GRAPH-BASED RULES FOR XML DATA CONVERSION TO OWL ONTOLOGY

Christophe Cruz and Christophe Nicolle

Laboratory Le2i, UMR-5158 CNRS, Université de Bourgogne, B.P. 47870, 21078, Dijon Cedex, France

Keywords: Ontology population, Ontology enrichment, OWL ontology, XML data, RDF, Semantic annotation.

Abstract: The paper presents a flexible method to enrich and populate an existing OWL ontology from XML data based on graph-based rules. These rules are defined in order to populate automatically a new version of an OWL ontology. Today, most of data exchanged between information systems is done with the help of the XML document. Leading researches in the domain of database systems are moving to semantic model in order to store data and its semantics definition. This flexible method consists in populating an existing OWL ontology from XML data. In paper we present such a method based on the definition of a graph which represents rules that drive the populating process.

1 INTRODUCTION

Ontologies are aimed at representing knowledge about a specific domain that are understandable by both developers and computers. For this, ontologies enumerate concepts and relations between concepts (Guarino, 1995) and define properties, functions, constraints and axioms (Studer, 1998). The major issues in ontology development include ontology representation, ontology acquisition, evaluation and ontology maintenance (zhou, 2007). Ontology representation is the main issue in ontology development because its representation has to be understandable by computers and humans. Consequently, an ontology representation language should provide representation adequacy for humans and inference efficiency for computers. Ontology dialects based on description logic (DL) provide a frame-based knowledge representation and profit from the expressiveness of DL reasoning systems. Ontology acquisition refers to the process of the ontology creation such as concepts, relations, individuals and axioms. From an empirical point of view, there are two kinds of ontology modeling processes. The first one is the ontology modeling, which is traditionally carried out by knowledge engineers or domain experts. Actually, these ontologies are built by humans for humans. The second one is in fact the point of view of the semantic Web according to which ontologies are built automatically by computers for computers

within sources such as dictionaries, Web documents and database schemas. It has to be noticed that the resulting ontologies are still understandable by humans. As a result, ontology acquisition can benefit significantly from ontology learning (Ding, 2002). Ontology evaluation aims at enhancing the quality of ontologies in order to improve the interoperability among systems and to increase the adoption of ontologies. Ontologies can be evaluated in different ways (Staab, 2004) using measures such as completeness, consistency and correctness (Gomez, 1995). Ontology maintenance concerns the organization, the search and the update process on existing ontologies. The constant evolution of the environment of ontologies makes it very important for ontologies to be evaluated and maintained (Sure, 2002) in order to keep up with the change.

To reach this goal, this article presents an automatic population process from XML data to OWL ontologies, a process which is based on a manual mapping between the XML schema elements and the OWL schema elements. If the OWL schema does not contain the required elements then the ontology has to be enriched by the system manager. The ontology enrichment is the activity of extending an ontology by adding new elements (e.g. concepts, relations, properties, axioms) (Castano, 2007). Our enrichment process consists in annotating knowledge which is contained in XML schemas in order to define the ontology schema (Faatz, 2004). Some automatic processes from
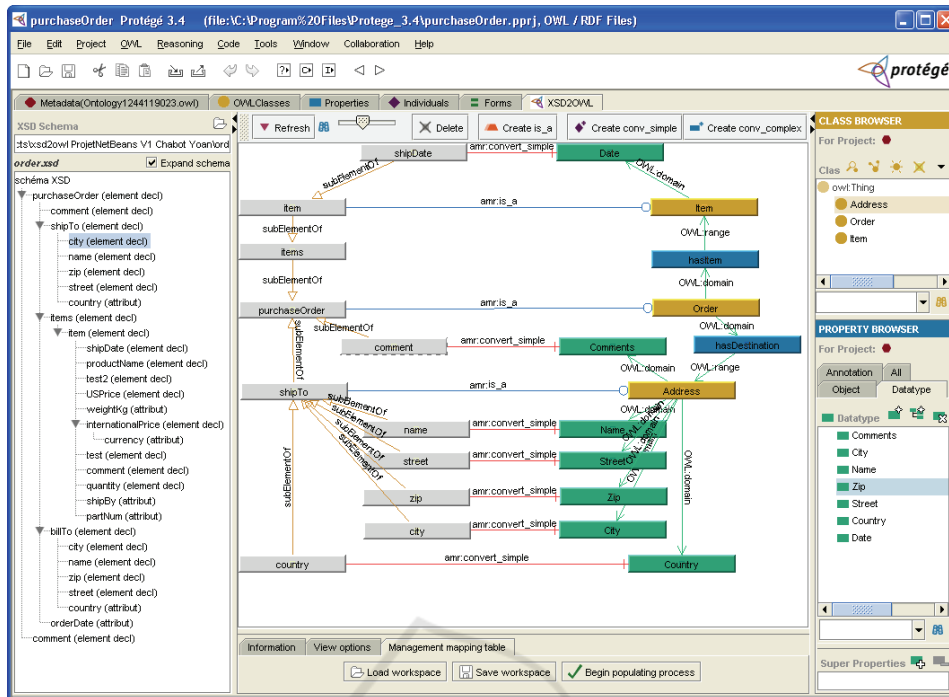
Figure 1: Snapshot of the XSD2OWL plug-in.

ontology learning can be used but this point is beyond the scope of this paper. The ontology population is the activity of adding new instances or individuals to an ontology (Castano, 2007).

2 THE XSD2OWL TOOL

The principle of our solution (Matthias, 2004), (Bohing, 2005) consists in annotating and linking the semantic level (OWL schema) and the schematic level (XML schema). The graphical interface used to realize this is incorporated in the tool "protégé" from Stanford as a plug-in (e.g. fig 1) in order to populate an existing ontology. Once the graph of mapping rules has been defined, the population process is automatic. The user has only to select a list of XML documents which can be validated by the XSD schema.

2.1 The Graph-based Rule Definition

The process of annotating consists in defining "Basic Mapping Rules" (BMR) which appear in the graph as nodes or boxes. These boxes represent annotations on the XSD and OWL documents. Some of the annotations are defined on the XSD schema and are represented by grey boxes. ("xsd:element", xsd:attribute"). The other boxes are annotations on

the OWL-DL schema ("owl:Class", "owl:ObjectProperty", "owl:DatatypeProperty"). The color of these boxes follows the colors defined in the application "protégé" (orange for "Concept", blue for "ObjectProperties" and green for "DatatypeProperties", e.g. figure 1).

The links between XSD annotations are "subElement" relationships which are added automatically by the process because these relationships already exist in the XSD schema. In addition, links between OWL annotations are also added automatically because these relationships already exist in the OWL ontology.

The process that consists in defining links between annotations of the XSD schema and annotations of the OWL-DL is called "Advanced Mapping Rules". These rules which are represented graphically are added manually by the user.

An RDF document is generated from the defined rules. This document is used to store all information required during the population process.

The objective of figure 2 is to describe the relationships between the components of the RDF rules. They are composed of BMR on XML schemas and BMR on OWL schemas. These BMR are used to identify elements required for the mapping process. Advanced Mapping Rules are defined in order to allow the conversion of data from XML data to OWL instances.
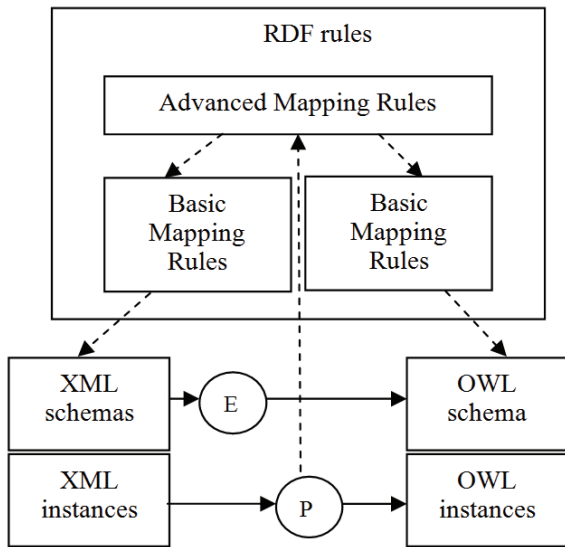
Figure 2: Relations between processes and RDF rules.

2.2 Population Process

Concerning the population process more than ten cases of use have been identified, but due to a lack of space only few of them can be presented.

Case 1. Population of an isolated concept from an XSD element

Figure 3: Use case 1.

In this graph the grey box "purchaseOrder" contains an annotation of the XSD element "purchaseOrder". In addition the orange box "Order" contains an annotation to the concept "Order" in the OWL schema. In order to populate the concept "Order" from the "purchaseOrder" element data, a link "amr:is_a" between both boxes is defined. This link is an advanced mapping rule.

Case 2. Population of a concept associated to n "DatatypeProperties" (n not null) from an XSD complex element that contains m sub-elements simple and (n-m) attributes.

The population of the ontology is an automatic process based on the mapping graphs. To realize this process, we have defined an algorithm that takes into account the type "bmr:id" in order to avoid duplicated instances of the ontology. First, it determines all classes that have to be populated. Secondly, all "datatypeProperties of each concept are provided to the instances. In the example given in this paper no references are made to the

management of restrictions on the properties. Some rules can be defined in order to specify which constraints have to be verified. If these rules are not defined then the restrictions are not checked. The limitation of our solution becomes apparent by the fact that we do not generate an XSL document in order to enrich and populate the ontology. However, the process is complex enough so that, for the moment, it is not relevant to add the generation of an XSL document for the population process.
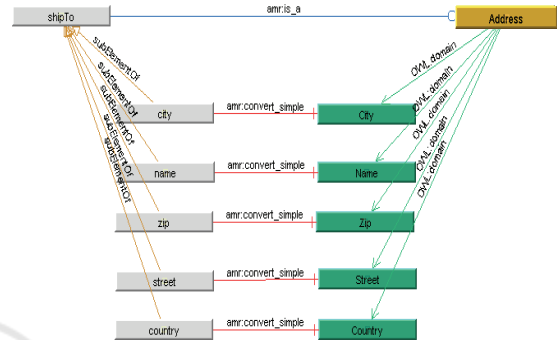

Figure 4: Use case 2.

3 CONCLUSIONS

This paper presents a flexible method to enrich and populate an OWL ontology for the integration of XML data. Basic mapping rules and advanced mapping rules are defined by users and can be reused for other conversions and populations of ontologies. This conversion is the first part of our work. The second part consists in improving the process and in making some suggestions in order to facilitate the mapping to the user. The RDF rules can be used for the automatic extraction of certain elements of the XML schemas that can be converted in order to help users during the mapping. For instance, a string that contains a date can be detected automatically to guide the user during the conversion.

According to (Cruz, 2004), (Klein, 2002), (Lakshmannan, 2003), (Cruz, 2006), data integration can be undertaken by defining rules of mapping between information sources and the ontological level. These rules consist in adding a semantic layer to source elements. They thus provide these elements with semantic definition with regard to a consensual definition of the meaning. For that purpose, ontologies are useful in order to define a common semantic. Furthermore, schema matching is a well studied field that allows to find out automatically identical resources in the different

schemas. Schema matching is a manipulation process on schemas that takes two heterogeneous schemas as input and produces as output a set of mapping rules that identifies relations between the elements of the two schemas (Huynh, 2008). This is required in many database applications, such as integration of web data sources, data warehouse loading and XML message mapping. As a future work, we would like to focus to an automatic process by reusing a set of previous RDF rules. In fact, it consists in reusing the mapping knowledge capitalized during different mapping processes. In addition the concatenation rules and the regular expression rules are being prototyped. This implies that new boxes have to be defined and will be connected to XSD boxes and OWL boxes.

ACKNOWLEDGEMENTS

REFERENCES

Bohring, H.; Auer, S., 2005. *Mapping XML to OWL Ontologies*, Leipziger Informatik-Tage (LIT 2005), Sep. 21-23, 2005, Lecture Notes in Informatics

Castano, S., Espinosa, S., Ferrara, A., Karkaletsis, V., Kaya, a., Melzer, S., Moller, R., Montanelli S., Petasis, G., 2007. *Ontology Dynamics with Multimedia Information: The BOEMIE Evolution Methodology*, International Workshop on Ontology Dynamics (IWOD) ESWC 2007 Workshop, Innsbruck, Austria

Cruz, C., Nicolle, C., 2006. *Ontology-Based Integration of XML data*, Webist, Setubal, Portugal, pp. 30-37

Cruz, I. F., Xiao, H., Hsu, F., 2004. *An Ontology-based Framework for Semantic Interoperability between XML Sources*, In Eighth International Database Engineering & Applications Symposium (IDEAS)

Ding, Y., 2002. *Ontology research and development part1 – A review of ontology generation*, Journal of Information Science, 28, 123–136

Faatz, A., and Steinmetz, R., 2004. *Precision and recall for ontology enrichment*. ECAI-2004 Workshop on Ontology Learning and Population, Valencia, Spain, Aug.

Gomez-Perez, A., 1995. *Some ideas and examples to evaluate ontologies*, Artificial Intelligence for Applications

Guarino, N., 1995. *Formal ontology, conceptual analysis and knowledge representation*, International Journal of Human-Computer Studies 43, 625–640

Huynh Quyet Thang, Vo Sy Nam, 2008. *XML Schema Automatic Matching Solution*, International journal on Information Systems Science and Engineering, vo.l 4, number 1

Klein, M., 2002. *Interpreting XML via an RDF schema*. In ECAI workshop on Semantic Authoring, Annotation & Knowledge Markup (SAAKM 2002), Lyon, France

Lakshmannan, L. V., Sadri, F., 2003, *Interoperability on XML Data*, In Proceeding of the 2nd International Semantic Web Conference.

Matthias Ferdinand and Christian Zirpins and D. Trastour, 2004, *Lifting XML Schema to OWL*, 4th International Conference, ICWE 2004, Munich, Germany, July 26-30, Proceedings, Springer Heidelberg, pp. 354-358

Staab, S., Gomez-Perez, A., Daelemana, W., Reinberger, M.-L. and Noy, N.F., 2004. *Why evaluate ontology technologies? Because it works!*, Intelligent Systems, IEEE 19, 74–8

Studer, R. Benjamins, R. and Fensel, D., Knowledge engineering: Principles and methods, Data and Knowledge Engineering 25, 161–197

Sure, Y., Staab, S. and Studer, R., 2002. *Methodology for development and employment of ontology based knowledge management applications*, SIGMOD Rec 31, 18–34

Zhou, L., 2007, *Ontology learning: state of the art and open issues*, Information Technology and Management archive Volume 8, Issue 3, 241 – 252