Semion: a smart triplification tool

Andrea G. Nuzzolese
Semantic Technology Lab,
ISTC-CNR
Via Nomentana 56
Rome, Italy
nuzzoles@cs.unibo.it

Aldo Gangemi
Semantic Technology Lab,
ISTC-CNR
Via Nomentana 56
Rome, Italy
aldo.gangemi@cnr.it

Valentina Presutti
Semantic Technology Lab,
ISTC-CNR
Via Nomentana 56
Rome, Italy
valentina.presutti@cnr.it

Paolo Ciancarini
Dept. of Computer Science,
Università di Bologna
Mura Anteo Zamboni 7
Bologna, Italy
cianca@cs.unibo.it

ABSTRACT

The Web of Data is fed by “triplifiers”, tools able to transform content (often databases) to linked data. Triplifiers implement different methods and typically are based on bulk recipes which allow for no or limited customization of the process. Furthermore, their consumption or refactoring is often difficult due to mismatches between the semantics embedded in original structures, and the RDF or OWL semantics obtained thorugh the recipes. Semion is a method and a tool for customizing and expliciting the semantics of data reengineering and refactoring.

1. INTRODUCTION

Commonly accepted solutions for tranforming non-RDF data sources into RDF are based on ad hoc semantics-driven approaches, that make implicit assumptions on the domain semantics of the non-RDF data source (e.g. a relational database is trasformed mapping a table into an *rdfs:Class*, a table column into an *rdf:Property* and a table record into an instance of the specific RDF table class). The tool described here, Semion, implements a method that firstly makes no semantic assumption at the domain level, and just transforms the data source into RDF triples driven by an OWL description of the data source structure (a source meta-model), which can be defined and customized by the user. Secondly, the RDF triples can be modelled by aligning them to any additional RDF or OWL ontology, which acts as either a meta-level “mediator” to the required semantics (e.g. SKOS [4] or the OWL metamodel [3, 1]), or as a reference domain ontology (e.g. DOLCE, FOAF, or the Gene Ontology). In particular, we exemplify the alignment of triplified data with the Linguistic Meta-Model (LMM) [5], an OWL-DL ontology that formalizes the distinctions of the semiotics.

2. METHOD

The method implemented in Semion is based on an approach that substantially divides the reengineering process from the modelling one. The reengineering process performs the semantic lifting just extracting RDF triples driven by the OWL description of the structure of the datasource provided as input. On the other hand, the modelling process allows to introduce semantics in the extracted data set, by using a semiotic-cognitive approach based on the Linguistic-Meta Model (LMM) [5]. The most important feature of LMM is its ability to support the representation of different knowledge sources developed according to different underlying semiotic theories [5]. Figure 1 shows the basic key concepts that

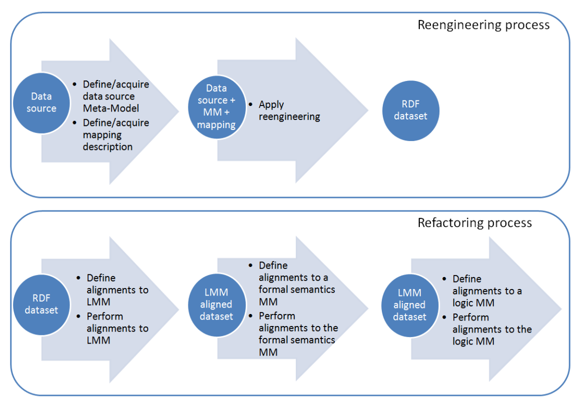

Figure 1: Tranforming method: key concepts.

are behind our transforming method. The “Data source” bubble represents the input consisting of a non-RDF data source that is reengineered into an RDF data set according to its type, to its structure described by an OWL metamodel and to a defined mapping. The RDF dataset is then refactored (“Refactoring process” frame in figure 1) to the LMM framework according to specific customizable alignment rules. Once the RDF dataset is aligned to LMM it is possible to grounds it to a formal semantics and finally to express its logics.

3. TOOL

The method described in the previous section is implemented in Semion. Currently the tool is still a prototype and has been tested only for transforming relational databases, but it was designed to perform the transormation of any kind of data source. The figure 2 shows the reengineering

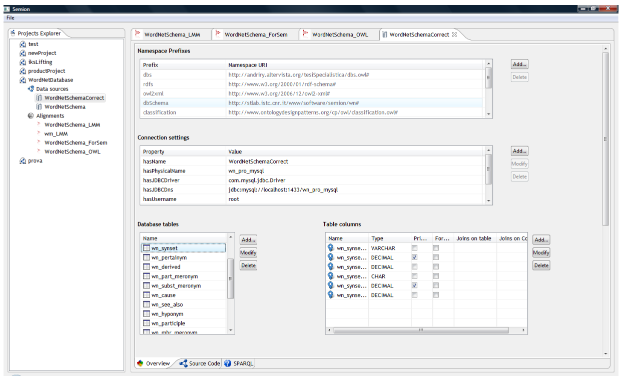

Figure 2: Semion tool: view of the reengineering interface.

interface of the Semion tool. It helps the user to define the schema of the database structure that is described by using the meta-model provided for the structure of the database itself. Both because the database could be large and because the user could not know exactly how the database was designed, the tool provides a wizard interface that automatically extracts the RDF of a database's schema. Once the RDF of the database's schema is available, the interface allows the user to transorm the data from the database to the RDF format. Before performing data extraction from the database it is also possible to correct possible issues derived from a bad design or a bad mantainance of the database. In fact, the tool provides functionalities to set in the RDF of the database's schema primary and foreing keys and eventually related relations. The refactoring interface allows the user to

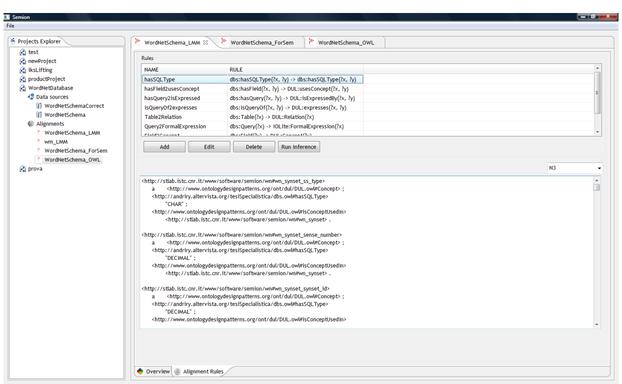

Figure 3: Semion tool: view of the refactorer interface.

align the dataset to specific ontologies for adding semantics to data. The alignment ontologies can by chosen following the method that Semion implements i.e. first the dataset is aligned to the LMM framework, then to an ontology that contains the distinctions of the formal semantics and finally to an ontology that contains the logics. Semion performs ontology alignments through SPARQL CONSTRUCT, that are obtained from the rules written in a human-readable syntax (see figure 3), that are based on the form:

$$antecedent \rightarrow consequent$$

Using this syntax, a rule e.g. asserting that being an instance of class *Table* in dataset meta-model implies to be a *Concept* of DOLCE [2] would be written:

$$dbs : Table(?x) \rightarrow DUL : Concept(?x)$$

This rule will be interpreted as the SPARQL query:

```
CONSTRUCT { ?x rdf:type DUL:Concept. }
WHERE { ?x rdf:type dbs:Table. }
```

With the same syntax can be written, through the Semion tool, rules for transforming LMM to the FormalSemantics vocabulary. The rules could be

IOLite:FormalExpression(?x) → FormalSemantics:Query(?x)
DUL:Relation(?x) → FormalSemantics:Class(?x)

Rules for aligning the FormalSemantics vocabulary to OWL can be written as the following

FormalSemantics:isSubsumedBy(?x, ?y) → rdfs:subClassOf(?x, ?y)
FormalSemantics:Class(?x) → owl:class(?x)

The Semion tool can be downloaded from the following URL `http://stlab.istc.cnr.it/software/semion/tool`

4. REFERENCES

[1] B. Cuenca-Grau and B. Motik. OWL 2 Web Ontology Language: Model-Theoretic Semantics. `http://www.w3.org/TR/2008/WD-owl2-semantics-20080411/`, 2008. Visited on April 2010.

[2] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening Ontologies with DOLCE. In *Proceedings of 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, volume 2473 of Lecture Notes in Computer Science, page 166 ff, Sigünza, Spain, Oct. 1–4 2002.

[3] F. v. Harmelen and D. L. McGuinness. OWL Web Ontology Language Overview. W3C recommendation, W3C, Feb. 2004. http://www.w3.org/TR/2004/REC-owl-features-20040210/.

[4] A. Miles and S. Bechhofer. SKOS Simple Knowledge Organization System Reference. W3C working draft, W3C, June 2008. http://www.w3.org/TR/2008/WD-skos-reference-20080609/.

[5] D. Picca, A. M. Gliozzo, and A. Gangemi. LMM: an OWL-DL MetaModel to Represent Heterogeneous Lexical Knowledge. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May 2008. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/.