Marley. Binding ontologies and coding systems to electronic health records and messages. *Applied Ontology*, 4:51–69, 2009.
- 19. Schulz S., Markó K., and Suntisrivaraporn B. Formal representation of complex SNOMED CT expressions. *BMC Med. Inf. & Decision Making*, 8(S-1):S9, 2008.
- S. Schulz, R. Cornet, and K. A. Spackman. Consolidating snomed ct's ontological commitment. *Applied Ontology*, 6(1):1–11, 2011.
- S. Schulz, B. Suntisrivaraporn, F. Baader, and M. Boeker. SNOMED reaching its adolescence: Ontologists' and logicians' health check. *I. J. Medical Informatics*, 78:S86–S94, 2009.
- M. Q. Stearns, C. Price, K. A. Spackman, and A. Y. Wang. SNOMED clinical terms: overview of the development process and project status. *Proceedings of the AMIA Sym*posium, pages 662–666, 2001.
- 23. B. Suntisrivaraporn, F. Baader, S. Schulz, and K. A. Spackman. Replacing SEP-triplets in SNOMED CT using tractable description logic operators. In *Proceedings of AIME 2007*, pages 287–291, 2007.
- W3C. OWL 2 Web Ontology Language Profiles (Second Edition), Last date accessed May 2016