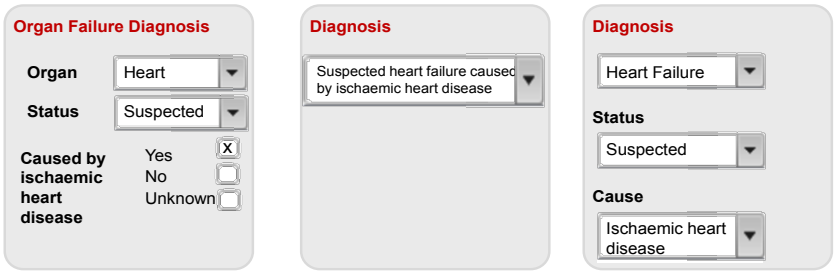
Approximating DL reasoning by ontology-aware RDF querying

Catalina Martínez-Costa, Stefan Schulz

IMI, Medical University of Graz, Austria
{catalina.martinez,stefan.schulz}@medunigraz.at

Abstract. We present initial experiments of transforming an expressive OWL DL model into RDF, with the goal of maximally preserving its semantics. Based on a semantic interoperability use case (three different ways of encoding the same medical information), we use OWL DL representation and reasoning results as gold standards and compare them with the results of querying two derived RDF representations enhanced by pre-computing of inferred triples by using the RL profile and approximate custom rules.
Keywords: OWL, Ontology, SPARQL, RDF, electronic health records

We have proposed a methodology of enriching data in electronic health records by semantic annotations that account for both ontological (represented by clinical terminologies) and epistemological (represented by health record information models) aspects [1]. The figure depicts three hypothetic forms to record the same statement "suspected heart failure caused by ischaemic heart disease". The content of each form is represented as an instance of the class *Diagnostic statement*. These instances are typed by OWL DL TBox expressions, as shown for the rightmost form (**d3**):



d3 rdf:type '*Diagnostic statement*' and '**is about situation**' only '*Heart failure*'
d3 rdf:type '*Diagnostic statement*' and '**has information attribute**' some '*Suspected* '
d3 rdf:type '*Diagnostic statement*' and '**is about situation**' only ('*Disorder*'
and '**is caused by**' some '*Ischaemic heart disease*')

Anticipating scalability problems of expressive OWL DL T-boxes we approximate the results obtained by OWL DL reasoning by an alternative approach that uses RDF,

SPARQL and materializations, triggered by axioms from the underlying OWL ontology. We present two RDF-based representations: (i) the OWL/RDF serialization of the original OWL DL axioms, and (ii) an RDF representation obtained as a simplification of the OWL DL one, in which the subject and object of the RDF triple correspond to ontology classes and the predicate to an OWL DL expression. For the **d3** instance, which is of the type '*Diagnostic statement*':

<**d3**> <IS ABOUT SITUATION> <**HeartFailure#123**>
<**d3**> <HAS INFORMATION ATTRIBUTE> <**Suspected#3826**>
<**d3**> <IS ABOUT SITUATION> <**Disorder#9981**>
<**Disorder#9981**> <IS CAUSED BY> <**IschaemicHeartDisease#0951**>

Each representation is then checked against a set of SPARQL 1.1 queries, by using Virtuoso 7.0 and the implementation of the OWL 2 RL profile as a set of SPARQL queries [2], as well as custom rules to address the needs of the specific representation. Not all the queries retrieve the three diagnosis instances even after the application of the OWL 2 RL rules. As an example, a query asking explicitly about heart failure diagnoses does not retrieve the instance of the leftmost form, where its type is described by using the axiom:

d1 rdf:type '*Diagnostic statement*'
and '**is about situation**' only ('*Organ failure*' and '**is included in**' some '*Heart*')

A DL reasoner is able to infer equivalence between '*Heart failure*' and the expressions '*Organ failure*' and **isIncludedIn** some *Heart.* However, in RDF it would be required to create a specific rule that each time that there is a triple ('Organ Failure' 'included in' Heart), then another one including 'Heart Failure' would be also created.

The drawback of the representation (i) is its complexity due to numerous intermediate blank nodes. The main drawback of (ii) is the need of keeping the track of the instances when generating the triples.

Although the OWL 2 RL captures a substantial fragment of OWL 2, it restricts expressiveness and is not able to capture existentially quantified expressions [3]. This entails that not all queries retrieve the desired data.

Our work presented here is preliminary and based on a concrete pattern of a semantic interoperability use case that is highly relevant for health care data. Further work will address more use cases and explore other ways of approximating DL reasoning, taking into account recent developments of triple stores.

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

References

1. SemanticHealthNet. http://www.semantichealthnet.eu/ (accessed September 2014).
2. W3C SPIN. http://www.w3.org/Submission/spin-overview/ (accessed September 2014).
3. Zhou, Y. et al. (2013). Making the most of your triple store: query answering in OWL 2 using an RL reasoner. *Proc. of the 22nd international WWW conference ,* 1569-1580.