A GRAPH-BASED TOOL FOR THE TRANSLATION OF XML DATA TO OWL-DL ONTOLOGIES

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: Today most of the data exchanged between information systems is done with the help of the XML syntax. Unfortunately when these data have to be integrated, the integration becomes difficult because of the semantics' heterogeneity. Consequently, leading researches in the domain of database systems are moving to semantic model in order to store data and its semantics definition. To benefit from these new systems and technologies, and to integrate different data sources, a 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. The graph of rules facilitates the mapping definition that consists in mapping elements from an XSD schema to the elements of the OWL schema.

1 INTRODUCTION

Ontologies are widely used to capture and organize knowledge about a particular domain. In addition, the definition of ontolgies is used as an index to retrieve specific data (Garcìa, 2005), to infer new knowledge (SWRL, 2004), to semantically annotate multimedia data (Castano, 2007), to find out Web Services automatically (Martin, 2007), or to match knowledge with other knowledge for a more general purpose.

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, 1998) 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-Perez, 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.

This article presents an automatic population process from XML data to OWL ontologies, a

Cruz C. and Nicolle C..
A GRAPH-BASED TOOL FOR THE TRANSLATION OF XML DATA TO OWL-DL ONTOLOGIES.
DOI: 10.5220/0003629603610364
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2011), pages 361-364
ISBN: 978-989-8425-80-5

Table 1: Synthesis of related projects.

Paper	*XSLT*	*Automatic XSD mapping*	*Automatic XML instances*	*Multiple XSD integration*	*OWL-DL*	*Mapping in RDF*	*XSD2OWL and XML2RDF*
Ferdinand & al.	no	yes	yes	no	yes	no	yes
García & al.	no	yes	yes	no	no	no	yes
Bohring & al.	yes	yes	yes	no	yes	no	no
Rodrigues & al.	yes	no	yes	yes	yes	no	no
Aninic & al	no	no	no	no	yes	yes	no
Kim & al.	yes	no	no	no	yes	no	yes
Bedini & al	no	yes	no	no	yes	no	no
Cruz & Nicolle	no	no	yes	yes	yes	yes	no

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 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 BACKGROUND

In order to populate an ontology, it is first necessary to define which elements in an XML document will be processed. In addition, it is also necessary to identify the nature of the XML element in order to generate the individuals of the corresponding ontology. The first step consists in defining mapping rules which define the mapping of an XML element to an OWL element. If the required OWL elements do not exist in the ontology then the ontology has to be enriched accordingly. Once the mapping rules have been defined, the second step consists in populating (automatically) the ontology by using XML documents validated by the XML schema.

We extend the solution described in (Rodrigues, 2006) by mapping several XML schemas in an existing OWL ontology (Cruz, 2008), which consists in defining mapping rules between each XML schema to a common existing OWL ontology. In addition, we allow users to define partial mapping of XML schemas in order to enrich and populate a relevant OWL-DL ontology for a specific data management process. Furthermore, we define the transformation rules in RDF that allow a more flexible and a more fine-grained rule definition in order to allow a partial reuse of annotated XML schemas and data type conversions. Our method allows also the definition of advanced rules of transformation in RDF that can be reused for other mappings. The most important point is that rules are represented by a graphical graph which is managed directly by the user. This graph-based rules method facilitates the definition of rules and the corresponding results. Some work has been done specifically on the translation of XML schemas into OWL ontology (Garcìa, 2005), (Do, 2007), (Ferdinand, 2004), (Bohring, 2005), (Rodrigues, 2006), (Anicic, 2007), (Kim, 2007), (Bedini, 2008), (e.g. table 1)

Table 1. summarize all properties of studied projects . "XSLT" means that the method uses an XSL style sheet such as XML data for the conversion process. "Automatic XSD mapping" means that the user cannot intervene in the mapping process between XML schema elements and the OWL ontology. "Automatic XML instances" means that if instances are generated by the studied method then the process is automatic. "Multiple XSD integration" means that the project allows users to integrate several XML schemas to an OWL ontology. "OWL-DL" means that the generated ontology is a description logic ontology which allows inference and consistency checking. "Mapping in RDF" means that the method uses RDF to specify the mapping between schemas. The last column implies that if the value is "yes" then XML schemas are mapped to an OWL ontology and the instances of the XML schemas are translated in an RDF document. It means that instances are not OWL instances. The last row of the table presents the properties of our method. Our method does not use XSLT because the process is too complex to be used with an XSLT processor (e.g. regex). The mapping is done manually but the population of the ontology is automatic. We also allow the user to integrate

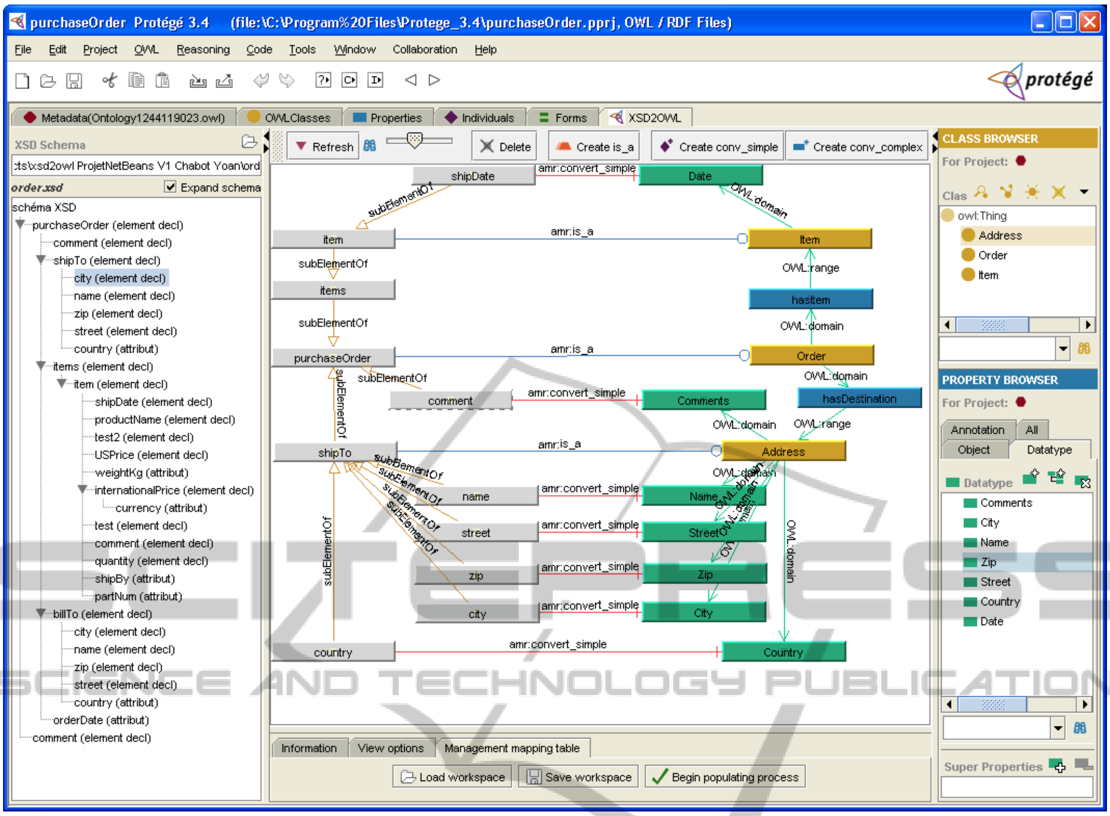

Figure 1: Snapshot of the GXSD2OWL plug-in.

several XML schemas into an existing OWL-DL ontology. In addition, in order to specify the mapping between schemas and the ontology, we use the RDF language in order to permit an advanced management of mapping rules. Finally, the instances of the ontology are obviously defined in the model of the ontology schema.

3 THE GXSD2OWL TOOL

The principle of our solution 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.

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.

4 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.

ACKNOWLEDGEMENTS

Authors would like to thank Yoan Chabot, for its contribution.

REFERENCES

Anicic, N., Ivezic, N. and Marjanovic, Z.: Mapping XML Schema to OWL, *Enterprise Interoperability, Springer London* (2007)

Aumueller, D., Do, H. H., Massmann, S., Rahm, E.: Schema and ontology matching with COMA++, *SIGMOD Conference* (2005)

Bedini, I. Georges Gardarin and Benjamin Nguyen, Janus: Automatic Ontology Construction Tool, *EKAW, 29th September-3rd October* - Acitrezza, Catania, Italy (2008)

Berstel, J., Boasson, L.: XML Grammars, *MFCS* 2000: 182-191 (2000)

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

Bowers S., Delcambre L.: Representing and Transforming Model-Based Information, *In Proceedings of the Workshop on Semantic Web at ECDL-00*, Lisbon, Portugal (2000)

Castano, S., Espinosa, S., Ferrara, A., Karkaletsis, V., Kaya, A., Melzer, S., Moller, R., Montanelli S., Petasis, G.: Ontology Dynamics with Multimedia Information: The BOEMIE Evolution Methodology. *In Proc. of International Workshop on Ontology Dynamics (IWOD) ESWC 2007 Workshop* - 7 June - Innsbruck, Austria (2007)

Cruz, C., Nicolle, C.: Ontology Enrichment and Automatic Population From XML Data, *4th International VLDB Workshop on Ontology-based Techniques for DataBases in Information Systems and Knowledge Systems, ODBIS 2008*, Auckland, New Zealand, August 23, 2008, Co-located VLDB, 17-20 (2008)

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

Do, H. H., Rahm, E.: COMA - A System for Flexible Combination of Schema Matching Approaches, Proc. *28th Intl. Conference on Very Large Databases (VLDB)*, Hongkong, Aug (2002)

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

Ferdinand M., Zirpins, C.and Trastour, D., Lifting XML Schema to OWL, in: Koch, Nora and Fraternali, Piero and Wirsing, Martin (Hrsg.): *Web Engineering - 4th International Conference, ICWE 2004*, Munich, Germany, July 26-30, 2004, Proceedings, Springer Heidelberg, pp. 354-358 (2004)

García, R., Celma, O.: Semantic Integration and Retrieval of Multimedia Metadata, *Proceedings of 4rd International Semantic Web Conference*, Galway, Ireland (2005)

Gomez-Perez, A.: Some ideas and examples to evaluate ontologies, *presented at Artificial Intelligence for Applications* (1995)

Gruber, T.: Ontology, *Encyclopedia of Database Systems*, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag (2008)

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

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

Kim, M., Sengupta, A., Extracting knowledge from XML document repository: a semantic Web-based approach, *Inf. Technol. and Management* 8, no. 3, 205-221 (2007)

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

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

Martin, D., Paolucci, M., Wagner, M.: Towards Semantic Annotations of Web Services: OWL-S from the SAWSDL Perspective, *In OWL-S Experiences and Future Developments Workshop at ESWC 2007*, June, Innsbruck, Austria (2007)

OWL XML schema, http://www.w3.org/2007/OWL/wiki/OWL_XML_Schema (2007)

Rahm, E. and Bernstein, P.: A survey of approaches to automatic schema matching. *The VLDB Journal 10*, 334-350 (2001)

Rodrigues, T., Rosa, P. and Cardoso, J.: Mapping XML to Exiting OWL ontologies, *International Conference WWW/Internet 2006*, (Eds) Isaías, Pedro and Nunes, Miguel Baptista and Martínez, Inmaculada J., pp.72-77, ISBN:972-8924-19-4 (2006)

Staab, S., Gomez-Perez, A., Daelemana, W., Reinberger, M.-L. and Noy, N.F.: Why evaluate ontology technologies? Because it works*!, Intelligent Systems, IEEE 19*, 74–81 (2004)

Studer, R. Benjamins, R. and Fensel, D., Knowledge engineering: *Principles and methods, Data and Knowledge Engineering* 25, 161–197 (1998)

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

SWRL: A Semantic Web Rule Language Combining OWL and RuleML, http://www.w3.org/Submission/SWRL/, (2004)

TriG syntax, This document describes TriG, a syntax for serializing Named Graphs and RDF Datasets. http://www4.wiwiss.fu-berlin.de/bizer/TriG/

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