e-Health – For Continuity of Care
C. Lovis et al. (Eds.)

doi:10.3233/978-1-61499-432-9-1018

Ontology patterns-based transformation of clinical information

María del Carmen LEGAZ-GARCÍA[a1], Catalina MARTÍNEZ-COSTA[b], José Antonio MIÑARRO-GIMÉNEZ[c], Jesualdo Tomás FERNÁNDEZ-BREIS[a2], Stefan SCHULZ[b], Marcos MENÁRGUEZ-TORTOSA[a]
[a] *Department of Informatics and Systems, University of Murcia, Spain*
[b] *Institute of Medical Informatics, Statistics, and Documentation, Medical University of Graz*
[c] *Medical Expert and Knowledge-Based Systems, Medical University of Vienna, Austria*

Abstract. The semantic interoperability of clinical information requires methods able to transform heterogeneous data sources from both technological and structural perspectives, into representations that facilitate the sharing of meaning. The SemanticHealthNet (SHN) project proposes using semantic content patterns for representing clinical information based on a model of meaning, preventing users from a deep knowledge on ontology and description logics formalism. In this work we propose a flexible transformation method that uses semantic content patterns to guide the mapping between the source data and a target domain ontology. As use case we show how one of the semantic content patterns proposed in SHN can be used to transform heterogeneous data about medication administration.

Keywords. Ontology, Ontology patterns, semantic interoperability

Introduction

Semantic interoperability of clinical information has become a major challenge in the medical informatics community [1]. The current state of the art shows that clinical data sources are heterogeneous in two ways: (1) different technologies are used to represent the data (i.e., relational databases, XML, etc.); (2) different types of data structures and terminologies are used within technologically homogenous environments. Both issues have to be addressed, since they are an impediment to reach the semantic interoperability of clinical information.

The EU network SemanticHealthNet (SHN, http://www.semantichealthnet.eu/) aims to improve semantic interoperability and proposes the representation of the meaning of clinical data by ontologies and semantic content patterns. Such patterns aim at assisting in information modelling, preventing users from fully understanding the underlying, complex, ontological expressions. Additionally, SHN proposes their formal representation based on description logics to detect semantic equivalences across

[1] Corresponding Author.
[2] Corresponding Author.

content expressed in a variety of information models and terminologies. Traditionally, popular formats like XML or relational databases have limitations to fully represent the meaning of clinical data, which complicates their exploitation in semantic interoperability settings. Given that clinical data is mainly available in the traditional formats (i.e., XML, relational databases), approaches and tools that allow their semantic representation are needed.

Most of existing approaches and tools [2-4] perform a generic transformation of the data, without taking into account their underlying model of meaning. Transformation methods are usually based on defining the mappings between the corresponding schemas. In relational databases one might propose to transform tables into ontology classes, records into individuals, and columns into properties. Such a transformation is not more than a change of format, because the real meaning of the entities represented is completely ignored in such a process, which is therefore not enough to achieve semantic interoperability.

In previous work, we have proposed an ontology-driven method for transforming relational and archetype-based data into semantic representations [5]. Our experience in [6] revealed that the definition and execution of the mappings required a significant repetitive effort, that simple mappings between source and target entities (i.e., database table to ontology class) were not enough, and that the reuse of those mappings was not easy. In this paper, we describe a more powerful approach for the definition and execution of the mappings, based on the SHN proposal, which uses semantic content patterns as proxy for representing clinical information based on a formal model of meaning. Our use case will be one pattern created within SHN to transform heterogeneous data about medication administration [7].

1. Methods

Our methodological approach is illustrated in Figure 1. The definition of the mapping rules is the central part of the approach. The rules permit to associate entities of the source data schema with ontology classes and properties, and they define how the source data are transformed into OWL ontology individuals. OWL individuals have a set of axioms about class membership, property values and individual identity associated, and such axioms are generated in the transformation process.

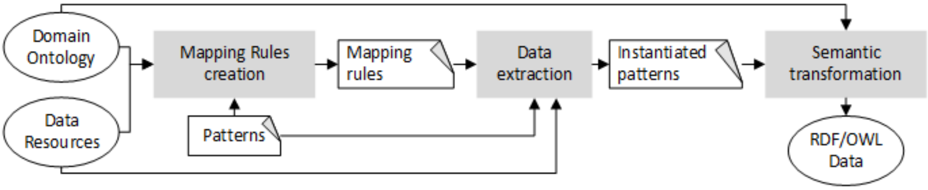

Figure 1. Overview of the methodological approach

Besides, we propose defining more complex mapping rules by combining some basic ones, which we call semantic patterns. Such patterns define the template for the creation of OWL individuals with specific types of axioms. A pattern is created from an ontological representation of a domain, which includes formal descriptions of types of entities and their relations. We use them to obtain an abstraction of the whole domain, including, on the one hand (types of) entities of the domain proper, such as drugs, disorders, procedures etc., as well as information entities (informational artefacts like record entries) on the other hand. Table 1 shows an excerpt of the *Medication Administration Pattern* defined within SHN, represented as Subject-Predicate-Object (SPO) triples. It gives the complete definition of a medication administration according to the ontological framework defined in SHN. It allows drug prescription either by product or by active ingredient.

Table 1. Excerpt of the representation of the SHN *Medication Administration Pattern*.

Subject	Predicate	Object
btl:Plan	isRealizedBy	sct:MedicationAdministration
sct:MedicationAdministration	hasFocusOn	sct:PharmaceuticalProduct
sct:PharmaceuticalProduct	hasComponent	sct:Substance
sct:MedicationAdministration	hasStartTime	btl:PointInTime
sct:MedicationAdministration	hasEndTime	btl:PointInTime
sct:MedicationAdministration	hasDuration	btl:Duration

We represent the patterns using the Ontology Pre-Processing Language version 2 (OPPL2), which is capable of modifying OWL ontologies and allows the definition of reusable, executable patterns [8]. An excerpt of the OPPL2 pattern for the SHN Medication Administration one is shown in Figure 2. It has two input variables, namely, ?medRecord and ?product, which are, respectively, a medical record and a subclass of PharmaceuticalProduct. This pattern associates the product with the medical record through the ontological axiom shown in the ADD clause. The result of this pattern is that ?medRecord is a *Plan* 'realized by' a *Medication Administration* 'to a patient' of a product parametrized by the variable ?product. The prefixes btl, sct and shn refer to BioTopLite, SNOMED CT and SHN ontologies respectively. Finally, the mapping rules defined between the pattern and the data source contain the access routes to the data source elements that will be used to create the ontology instances. Once the data is extracted, we instantiate the templates represented by the patterns with it. These instantiated templates are the new individuals to be created. This will be explained later.

```
?medRecord:INDIVIDUAL, ?product:CLASS[subClassOf sct:PharmaceuticalProduct]
  BEGIN
    ADD ?medRecord Type btl:Plan
          and btl:hasRealization only (sct:MedAdmin and btl:hasPatient some ?product);
  END
```

Figure 2. Example of OPPL2 pattern

2. Results

Next, we illustrate how we use semantic patterns to guide the mapping definition assuming medication administration data captured using two heterogeneous but similar models, the SHN and epSOS (http://www.epsos.eu/) archetypes, whose structures are described in Figure 3. Both models are represented according to OpenEHR, but they have a different structure. For example, in epSOS the focus of the medication order is the drug product, which can be optionally refined by one or more active ingredients. In the SHN model the focus is on active ingredients. As data exemplar we will represent "Aspirin + caffeine" product administration starting on October 10th, 2013. Figure 4 (1) shows an excerpt of the OpenEHR data extract. In order to transform the data rendered according to both SHN and OpenEHR into their semantic representation (i.e. target ontology), we have represented the medication administration pattern in OPPL (Figure 4 (2)). The pattern has fixed and variable parts. The variables depend on the source data and the OPPL pattern represents them like ?product or ?substance and will be bound to the clinical data source. The scope of these variables is restricted to entities of the target ontology.

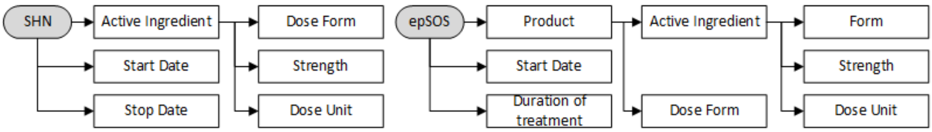

Figure 3. Structure of medication administration models created in SHN and epSOS projects

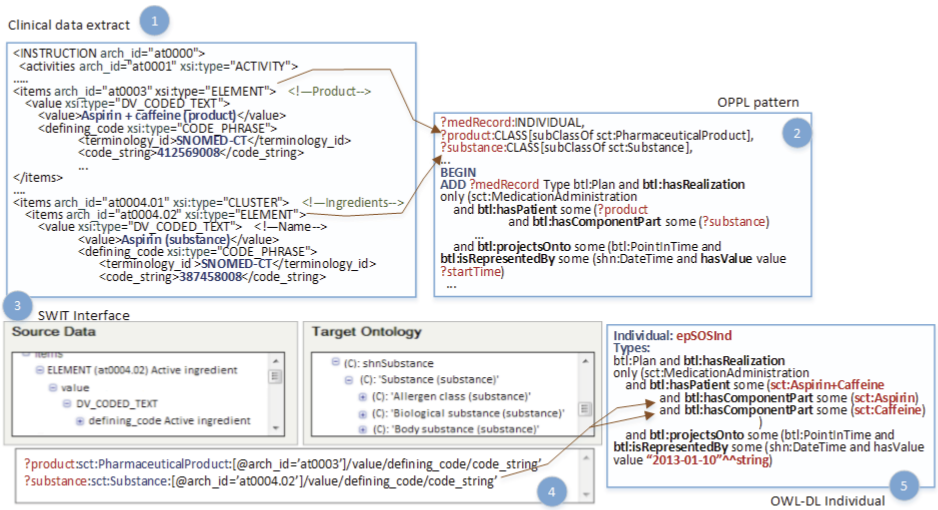

Figure 4. Overview of the mapping process

Figure 4 depicts the process for transforming the data for the epSOS model into the SHN semantic representation. For each relevant entity of the source data schema (3), we define its mapping onto the variables of the pattern. For instance, the variable ?product is mapped onto the ELEMENT at0003, which contains the label and code for 'Aspirin + Caffeine (product)'. The representation of such mapping can be found in (4) and for the case of archetypes the variable has an XPath expression associated. The OWL individual obtained is shown in (5). For the SHN model, the mapping is partially different, since the SHN model does not represent product, only substances. When a variable cannot be mapped, as is the case of ?product, such variable is bound to sct:PharmaceuticalProduct.

We have developed the SWIT tool (http://sele.inf.um.es/swit), which implements this methodology for the mapping and transformation of clinical data, and provides users with a step by step process to perform the final transformation and allows reusing the mappings defined. Figure 4, see (3) and (4), provides a partial view of the interface for defining the mappings.

3. Discussion and Conclusions

We have presented a method, which allows the semantic transformation of data rendered according to representations with implicit semantics, and it is based on mappings between the source data and a target ontology representing the domain knowledge. Here, we have enhanced our approach with semantic patterns that provide a friendlier view on the ontology by avoiding the handling of complicated representational formalisms at the target side, thus guiding and facilitating the mapping process. We have used the semantic content patterns developed within SHN to facilitate the representation of clinical data based on formal principles and improve semantic interoperability across heterogeneous data representations. Given that the transformation process is able to create content in OWL format, that is, OWL axioms, a constructive language like OPPL2 is appropriate for this task, because: (1) it has enough expressivity; (2) its API reduces the implementation and execution effort (the patterns can be reused in the tool); (3) it permits exploiting inferencing and automated reasoning in the transformation process; (4) the patterns are directly connected to the underlying ontology.

Although as a proof of concept, in this work the use of the medication administration pattern as guide simplifies the mapping process. The effectiveness of such patterns to guide the mapping, will depend on how close-to-user their representations are. However, their building and maintenance should be ideally supported by a community of experts. Further work would be required to ensure the quality of manually defined mappings. In the case of archetype-based data sources, the availability of quality terminology bindings could facilitate the definition of the mappings and developing quality assurance methods.

Acknowledgements

This work has been funded by the Spanish Ministry of Science and Innovation and the FEDER programme through grant TIN2010-21388-C02-02, the Fundación Séneca through grants 15555/FPI/2010 (MCLG) and 15295/PI/10, and the *SemanticHealthNet Network of Excellence* within the EU 7th Framework Program, Call:FP7-ICT-2011-7, agreement no.:288408.

References

[1] Stroetmann VN, et al. Semantic Interoperability for better health and safer healthcare. Deployment and Research Roadmap for Europe. 2009. http://www.empirica.com/publikationen/documents/2009/semantic-health-report.pdf (accessed Feb. 2013)

[2] D2RQ. Accessing Relational Databases as Virtual RDF Graphs. http://d2rq.org/ (accessed Feb. 2013)

[3] Triplify. http://triplify.org/ (accessed Feb. 2013)

[4] Mapping Relational Data to RDF in Virtuoso. http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSSQLRDF (accessed Feb. 2013)

[5] Miñarro-Giménez JA, Fernández-Breis JT. Ontology-Driven Method for Integrating Biomedical Repositories. *Advances in Artificial Intelligence*. Springer Berlin Heidelberg; 2011. 473–82

[6] Fernández-Breis JT, et al. Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts. *J Am Med Inform Assoc*. 20(2013), 288–296.

[7] SemanticHealthNet deliverable 4.2: Ontology/Information models covering the HF use case. 2013.

[8] Egaña M, et al. Applying Ontology Design Patterns in Bio-ontologies. *Proceedings of the 16th International Conference on Knowledge Engineering: Practice and Patterns*. Berlin, Heidelberg: Springer-Verlag; 2008. 7–16.