UMLS-OWL: an OWL 2 translation of the Unified Medical Language System (UMLS®) Semantic-Network and Metathesarus for publishing in the semantic web

Michel Héon[1], Joseph Aubut[2], Sheila Gaudreau[2]

[1] Cotechnoe, Lachute (QC), Canada
heon@cotechnoe.com
[2] Plexus Med, Trois-Rivières (QC), Canada
Joseph.aubut@plexusmed.com
Sheila.gaudreau@plexusmed.com

Abstract

Background: The Unified Medical Language System is a formal ontology of the biomedical and health domain. Its goal is to offer a vocabulary that would facilitate the exchange of information between informatics systems. The semantics and basic facts of the UMLS are contained in the Semantic Network of Metathesaurus. The objective of this research is to translate their content into a OWL 2 Ontology that can be interrogated and inferenced through a SPARQL Endpoint.

Methods: The translated model of the UMLS towards the UMLS-OWL was done through the classification of the conceptual entities, the atoms and the relationships of the UMLS into entities of the description logic T-BOX, R-BOX, and A-BOX.

Results: The experimentation will validate the interoperability of the transformation by using tools of the semantic web to consult the UMLS-OWL.

Keywords: UMLS, UMLS-OWL, UMLS on semantic web, Ontology transformation, Biomedical ontology, Biomedical linked open data

Introduction

Context: Created in 1986 by the U.S. National Library of Medicine, the Unified Medical Language System (UMLS®)[1] is a database that regroups numerous vocabularies and biomedical health norms so as the facilitate interoperability between informatics systems. The UMLS is used to ameliorate or develop applications such as Electronic Health Record's classification tools, dictionaries and their translations into many languages. The UMLS is a formal ontology, but is not an ontology of the semantic web. To increase the interoperability of this vocabulary towards a use by the agents of the web, it is necessary to express the UMLS in a semantic web language such as OWL.

Goal: The goal of this research is to translate the entire UMLS's Semantic Network and Metathesarus in a relational representational model towards a representation form of OWL 2. To do this, we are relying on the hypothesis that the translation is possible since the UMLS is structured with an ontological architecture.

Related work. Many past works [2-5] suggested methods of translation of the UMLS in ontologies of the semantic web by addressing the formalization of the Semantic Network of the UMLS either so as to emphasize the semi-formal character of the UMLS [2], either to use the description logic (DL) so as to reveal inconsistencies between statements originating from different data sources [3] or to evaluate the capacity of an OWL-UMLS to interconnect different agents to UMLS [4], and finally, to formalize the semantic network module of the UMLS so that it serves as a basic vocabulary to an upper-ontology [5].

1 Methodology

Procedure: Fig. 1 presents the process divided in three steps, from the transformation of the semantic network and of the ontological UMLS Metathesarus in text format to the OWL 2 syntax. The first step consists in using MetamorphoSys, the installation wizard suggested by UMLS, to convert the text format UMLS to a UMLS in a MySQL Database. The second step aims to produce an UMLS-OWL by the translation of the UMLS relational representation to an OWL representation according to the architecture of the mapping in Fig. 2 with a combination of Java Persistence API and Apache Jena OWL-API[1] technologies. Third and last step consists in storing the UMLS-OWL in a triplestore and to making it available on the web through a Linked Open Data (LOD)[2] server so that it can be accessed by a SPARQL-endpoint.

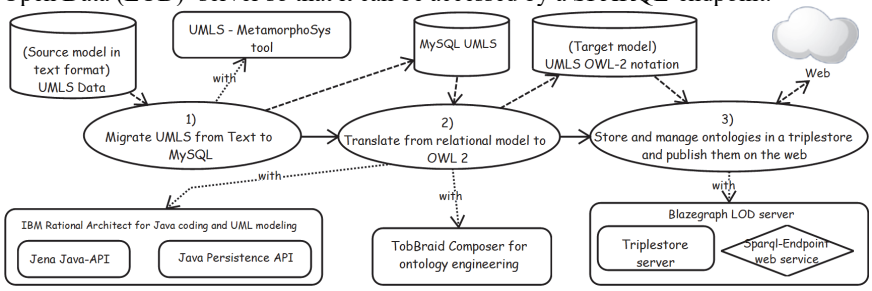

Fig. 1. Transformation process of UMLS ontology from relational models in text mode to an ontological model in OWL 2

UMLS primitive classification in the DL architecture: Fig. 2 introduces the repartition of the UMLS primitives in the architecture of the DL architecture who they contain four levels of abstraction:

1. *Metadata level*: it encompasses all the primitives associated with all table names and columns of the different tables of the UMLS

[1] https://jena.apache.org/

[2] Blazegraph (https://www.blazegraph.com/) in our case

2. *Semantic level*: it includes the TUI associated to the typological referencing
3. *Conceptual level*: it consists of the conceptual identities (CUI)
4. *Factual level*: it contains the different atomic entities (AUI) and roles (RUI)

Each primitive of the UMLS is classified according to the architecture of the DL that is divided in 2 abstraction levels: the *DL abstract level* that includes the T-BOX and R-BOX primitives and the *DL factual level* that stores the elements of the A-BOX.

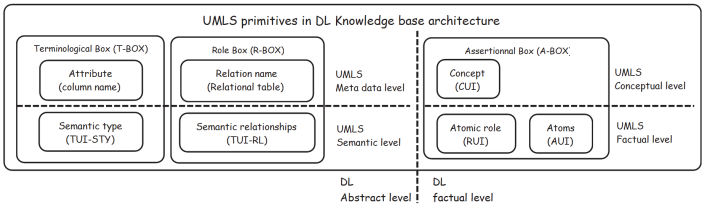

Fig. 2. The primitives' repartition of the UMLS within the architecture of the DL

2 Result

The experimentation that was performed consisted in validating the usage of the UMLS-OWL as a fitting vocabulary for the semantic use of a database. It was also performed to validate the diffusion of the database on a LOD server. Fig. 3 presents an overview of the programed scenario usin g the UMLS-OWL vocabulary. In the left column, a taxonomic view the UMLS semantic networking is presented. In the right part of Fig. 3, in G-OWL [6] notation, is presenting the association `has_manifestion` of 5 different illnesses (names in bold) with the `Findings` and `Signs or Symptoms` .

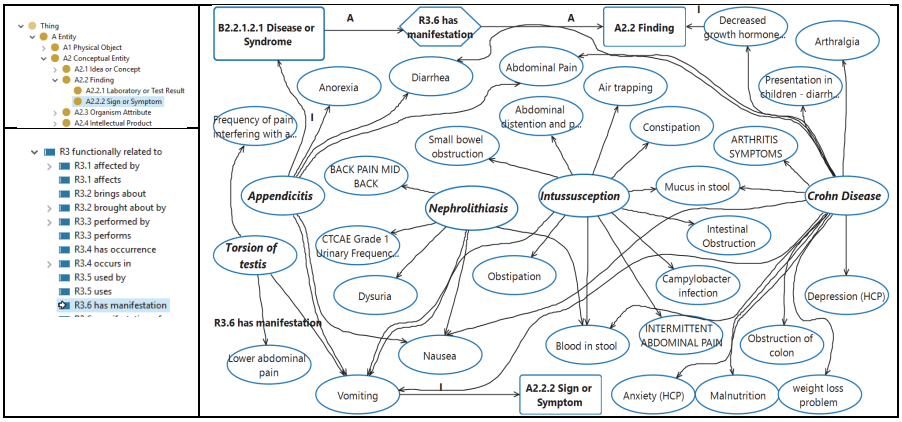

Fig. 3. Semantic level of UMLS-OWL and factual level of a specific scenario

In order to demonstrate the interoperability of the UMLS-OWL, the above scenario has been put online in the LOD [3] server and interrogated by RelFinder[7][4] (see Fig 4).

[3] http://iswc2017.plexusld.com:8080/bigdata/#query
[4] http://iswc2017.plexusld.com/relfinder

Fig. 4. Evaluation scenario presented by the LOD with RelFinder

3 Conclusion

This article presents UMLS-OWL, a version of the UMLS Semantic Network and the Metathesaurus for the semantic web and the LOD server. Each UMLS primitive is classified in the T-BOX, R-BOX or A-BOX according to the DL and the semantic level which they correspond to. The UMLS-OWL's ontological architecture's viability has been proved by the successful use of semantic web tools, its interoperability using a LOD server. The automatization of the process of construction and consolidation of the database is being currently worked upon using a Natural Language Processor in order to identify and classify texts of the Metathesaurus according to the UMLS-OWL's semantic. We plan to use this work to develop a knowledge base that will be used for the creation of a diagnostic assistant application.

Reference

1. https://www.nlm.nih.gov/research/umls/
2. Kashyap, V., and Borgida, A.: 'Representing the UMLS® Semantic Network Using OWL', in Fensel, D., Sycara, K., and Mylopoulos, J. (Eds.): 'The Semantic Web - ISWC 2003: Second International Semantic Web Conference, Sanibel Island, FL, USA
3. Jiménez-Ruiz, E *et al.*: 'Logic-based assessment of the compatibility of UMLS ontology sources', Journal of Biomedical Semantics, 2011
4. Bhatt, M. *et al.* 'Ontology driven semantic profiling and retrieval in medical information systems', Web Semantics: Science, Services and Agents on the World Wide Web, 2009,
5. Schulz, S *et al.* 'BioTopLite: An Upper Level Ontology for the Life Sciences Evolution, Design and Application', in Editor (Ed.): 'Book BioTopLite: An Upper Level Ontology for the Life Sciences Evolution, Design and Application'
6. Héon, M *et al.*: 'Toward G-OWL: A Graphical, Polymorphic And Typed Syntax For Building Formal OWL 2 Ontologies'. Proceedings of the 25th International Conference Companion on World Wide Web, Montréal, Québec, Canada 2016
7. Heim, P *et al*., T.: 'Interactive Relationship Discovery via the Semantic Web', in Aroyo, L., Antoniou, G., Hyvönen, E., ten Teije, A., Stuckenschmidt, H., Cabral, L., and Tudorache, T. (Eds.): 'The Semantic Web: Research and Applications: 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 –June 3, 2010, Proceedings, Part I' (Springer Berlin Heidelberg, 2010), pp. 303-317