

Translating XML Models into OWL Ontologies for Interoperability of Simulation Systems

He Tan, George Barakat, and Vladimir Tarasov

Tekniska Högskolan, Högskolan i Jönköping, Sweden
{he.tan, Vladimir.Tarasov}@jth.hj.se, sbarakat.george@gmail.com

Abstract. Today XML is a common format supporting interoperability and information exchange between systems in the modeling and simulation field. Although XML enables systems to agree on a common syntax and understand the exchanged information, systems can misinterpret them due to their different conceptualizations of the domain of interest. In this paper, we present a framework for automatic translation of XML simulation models which follow the High Level Architecture (HLA) object model template specification, into OWL ontologies. In OWL ontologies the semantics of information is formally defined. It provides the basis for interoperability and information exchange between simulation systems on semantic level.

Key words: Semantic Interoperability, Ontology, Ontology Language, OWL, XML

1 Introduction

One of the major problems in the modeling and simulation (M&S) field is to improve interoperability and reusability of systems, especially when systems are distributed, autonomous and heterogeneous [1]. The reason for this is that many military and civil organizations need not just a single system that meets an exactly specified set of requirements, but need solutions to meet diverse and changing needs of users. WISE (Widely Integrated Systems Environment) developed by SAAB is a generic integration platform that allows connection of simulation systems into a common environment and supports information exchange between the individual systems, regardless of individual architecture, communication standards and protocols [2].

In WISE the information to be exchanged is described in a XML-based common representation which follows the High Level Architecture (HLA) Object Model Template (OMT) specification [3]. Although the XML-based information representation enables the systems to agree on a common syntax and understand the exchanged information, the systems can misinterpret the information due to their different conceptualizations of the subject of a domain. This may result in a mismatch between the intended and actual effect of the information. The correctness of interoperability and integration depends completely on the knowledge of human engineer about low-level data structures and his implemen-

tation of the semantically correct integration. In this paper we present our effort to extend WISE to support interoperability and information exchange between simulation systems on semantic level. In particular, we propose a strategy for how OWL ontologies can be generated automatically out of the XML models of simulation systems. When the semantics of the exchanged information is formally defined in OWL ontologies, it is possible to provide automated support in defining integration and/or formal automatic verification of manual integration using the ontologies.

The rest of the paper is organised as follows. Section 2 presents the background and the related work. In section 3 we describe our strategy for translating the XML models of simulation systems into OWL ontologies. We evaluate and discuss the quality of the ontologies generated from XML models in section 4. In the end we present our conclusion and the future work.

2 Background and Related Work

2.1 Interoperability of Systems in the Modeling and Simulation Field

The M&S community has focused on technical layer for interoperability between systems [4]. The development of the standards, such as HLA, has simplified the integration tasks. The HLA defines a standard for the interoperation of simulation systems through the communication of the models of objects and events of interest. HLA OMT standardizes the common syntax specification for the models. The semantics of the models is defined simulation objective specific in the federation development and execution process and documented in the federation object model (FOM) lexicon as well as in the federation agreements. The problem for the standardized solution is that it is complex and high-cost to integrate a new system having different information model. The more systems integrated, the more complex it becomes to integrate an additional system.

WISE goes beyond the standardized solutions. The focus of system integration is switched from implementation details to the flow of information. The integration platform is seen as an information infrastructure. It is responsible for collecting, translating and delivering information. Each system only needs communicate with the infrastructure. The engineer who builds a connection of a system to the information infrastructure, manually defines the mappings between the model of the information the system needs, to a common model. The WISE currently includes the common models for command and control information in military, real-time simulation, and emergency alerts and public warnings for the civil domain.

In WISE the information exchange models still follow the HLA OMT syntax specification, and are implemented in XML. Since the semantics of information exchange is not formally defined in the XML models, it is not possible to provide formal automated verification of the manual mapping, and/or automated support in defining the mapping. The correctness of the connection depends

completely on the knowledge of the engineer about low-level data structures and his implementation of the semantically correct integration.

2.2 Ontology for Interoperability of Simulation Systems

Over the last few years the community has found that many challenges in the area are on higher levels, underlying concepts and models that have to be aligned, not the implementation questions [4]. Many researchers started to study how to utilize ontology to promote semantic and pragmatic interoperability in the field. One direction of the research is toward enhancing the semantic expressiveness of the base object model (BOM) using ontology (e.g. [5, 6]). The BOM [7] is a SISO standard to response to the semantics lacking in the HLA simulation object model (SOM) and FOM, but itself does not contain sufficient information for defining conceptualizations. While the semantics of SOM and FOM can be defined in BOM using this kind of methods, it is usually very complex do the alignment between BOMs.

Another direction is to develop complete ontology-based framework to achieve semantic interoperability [Ra04]. To build up a simulation among distributed systems, developers have to develop domain ontologies for the individual systems and an ontology for certain simulation objective. Although there are many methods, tools and guidelines for ontology development, building ontologies is still not a simple task, particularly when engineers have no background knowledge on ontology engineering techniques and/or they have not much time to invest in domain conceptualization. In this paper we propose a strategy for how OWL ontologies can be generated automatically out of the XML models, so that there is no requirement on engineers to developing ontologies.

2.3 XML to OWL Strategies

Several strategies have been proposed for translating XML to OWL or RDF meta-data (e.g. [9, 10]). All existing XML to OWL mapping strategies assume that an XML document always contains instance data, so that the information from an XML document is always mapped to OWL instances and an XML schema is translated into an OWL model. Although most of the work deal with creating an ontology from a single XML source, Janus framework [11] presents a method for generating an ontology from a large source of XML schemas based on pattern recognition. The work in [12] also proposes the solutions for generating a local OWL model from heterogeneous XML data sources.

In the case of WISE project, every XML document represents either a model for individual simulation system, or a common model for a simulation domain. No actual data is included in XML documents. All XML documents are validated against an XML schema to verify the structure of the documents. As a result, the adoption of any of the existing strategies is not sufficient to produce the intended OWL ontologies in our project. We aim to propose a method focusing on translating XML to OWL on conceptual level. Every XML document

is translated into an ontology representing a more specific simulation domain. The XML schema is translated into an ontology that can be considered as the top domain ontology for the M&S field.

3 The Translation

The XML model is based on the labelled tree, where the meaning of the tag-nestings are interpreted by the program undertake on it. The OWL model is based on the subject-predicate-object structure from RDF/RDF-S, where objects, their attributes and relations are naturally represented, and the semantics is specified [13]. Translation from XML to OWL is to interpret the tree structure of XML, and to represent the intended model in the subject-predicate-object structure. In this work we implemented the translation from XML to OWL Full language [14].

3.1 Mapping Rules

Two kinds of entities present in WISE XML models: objects and events. An object is an entity which persists over time (e.g. "device", and "system") while an event is an entity which only exists momentarily (e.g. "act", and "interaction"). Every entity, either an object or an event, is translated into a OWL class. Each entity always has a set of properties. Many of the properties are datatype properties, describing relations between individuals and data values. All the datatypes in WISE are user-defined types that always restrict the value of a property to a subset of an existing type, not the build-in ones from XML Schema. They are important in capturing the intended meaning of the elements in the models. For example, the "HLAfloat32LE" is a HLA standard datatype: `<float description="Single-precision floating-point number." name="HLAfloat32LE"/>`. They are translated to RDF datatype classes.

Either a model for individual simulation system, or a common model for a simulation domain, is described in a XML document. All XML documents are validated against a XML schema to verify the structure of the documents. The XML schema indeed describes very general concepts that are the same across all XML object models. It defines what object is, what event is, their general properties, datatypes specific in simulations, etc. It supports semantic interoperability between the XML models of more specific simulation domains. Therefore, the general model described in the XML schema can be considered as a top domain ontology for the M&S field. Table 1 gives the principle mapping rules for the translation of the XML schema to the OWL top domain ontology.

3.2 The Translation Process

Figure 1 shows the process for the translation. First, a set of manually defined mapping rules are used to generate an XSLT stylesheet to translate the XML

Table 1. The principle mapping rules for the correspondences between XML schema model and OWL ontology

XML Schema	OWL Full
named xsd:elements (not within the element "datatypes")	Class
<xs:element name="attribute" type="X" ...>	ObjectProperty, DatatypeProperty or AnnotationProperty (determined by the value of the type).
named <xsd:attribute type="X" ...>	ObjectProperty, DatatypeProperty or AnnotationProperty (determined by the value of the type).
named xsd:complexType on the top level	rdfs:Datatype
named xs:element within the element "datatypes"	rdfs:Datatype
xsd:minOccurs, xsd:maxOccurs	owl:minCardinality, owl:maxCardinality

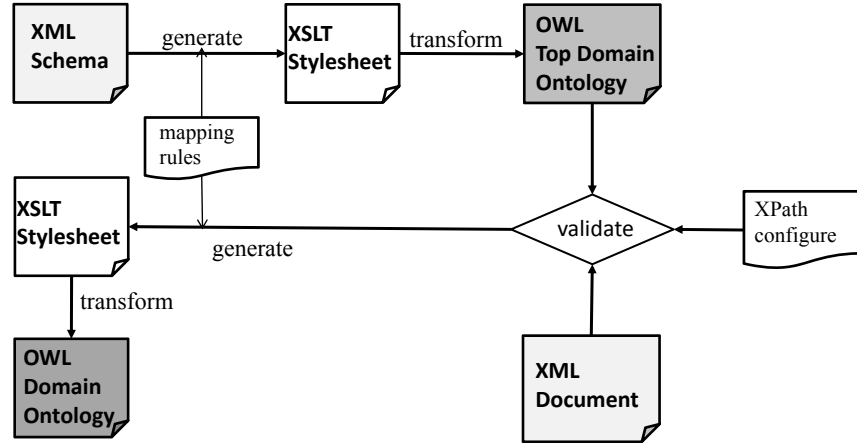


Fig. 1. The Translation Process

schema into the top domain ontology. This is a one time translation when all XML documents share the same schema. A produced XSLT stylesheets can be used by any XSLT processor to automatically generate the desired ontology. The second step is to translate the source XML document to an domain ontology. Not all schema constructs defined in the XML schema appear in every XML document. Since each XML document contains a different set of schema constructs, XSLT stylesheet for translating each XML document is different. A configuration file maintains the XPath expressions for each kind of OWL expressions in the top domain ontology, including classes, properties and datatypes. If an

XPath expression is evaluated as a valid XPath in the source XML document, a corresponding XSLT stylesheet is generated based on the manually defined mapping rules. In the end, the produced XSLT stylesheet is used to automatically generate the domain ontology from the source XML document.

4 Evaluation and Discussion

The goal of the work is to capture the semantics of exchanged information between simulation systems in XML models and represent them formally in OWL ontologies, so the focus of our evaluation is to verify and validate the structure and semantics of the generated ontologies against XML data sources. One proper method to perform this evaluation is relying on domain expert assessment against a set of criteria [15]. In the evaluation we considered five ontology quality criteria: accuracy, consistency, completeness, clarity and adaptability [16]. Table II shows the criteria, assessment questionnaire and the assessment results according to a 1-to-5 scale, where 1 corresponds to 'totally disagree' and 5 to 'fully agree'. Three resulted ontologies are assessed by an ontology expert. The ontology expert is also acquainted with the relevant XML data sources. One of the three ontologies is generated from the XML schema, and the other two are generated from XML models of two specific simulation domains.

The results from the assessment show that the translation process provides accurate conceptualization during automatic translation of XML object models into OWL ontologies. The assessment also shows that the resulted ontologies are consistent. With regard to completeness, the expert 'fully agrees' that the domain of interest is appropriately covered in the resulted ontologies and is 'neutral' about any implicit knowledge in XML object models. The latter can be explained by the fact that the expert was also 'neutral' about his level of expertise in the simulation domain. However, the results still show soundness of the translation process and the mapping rules. Regarding clarity, the expert found no ambiguity in the names of classes and properties and that naming conventions were properly applied. However, the expert was 'neutral' about the ease of understanding of the conceptualization of the constructed ontologies. Thus, more effort can be made to improve the descriptions of classes and properties and their clarity level. Finally, the expert 'agrees' that the resulted ontologies can be adapted to different usages but comments that it is difficult to judge without being able to evaluate the use of the constructed ontologies within their applications.

5 Conclusion

In this paper we presented an effort to enable interoperability and integration in the M&S field on the semantic level. XML is the common format for information exchange between systems in the field. We proposed the method for automatically translating XML models in the domain into OWL ontologies. The translation intends to capture the semantics of exchanged information in the

Table 2. Ontology Evaluation Criteria and Result

Accuracy: determines if the asserted knowledge in the ontology agrees with the expert's knowledge about the domain. A higher accuracy comes from correct definitions and descriptions of classes and properties.	Result
1. Are [rdfs:Datatype]s well structured and do they properly represent the datatypes in XML object model?	5
2. Are [owl:Class]s well structured and do they properly represent the entities in XML object model?	5
3. Are [owl:ObjectProperty]s well structured and do they properly represent the possible relations between defined datatypes?	5
4. Are [owl:DatatypeProperty]s well structured and do they properly represent the attributes of the defined datatypes, and entities?	5
5. Are [owl:AnnotationProperty]s well structured and do they properly provide descriptions about the defined entities and their attributes?	5
Consistency: describes if the ontology does not include any contradictions. Not only asserted knowledge must be logically consistent, but also the formal and informal descriptions in the ontology should be consistent	
6. Is there any logical contradictions inferred in the constructed ontologies. e.g. by running a reasoner?	1
7. Is there any other contradictions, e.g. the documentation and comments of an entity should be aligned with its definition in ontology?	1
Completeness: determines if the domain of interest is appropriately covered.	
8. Are all the entities and datatypes described in XML object model represented in the ontology?	5
9. Is implicit knowledge in XML object model captured by the resulted ontology?	3
Clarity: measures if the resulting ontology communicates the intended meaning of the defined terms.	
10. Is it easy to understand the conceptualization of the resulting ontology?	3
11. Are the names of classes and properties unambiguous?	5
12. Are the descriptions of classes and properties unambiguous?	4
13. Are the naming conventions properly applied to classes and properties?	5
Adaptability: measures how far the ontology can be adapted to anticipated usages. The resulting ontology should offer the conceptual foundation for a range of anticipated tasks	
14. Could I apply the constructed ontologies in anticipated M&S tasks?	4
15. Could I extend or specialize the resulted ontology monotonically, i.e. without the need to remove axioms?	4

XML models and represent them formally in OWL ontologies. The evaluation has showed that the ontologies generated by the automatic translation successfully capture the semantics of the information and correctly represent them. One direction of our future work is to develop algorithms and methods for automatic integration support. When the semantics of exchange information is formally defined in OWL ontologies, it is possible to provide automated support for defining integration and formal automatic verification of manual integration.

Acknowledgment

We thank Kurt Sandkuhl for comments on the project. We also acknowledge the technical support of Saab Training Systems AB.

References

1. Zeigler, B. P., Praehofer, H., Kim, T. G. (2000). *Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems*. Academic press.
2. Gustavsson, P. M., Lundmark, S., Wemmergård, J. (2009). *Interoperability in the Next Generation Training Systems*. In Spring Simulation Interoperability Workshop.
3. HLA Working Group. (2010). *IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-Framework and Rules. Standard IEEE STD 1516-2010*. IEEE CS Press.
4. Andreas, T., Saikou, D., Charles, T. (2007). *Applying the levels of conceptual interoperability model in support of integrability, interoperability, and composability for system-of-systems engineering*. In Journal of Systemics, Cybernetics and Informatics.
5. Mojtahed, V., Svee, E. O., Zdravkovic, J. (2010) *Semantic Enhancements when Designing a BOM-based Conceptual Model Repository*. In Proc. 2010 European Simulation Interoperability Workshop.
6. Yanjun, Y., Fengju, K. (2009). *HLA System Development Based on Semantic Enhanced BOM*. In Information Engineering and Computer Science, 2009.
7. *Guide for Base Object Model (BOM) Use and Implementation*. Simulation Interoperability Standards Organization (SISO). SISO-STD-003.1-2006.
8. Rathnam, T. (2004). *Using ontologies to support interoperability in federated simulation*. Master Thesis, Georgia Institute of Technology.
9. Bohring, H., Auer, S. (2005). *Mapping XML to OWL Ontologies*. Leipziger Informatik-Tage, 72.
10. Rodrigues, T., Rosa, P., Cardoso, J. (2006). *Mapping XML to Existing OWL ontologies*. In International Conference WWW/Internet.
11. Bedini, I., Matheus, C., Patel-Schneider, P. F., Boran, A., Nguyen, B. (2011). *Transforming XML schema to OWL using patterns*. In 5th IEEE International Conference on Semantic Computing (ICSC).
12. Yahia, N., Mokhtar, S. A., Ahmed, A. (2012). *Automatic generation of OWL ontology from XML data source*. In International Journal of Computer Science Issues (IJCSI) 9(2).
13. Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Horrocks, I. (2000). *The semantic web: The roles of XML and RDF*. Internet Computing, IEEE, 4(5).
14. McGuinness, D. L., Van Harmelen, F. (2004). *OWL web ontology language overview*. W3C recommendation, 10(10), 2004.
15. Obrst, L., Ceusters, W., Mani, I., Ray, S., Smith, B. (2007). *The evaluation of ontologies*. In Semantic Web.
16. Vrandečić D. (2010). *Ontology evaluation*. PhD dissertation, Karlsruhe Institute of Technology.