# Ontology-based reinterpretation of the SNOMED CT context model

Catalina Martínez-Costa[1,*] Stefan Schulz[1,2]

[1] Institute of Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria

[2] Institute of Medical Biometry and Medical Informatics, Freiburg University Medical Center, Germany

## ABSTRACT

SNOMED CT includes concepts that encode complex expressions in its context model under *Situation with explicit context*, which blends characteristics of information models with characteristics of ontologies. In order to improve interoperability of isosemantic expressions that are constituted by different information model / ontology combinations, we propose an ontology-based reinterpretation of four of the most representative patterns found in the SNOMED CT context model. The formalizations provided require the use of negation and universal quantification, thus requiring a shift from OWL-EL to OWL-DL.

After a thorough analysis of the meaning of the SNOMED CT concepts that instantiate these patterns as well as ontological errors in the current OWL-EL of SNOMED CT rendering we transformed a module of SNOMED CT according to these patterns. The classification performance of the resulting ontology was benchmarked for several transformation steps. Classification times remained under 1s for FaCT++ and under 3s for HermiT. Although the SNOMED CT module comprised only those concepts which are needed for expressing the content of the context model (about 5% of the complete context model), the results are encouraging as they suggest that the limited inclusion of OWL-DL expressiveness does not lead to inacceptable performance results.

## 1 INTRODUCTION

SNOMED CT[1], a clinical terminology covering the whole clinical domain is an emerging standard for the representation of semantically explicit, structured information in electronic health records by providing more than 300,000 meaning-bearing representational units (concepts). SNOMED CT has increasingly been guided by principles of Applied Ontology (Schulz & Karlsson, 2011), as it has become clear that a thorough grounding of the meaning of domain terms using a formal language and rooted in an ontological framework is a key requirement for the communication of meaning between humans and machines. SNOMED CT does not only provide codes for clinical terms proper but also for contextual statements, for which its own hierarchy is reserved, named 'Situation with explicit context', commonly known as the SNOMED CT **Context Model**. Typical for this is the blending of epistemic with ontological content such as *Suspected deep vein thrombosis* or *No past history of venous thrombosis*. The purpose of such expressions is to facilitate the recording of complex clinical information by using a single code. Under ontological scrutiny these concepts correspond to information entities rather than clinical entities: whereas the concept *Deep vein thrombosis* can be conceived to be instantiated by real thromboses (or situations with thrombosis), this is not the case with instances of *Suspected deep vein thrombosis*. They are not thromboses but rather statements about the concept *Thrombosis*, and therefore categorially different. Information entities should represent the contextual and epistemic aspects of information (e.g. suspicion, past history); and they refer to types/ concepts of clinical entities which are not necessarily instantiated (Schulz et al., 2009): the statement "suspected vein thrombosis" makes perfect sense even if the patient has no thrombosis.

In theory, systems that represent the meaning of domain terms (terminologies / ontologies) should be kept separate from systems that place that meaning in a situational or epistemic context (information models) (Rector, 2001; Schulz et al., 2010). In practice, both approaches tend to overlap, which produces a plurality of syntactically and terminologically different possibilities to encode the same piece of clinical information. This plurality constitutes a severe barrier to semantic interoperability: for two or more systems to be semantically interoperable, they must interpret clinical information in the same way (Stroetmann et al., 2009). As ontologies have the potential of formalizing the meaning of clinical information, they could be seen as semantic building blocks for sharing information across systems.

Not only SNOMED CT but also the upcoming ICD-11[2] is based on formal ontology principles. In the context of harmonizing future versions of SNOMED CT and ICD we have provided an ontology-based re-interpretation of what SNOMED CT categorizes as *Findings* and *Disorders*, and what constitutes most parts of the foundation component of ICD 11. According to Schulz et al. (2012) the concepts in these hierarchies denote clinical *situations* rather than clinical *conditions*, i.e. as phases of a patient's life in which a clinical condition is present at all times. Both naming aspects and the makeup of the taxonomies provide evidence that the *Situation* interpretation corresponds more to the intuition of the terminology builders and users.

Confusingly, SNOMED CT has already categorized as *Situations* all concepts under *Situation with explicit context (situation)*, i.e the already mentioned context model, on which we concentrate our further deliberations. Context model concepts only partially correspond to our notion of *Situation*, whereas many of them fall into the category of *Information entities*.

SNOMED CT formally represents the content of the context model in the same way as the content of other hierar-

* To whom correspondence should be addressed: catalina.martinez@medunigraz.at

[1] http: //www.ihtsdo.org/snomed-ct

[2] http://www.who.int/classifications/icd/

chies, *viz*. using the SNOMED CT compositional syntax, which can be transformed in OWL EL expressions according to IHTSDO (2012). Past analyses (Rector and Brandt 2008, Schulz and Karlsson, 2011) have shown that this expressivity is not sufficient for appropriately representing the meaning needed for context model concepts.

In this paper we provide definitional patterns of four frequent types of SNOMED CT context model concepts. Some of them require the use of negation as well as of universal quantification, thus requiring OWL DL expressivity. We will develop representational patterns into which the current representations will be transformed. We will then evaluate the performance of a redesigned ontology.

## 2 MATERIALS AND METHODS

### 2.1 The SNOMED CT context model

When contextual information is added to a term, its meaning generally changes in a way that the resulting term no longer denotes a specialization of the referents of the original term. For instance, "suspected heart failure" does not denote a special kind of heart failure. Consequently, in SNOMED CT the concepts *Suspected heart failure* and *Heart failure* are not taxonomically related but by the relation **'associated finding'**. Ideally, epistemic and contextual aspects of clinical documentation should be represented in information models and not in ontologies, which should therefore not include terms like *Suspected X* (Jansen and Schulz, 2011). However, SNOMED CT was designed to be usable without being embedded in a specific information model. Its context model therefore plays the role of a simplified in-built information model, which uses dedicated relations and constraints such as illustrated in Table 1 for clinical findings.

| *Defining Attribute* | *Permissible Values* |
|---|---|
| Subject relationship context | *Person* 125676002 |
| Temporal context | *Temporal context value* 410510008 |
| Finding context | *Finding context value* 410514004 |
| Associated finding | *Clinical finding* 404684003 |

**Table 1.** Context model - Situation with explicit context

Accordingly, the concept *Heart failure excluded* (*situation*) is represented in the description logics interpretation of the SNOMED CT syntax (Spackman et al., 2002) as:

'*Heart failure excluded* (*situation*) ' **subclassOf**
**RoleGroup** some (('**associated finding**' some '*Heart failure*')
and ('**finding context**' some '*Known absent*')
and ('**temporal context**' some '*Current of specified*')
and ('**subject relationship context**' some '*Subject of record*'))

If the opposite is meant, *viz*. that the heart failure is *present*, the value of the relation '**finding context**' has to be changed to '*Known present*'. As it can be observed, in the same concept three types of information are merged: (i) epistemic information such as *Known absent*, *Suspected*, or *Changed,* (ii) other contextual information like temporal reference (*Current*, *Past*) or subject relationship (*Subject of record*, *person in the family*), and (iii) the reference to the clinical concept proper (*Heart failure*). From OWL semantics it is immediately clear that the existential statements 'some' contradict the intended meaning, as here '*Heart failure excluded* (*situation*)' would logically entail the existence of at least one instance of 'Heart failure'.

There are other cases, such as *Suspected heart failure*, in which we are not referring directly to a situation in which heart failure is clearly present or absent, but to the state of knowledge of the author of this statement. This justified our decision to categorize this type of SNOMED concepts under the BioTopLite[3] category *Information Object*.

### 2.2 Definition of patterns

We generally interpret SNOMED CT findings and diseases as clinical situations that include conditions. In BioTopLite, *Condition* is defined as the disjunction of disposition, material object and process (Schulz et al., 2011). A clinical situation is related to a clinical Condition via the relation **hasCondition**:

*ClinicalSituation_*$X_S$ **equivalentTo**
*Situation* and **hasCondition** some *ClinicalCondition_*$X_C$

| *Original SNOMED CT category* | *Reinterpretation based on the BioTopLite ontology* |
|---|---|
| *Clinical finding (finding)* (*Disorder* being a specification thereof) | *ClinicalSituation* and **hasCondition** some *ClinicalCondition* |
| *Situation with explicit context (situation)* ("context model") | *ClinicalSituation* and **hasCondition** some *ClinicalCondition* |
| | *InformationEntity* and **isAboutSituation** only *ClinicalSituation* |

**Table 2.** Reinterpretation of *Clinical findings* and *Situation* (concept model) concepts from SNOMED CT

The relation **hasCondition** corresponds to the Role Group relation in SNOMED CT. According to our reinterpretation, conditions are only indirectly present in SNOMED CT, namely as the post-coordinated expressions within the role groups. Table 2 reflects our current analysis of the ontological categories of SNOMED CT, restricted to findings and disorders, as well as to related situations. Sometimes SNOMED CT concepts indirectly refer to clinical situations by stating some information about them, such as *suspected* or *at risk*. These concepts are categorized as information entities, which refer to the clinical situation via the relation ***isAboutSituation***:

[3] http://purl.org/biotop

*InformationEntityAboutClinicalSituation_*$X_S$ **equivalentTo**
*InformationEntity* and **isAboutSituation** only *ClinicalSituation_*$X_S$
and **hasInformationObjectAttribute**
some *InformationObjectAttribute*

In this expression, we use the universal quantification ('only') for relating an information entity to a clinical situation, to avoid asserting the existence of an entity the existence of which cannot be guaranteed. We consider this an acceptable approximation, as second-order statements (information individual being about a type or concept) are not expressible within OWL DL.

Once clinical situations have been introduced we can focus on the definition and implementation of the following four patterns that correspond to four frequent design patterns found in the SNOMED CT context model, viz. (i) *Clinical finding present*, (ii) *Clinical finding absent*, (iii) *Suspected clinical finding*, and (iv) *No history of clinical finding in subject*. The four of them follow the context model shown in Table 1. These patterns consider that the values of the relations '**subject relationship context**' and '**temporal relationship**' are restricted by '*Subject of record*' and '*Current or specified*', respectively.

*Clinical finding present (situation)*: A clinical finding present is characterized by the qualifier value '*known present*'. As we consider SNOMED CT concepts in the findings hierarchy as situations, we here simply equate the situation concept with the finding concept.

*ClinicalFindingPresentSituation_*$X_S$ **equivalentTo**
*ClinicalFinding_*$X_F$

According to our interpretation of findings as clinical situations, *Night cough present (situation)* would be equivalent to *Nocturnal cough (finding)*.

*Clinical finding absent (situation)*: Originally, it had the **finding context** value '*known absent*'. We have re-interpreted it as a clinical situation in which the finding is not present at any time.

*ClinicalFindingAbsentSituation_*$X_S$ **equivalentTo**
*ClinicalSituation* and not
(**hasProcessualPart** some *ClinicalFinding_*$X_F$)

*No history of clinical finding in subject (situation)*: It defines the **temporal context** as '*all times past*' and the **finding context** as '*known absent*'. We interpret it by stating that there is not part of the patient's life in which the condition was present and we refer to it by using an information entity and the quantifier *only*.

*NoHistoryOfClinicalFindingSituation_*$X_s$ **equivalentTo**
*InformationEntity* and **isAboutSituation** only
(*BiologicalLife* and not (**hasProcessualPart** some
*ClinicalFindingPresentSituation_*$X_S$))

*Suspected clinical finding (situation)*: A suspected clinical finding includes the **finding context** value '*Suspected* '. We have defined it as an information entity that refers to a possible situation with some condition present during all time. In this case, the epistemic attribute '*Suspected* ' qualifies the information entity. Again, the universal quantifier (only) has been used to avoid asserting the existence of a situation with that condition, since it is a suspect.

*SuspectedClinicalSituation_*$X_s$ **equivalentTo**
*InformationEntity*
and **isAboutSituation** only *ClinicalFindingPresentSituation_*$X_S$
and **hasInformationObjectAttribute** some *Suspected*

## 2.3 Module creation

Subject to our study was a module extracted from the SNOMED CT July 2012 release, using the context model concepts as signature and following an *Upwards & Upwards from Links* strategy, in variation of the approach proposed by Seidenberg & Rector (2006), implemented as a Perl script, which takes the relational tables of the SNOMED CT distributions as input (Table 3). The algorithm traverses the hierarchy upwards to the root class, starting with the *Situation* concepts and adding all their parents. In addition it also add those concepts linked to them. It continues until no new concepts and links are added.

```
Input:
  sct_concepts;  //concepts from input file
  sct_rels;  //relations from input file
Vars:
  sit_concepts;  //all situation concepts
  rel_sit_concepts;  //all situation related concepts
  sit_rels;  //all relations with at least one situation concept

for each concept from sct_concepts
 if concept.label match '(Situation)' add concept to sit_concepts

for each rel from sct_rels
 if rel.subject.label | rel.object.label match'(Situation)' add rel to sit_rels

Do
 for each rel from sit_rels
    if rel.subject.label match '(Situation)' & !exist in rel_sit_concepts
      Add rel.subject to rel_sit_concepts
      Add parents of rel.subject to rel_sit_concepts
    if rel.object.label match '(Situation)' & !exist in rel_sit_concepts
      Add rel.object to rel_sit_concepts
      Add parents of rel.object to rel_sit_concepts
  end
  for each concept from rel_sit_concepts
    Add all rels from sct_rels to sit_rels
       if rels.type.label match 'is_a' & concept eq rels.subject
       if rels.type.label !match'is_a'& rels.subject | rels.object eq concept
   end

Until no new concept is added to rel_sit_concepts

Write sit_concepts and related_sit_concepts to concept_file.txt
Write sit_rels to rels_file.txt
```

**Table 3**. SNOMED CT module creation algorithm

## 2.4 Implementation of the patterns

The implementation of the four patterns presented in subsection 2.2 targeted an OWL-DL representation. We started with the module created according to the algorithm in Table 3, as input to the Perl script delivered with the SNOMED CT distribution. Applied to our module it transforms the relational format into an OWL-EL ontology, containing 10,773 classes, 48 object properties, 5,381 subclass axioms, and 5,391 equivalence axioms.

For the transformation of this ontology into an OWL-DL version we made an implementation in Java using the OWL API (Horridge et al., 2009). This implementation takes the concepts from each hierarchy and modifies their axioms according to the specific patterns proposed. The following tables show the representation of example concepts instantiating each of the four patterns studied.

It can be observed how the representation of the *Alexia and agraphia present* situation concept is simplified by interpreting the concepts from the *Clinical finding* hierarchy as *Clinical situations*. Besides, we maintain the right inferences since a *Situation with Agraphia and Alexia* is also classified as an *Agraphia situation* and as an *Alexia situation*, independently. Thus the representational patterns of the finding hierarchy coincide with the "default" patterns from the context model.

| CLINICAL FINDING PRESENT |
|---|
| **Original SNOMED CT representation** |
| *Alexia and agraphia present (situation)* **equivalentTo**<br>*'Clinical finding present (situation)'*<br>and **RoleGroup** some ('**associated finding**' some '*Agraphia (finding)*'<br>and '**finding context**' some '*Known present*'<br>and '**temporal context**' some '*Current or specified*'<br>and '**subject of record**' some '*Subject of record*')<br>and **RoleGroup** some ('**associated finding**' some '*Alexia(finding)*'<br>and '**finding context**' some '*Known present*'<br>and '**temporal context**' some '*Current or specified*'<br>and '**subject of record**' some '*Subject of record*') |
| **Modified SNOMED CT representation** |
| *Alexia and agraphia present (situation)* **equivalentTo**<br>'*Agraphia (finding)*' and '*Alexia(finding)*' |

**Table 4.** Clinical finding present pattern example

The *Clinical finding absent* pattern is exemplified by *Night cough absent*, which in SNOMED CT is defined as *No cough* and the expression **associated finding** some *nocturnal cough* nested in a role group (to be interpreted as **hasCondition**). At the beginning, we had applied this pattern maintaining the original taxonomy. This lead to numerous wrong inferences, which surfaced as numerous inferred equivalence statements. Therefore, we removed the original parents and obtained the inverted hierarchy. It means that the terminology concept *no cough* is now classified <u>under</u> *night cough absent*. The original SNOMED CT hierarchy, albeit intuitive at a first glance, appeared as faulty under closer scrutiny: *No cough* <u>above</u> *Night cough absent* would mean that a patient without night cough would have no cough at all, which is certainly not intended.

| CLINICAL FINDING ABSENT |
|---|
| **Original SNOMED CT representation** |
| '*Night cough absent (situation)*' equivalentTo<br>'*No cough (situation)*' and **RoleGroup**<br>some ('**associated finding**' some '*Noctural cough (finding)*'<br>and '**finding context**' some '*Known absent*'<br>and '**temporal context**' some '*Current or specified*'<br>and '**subject of record**' some '*Subject of record*') |
| **Modified SNOMED CT representation** |
| '*Night cough absent (situation)*' equivalentTo<br>*ClinicalSituation*<br>and not ('**has processual part**' some '*Nocturnal cough (finding)*') |

**Table 5.** Clinical finding absent pattern example

If we analyze the equivalent axiom defined for the context model concept *Suspected neuroblastoma,* it states that it is equivalent to a *Suspected brain tumor situation*, a *Suspected head and neck cancer situation,* a *Suspected central nervous system cancer situation* and a role group expression that includes *Neuroblastoma of brain* as **associated finding**. This definition is, again, incorrect. A neuroblastoma can be placed in many parts of the body. Therefore this concept does not encompass all kinds of suspected neuroblastomas but only the ones that are located in the brain. Then, it should have been better named *Suspected neuroblastoma of brain*. Moreover, instead of including in the definition of the situation concept all the situations that are a kind of brain neuroblastoma, this information should be inferred as it happens with the proposed representation, in which the three situation concepts are inferred as subclasses.

| SUSPECTED CLINICAL FINDING |
|---|
| **Original SNOMED CT representation** |
| '*Suspected neuroblastoma (situation)*' equivalentTo<br>'*Suspected brain tumor (situation)*' and<br>'*Suspected head and neck cancer (situation)*' and<br>'*Suspected central nervous system cancer (situation)*' and<br>**RoleGroup** some<br>('**associated finding**' some '*Neuroblastoma of brain (disorder)*'<br>and '**finding context**' some '*Suspected*'<br>and '**temporal context**' some '*Current or specified*'<br>and '**subject of record**' some '*Subject of record*') |
| **Modified SNOMED CT representation** |
| '*Suspected neuroblastoma of brain (situation)*' equivalentTo<br>*shn_information_item*<br>and **shn_hasInformationObjectAttribute** some *shn_suspected*<br>and **shn_isAboutSituation** only '*Neuroblastoma of brain (disorder)*' |

**Table 6.** Suspected clinical finding pattern example

| NO HISTORY OF CLINICAL FINDING IN SUBJECT |
|---|
| **Original SNOMED CT representation** |
| '*No past history of venous thrombosis (situation)*' equivalentTo<br>'*No history of cardiovascular system disease (situation)*' and<br>'*No thrombus (situation)*' and<br>**RoleGroup** some<br>('**associated finding**' some '*Venous thrombosis (disorder)*'<br>and '**finding context**' some '*Known absent*'<br>and '**temporal context**' some '*All times past*'<br>and '**subject of record**' some '*Subject of record*') |
| **Modified SNOMED CT representation** |
| '*No past history of venous thrombosis (situation)*' equivalentTo<br>*shn_information_item* and<br>**shn_isAboutSituation** only ('*biological life*' and<br>not ('**has processual part**' some<br>'*Disorder of cardiovascular system (disorder)*')) |

**Table 7.** No history of clinical finding in subject pattern example

Finally, let us examine *No past history of venous thrombosis*. It includes in its equivalent axiom two situation concepts, *No history of cardiovascular system disease* and *No thrombus situation*, together with a role group expression in which the **associated finding** is *venous thrombosis disorder*. Again, the situation concepts should not be in the equivalence axiom but in case they were right they should be inferred. In this case by using our proposed representation we could check that both of them were wrong. First, that the patient does not have a past history of venous thrombosis does not mean that he/she might have the disease now. It means that not having a past history of venous thrombosis is not a kind of *No thrombosis* situation, since this last one refers to the present patient status and therefore this concept should be removed from the equivalent axiom. Second, not having a past history of venous thrombosis does not imply that the patient might have had a past history of another cardiovascular system disease such as migraine. It means that the concept *No history of cardiovascular system disease* is also wrongly placed in the definition. By using the representation proposed we correctly get *No past history of venous thrombosis* as a superclass of *No history of cardiovascular system disease*. If querying for the first expression you also expect patients coded with the second one, because who has no history of cardiovascular system disease at all, won't have had a thrombosis. Especially in this hierarchy, by using our reinterpretation we easily found many incorrect hierarchical links as a result of analyzing the inferences produced by the classifier.

## 3 PERFORMANCE EVALUATION

As we moved away from an inexpressive logic to a more expressive one, performance issues are of utmost importance, as in the worst case, the ontology would not be scalable and the whole approach which appears promising in theory would become intractable in real-world scenarios.

The performance of the remodeled ontology will be measured by using two OWL DL reasoners, HermiT (Shearer et al., 2008) and FaCT ++ (Tsarkov et al., 2006). The experiments are limited to T-Boxes. Both reasoners were made accessible via the OWL API, which facilitated the execution of the experiments. The aim of these experiments is to see how the application of the different patterns affects the reasoning time. The rationale followed has been:

- Extraction of the context model module from the original ontology
- Application of the four patterns independently. As a result four versions of the ontology are produced.
- Application of the patterns in groups of two. As a result six versions of the ontology are produced.
- Application of the patterns in groups of three. As a result four versions of the ontology are produced.
- Application of the four patterns at the same time to the ontology.

Tables 8 and 9 show the reasoning times obtained for the FaCT++ and the HemiT reasoners considering classification and consistency check tasks and using a MacBook Air with 1.7 GHz Intel Core i5 and 4GB 1333 MHz DDR3 memory. Only the time for the two-fold and three-fold pattern combination are shown. The other times are shown in Figure 1. In the tables each letter corresponds to a pattern: (P) clinical finding present (82 classes); (A) clinical finding absent (322 classes); (S) suspected clinical finding (66 classes); (N) no past history of clinical finding in subject (25 classes).

| Two Patterns | **PA** | **PS** | **PN** | **AS** | **AN** | **SN** |
|---|---|---|---|---|---|---|
| | 0.935 | 0.874 | 0.536 | 0.670 | 0.721 | 0.608 |
| Three Patterns | **PAS** | **PAN** | **ASN** | **PSN** | | |
| | 0.752 | 0.691 | 0.677 | 0.579 | | |

**Table 8.** Reasoning time in seconds with FaCT++

| Two Patterns | **PA** | **PS** | **PN** | **AS** | **AN** | **SN** |
|---|---|---|---|---|---|---|
| | 2.888 | 2.339 | 1.646 | 1.430 | 1.442 | 2.763 |
| Three Patterns | **PAS** | **PAN** | **ASN** | **PSN** | | |
| | 1.288 | 1.420 | 1.552 | 1.637 | | |

**Table 9.** Reasoning time in seconds with HermiT

Finally, Figure 1 visualizes the effect of adding the patterns by plotting the reasoning time against the patterns in increasing complexity (two-fold and three-fold are calculated as the average of the time for each pattern combination). It has to be noticed that the number of concepts of each of the four hierarchies to which the patterns have been applied are different. The results obtained with FaCT ++ are better than the results obtained with HermiT. This is also true with the original version of the ontology, without changing any axiom in which reasoning takes 0.670s with FaCT ++ and 1.687s with HermiT. Therefore the reasoning times without changing any axiom is in most cases not lower than the obtained after applying any of the patterns. In general, the addition of patterns increases the reasoning time but not significantly, using less than a second for classifying the ontology that includes the four patterns modification. Both curves have a peak, for the two patterns case, but it is significantly greater in the case of HermiT. It looks counterintuitive that using two patterns entails longer reasoning times than using three. This might be caused by certain internal features of the reasoners such as optimization strategies. This phenomenon will be communicated to the developers.

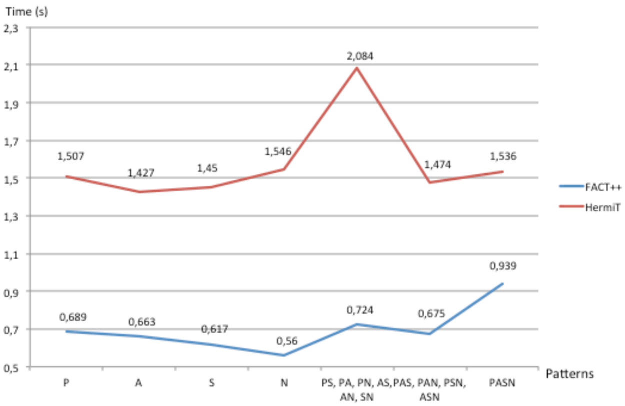


**Figure 1.** Reasoning time in seconds with FaCT++ and HermiT

## 4 CONCLUSIONS

SNOMED CT is a terminology partially built on formal ontological principles, but its top-level categories and relations are still influenced by the legacy of its predecessors. This is evident in its *Situation* hierarchy, which blends characteristics of information models with characteristics of ontologies. Its architecture strongly evokes old frame formalisms, which cannot be transformed into OWL ontologies in a straightforward way, especially when negated concepts have to be arranged in hierarchies. Our stepwise designing and testing of the transformation patterns have revealed some of the original defects; a process in which the iterative use of a description logics reasoner was of great heuristic value. Our analysis also supported our view that the SNOMED CT concept model is heterogeneous, i.e. only part of its concepts can be conceived as denoting clinical situations, whether many others include epistemic aspects and should therefore be categorized as information entities. An encouraging result was that – at least for a module 20 times smaller as the complete SNOMED CT – the shift from OWL-EL to OWL-DL had no devastating consequences with regard to reasoning performance. More scaling tests will have to be performed, but the results of this study should encourage the SNOMED CT developers to embark on more expressive formalisms in future versions. Future work will include the study of the impact on the performance when more patterns are applied to SNOMED CT and also to the whole ontology and not just to modules as here. This might require the use of modularization-based OWL reasoning techniques in which several reasoners are applied to different ontology fragments (Armas et al., 2012).

## ACKNOWLEDGEMENTS

This work has been funded by the SemanticHealthNet Network of Excellence within the EU 7th Framework Program, Call:FP7-ICT-2011-7, agreement 288408. http://www.semantichealthnet.eu/

## REFERENCES

Armas Romero A, Cuenca Grau B, Horrocks I. MORe: Modular Combination of OWL Reasoners for Ontology Classification. International Semantic Web Conference, (1) 2012:1-16.

Horridge M, Bechhofer S. The OWL API: A Java API for Working with OWL 2 Ontologies. In 6th OWL ED Workshop, 2009.

IHTSDO. SNOMED CT® Technical Implementation Guide July 2012. Copenhagen, Denmark, 2012. http://www.snomed.org/tig.pdf

Jansen L, Schulz S. The Ten Commandments of Ontological Engineering. Third Workshop of Ontologies in Biomedicine and Life Sciences (OBML),2011. http://www.onto-med.de/obml/ws2011/obml2011report.pdf

Rector AL, Brandt S. Why do it the hard way? The case for an expressive description logic for SNOMED. J Am Med Inform Assoc. 2008 Nov-Dec;15(6):744-751.

Rector AL. The interface between information, terminology, and inference models. Stud Health Technol Inform. 2001;84(Pt 1):246-250.

Schulz S, Karlsson D. Records and situations. Integrating contextual aspects in clinical ontologies. The 14th Annual Bio-Ontologies Meeting. Vienna, Austria: ISCB; 2011, 49-52.

Schulz S, Rector A, Rodrigues JM, Spackman K. Competing Interpretations of Disorder Codes in SNOMED CT and ICD. AMIA Annu Symp Proc. 2012;2012:819-827.

Schulz S, Schober D, Daniel C, Jaulent MC. Bridging the semantics gap between terminologies, ontologies, and information models. Stud Health Technol Inform. 2010; 160 (Pt 2): 1000-1004.

Schulz S, Spackman K, James A, Cocos C, Boeker M. Scalable representations of diseases in biomedical ontologies. J Biomed Semantics, 2011, 17;2 Suppl 2:S6

Schulz S, Stenzhorn H, Boeker M, Smith B. Strengths and limitations of formal ontologies in the biomedical domain. Rev Electron Comun Inf Inov Saude. 2009 Mar 1;3(1):31-45.

Seidenberg J, Rector AL: Web ontology segmentation: analysis, classification and use. WWW 2006: 13-22

Shearer R, Motik B, Horrocks I, HermiT: a Highly-Efficient OWL Reasoner. In 5th OWL Experienced and Directions Workshop, 2008.

Spackman KA, Dionne R, Mays E, Weis J. Role grouping as an extension to the description logic of Ontylog, motivated by concept modeling in SNOMED. Proc AMIA Symp. 2002:712-716.

Stroetmann VN et al. Semantic Interoperability for better health and safer healthcare - Deployment and Research Roadmap for Europe. SemanticHEALTH Report: European Communities, 2009. http://ec.europa.eu/information_society/activities/health/docs/publications/2009/2009semantic-health-report.pdf

Tsarkov D, Horrocks I. FaCT++ Description Logic Reasoner: System Description. In Third International Joint Conference on Automated Reasoning, 292-297. Springer, 2006.