Deriving Ontologies from XML Schema

Ivan Bedini*, Georges Gardarin**
Benjamin Nguyen**

*Orange Labs, Site of Caen - 14000, France
ivan.bedini@orange-ftgroup.com; http://www.orange.com

**Prism, Université de Versailles St-Quentin, 78035 Versailles Cedex 035
<prenom>.<nom>@prism.uvsq.fr; http://www.prism.uvsq.fr

Abstract. In this paper, we present a method and a tool for deriving a skeleton of an ontology from XML schema files. We first recall what an is ontology and its relationships with XML schemas. Next, we focus on ontology building methodology and associated tool requirements. Then, we introduce Janus, a tool for building an ontology from various XML schemas in a given domain. We summarize the main features of Janus and illustrate its functionalities through a simple example. Finally, we compare our approach to other existing ontology building tools.

1 Introduction

Ontologies appear as useful building blocks in several domains including the Semantic Web, data source integration, data visualization and zooming, text indexing and mining, etc. In Data Warehouse and On Line Analysis, ontologies are particularly useful for integrating multiple data sources and for zooming on data cube dimensions. They provide a deep understanding of data by composing and transforming it automatically. Though helpful and trendy, ontologies are still hard to build, to maintain and make evolve, notably when large (e.g., more than thousands of concepts).

The "nuts and bolts" of ontologies are concepts and attributes with *is-a* and *part of* relationships. In that respect, ontologies are similar to object models of specific domains. XML schemas also have similarities with object models. While there exist many XML schemas in certain domains (eg., B2B e-commerce are at standardization level), there are only very few *ontologies*. In this paper, we propose a methodology and a tool for semi-automatic derivation of ontologies from XML schemas. We present the main features of our tool called Janus and illustrate it through a simple example. Janus is unique in the sense that it mixes several technologies (language analysis, text mining, schema validation, graphical representation) into an incremental methodology of ontology construction and evolution. We further propose a classification of ontology building tools and their comparison in terms of functionalities.

The rest of this paper is organized in four main sections: i) ontology definition and requirements for building tools; ii) overview of our Janus ontology skeleton building tool; iii) presentation of other tools; and iv) comparisons of tools.