Mapping Pragmatic Class Models to Reference Ontologies

Heiko Paulheim #1, Roland Plendl #2, Florian Probst #3, Daniel Oberle *4

#SAP Research
Bleichstrasse 8, 64283 Darmstadt, Germany
1heiko.paulheim@sap.com
2roland.plendl@sap.com
3f.probst@sap.com

*SAP Research
Vincenz-Priessnitz-Strasse 1, 76131 Karlsruhe, Germany
4d.oberle@sap.com

Abstract—Capturing a domain of discourse in an object-oriented class model and in a reference ontology leads to different results. On the one hand, modeling decisions for class models are motivated by pragmatic and efficiency-related choices because modeling decisions are motivated by different choices. However, information integration scenarios require a coexistence between reference ontologies and class models at runtime. When implementing an integration scenario, objects need to be transformed between both types of models, thus requiring expressive, executable, and bidirectional mappings. In this paper, we present a novel approach for a bidirectional mapping between a class model and an ontology. The approach is both non-intrusive and dynamic. That means it can be integrated in existing IT systems without having to change the class model and it does not rely on static 1:1 mappings, respectively.

I. INTRODUCTION

The recent W3C Semantic Web recommendations, such as OWL, are increasingly used to specify reference ontologies. The idea of a reference ontology is to be a generic, commonly agreed upon conceptual model of a domain. To this end, modeling decisions are taken in a way such that the semantics of domain terms are captured as precisely as possible. As an example, consider the ISO 15926 Oil and Gas ontology standard whose purpose is to provide a lingua franca for Oil and Gas IT systems [1].

One goal of reference ontologies is to facilitate and enable information integration across IT systems. For example, the declared goal of the Norwegian Oil and Gas industry is to enable information integration based on ISO 15926 to support, e.g., the integration of different data sources and sensors, the validation of information delivered by different sources, and the exchange of information within and between companies via web services [2]. This requires that IT systems share information according to the reference ontology, and that they can interpret messages using that ontology.

However, existing IT systems are based on database schemas or class models that historically deviate from the reference ontology. In addition, we argue that even for new IT systems, it is likely that their class models or database schemas will deviate from the reference ontology as demonstrated by [3], [4]. The reason is that class models are task-specific, with the focus on an efficient implementation of an application. In contrast to reference ontologies, modeling decisions are geared towards a pragmatic and efficient model. Due to those differences, one often faces the situation where class models and reference ontologies are incompatible in the sense that a 1:1 mapping between them does not exist. Yet, it is required to use both the pragmatic class model and the reference ontology in parallel and to transform objects between class models and ontologies as illustrated by the Oil and Gas integration scenario. Practically, due to the dominant paradigm of object-oriented programming, this means that programming models supporting arbitrary mappings between class models and ontologies are needed. Despite the existence of sophisticated solutions for model mapping [5] and database integration [6], the existing approaches for mapping class models to ontologies are still limited with respect to supporting arbitrary mappings.

In this paper, we contribute a novel approach for bidirectionally mapping between reference ontologies and class models. As required for integrating existing IT systems, the approach is non-intrusive, i.e., it can be implemented without having to change the class model. The approach is dynamic, i.e., it does not rely on static 1:1 mappings between the class model and the reference ontology. Instead of using static mappings between the class model and the ontology, our approach uses rules which are dynamically evaluated at run-time.

We start by introducing a scenario of information integration in the area of emergency management in Section II. This scenario originates from building a prototype in the context of the German lighthouse project called SoKNOS.[1] The section also highlights typical deviations between class models and reference ontologies which go beyond a 1:1 mapping. Section III discusses how mappings for both directions are specified in the form of rules, followed by a discussion of the architectural design in Section IV. A scalability and performance evaluation can be found in Section V. Finally,
we discuss related work and give a conclusion in Sections VI and VII, respectively.

II. Motivating Scenario

Emergency management is a domain where semantically unambiguous information exchange between organizations is crucial. Fighting large incidents like floodings or earthquakes requires close cooperation between a large number of organizations. If such an incident occurs, an emergency headquarters is put into place. The command staff operating the headquarters has the task to plan counter measures and to coordinate the joint efforts of all involved organizations. Since more and more organizations use IT systems, the challenge of system interoperability arises.

In the project SoKNOS, a prototype of an emergency management infrastructure has been developed, which integrates both newly developed and existing IT systems, ranging from geographical information providers to databases for managing resources. The SoKNOS system enables the command staff to quickly integrate new information services, such as map overlays or sensor information, but it also enables the integration of complete modules from other systems, such as modules displaying the currently available resources [7], [8].

For facilitating information exchange between the IT systems, we have developed a reference ontology of the emergency management domain [9], which defines the categories and relations of relevant terms in the emergency management domain, based on the formal top level ontology DOLCE [10].

The SoKNOS prototype includes a number of system components developed in Java, which use a simplified, pragmatic class model for information processing. Several deviations do occur between such a class model and a reference ontology. For example, measures in emergency management, such as extinguishing a particular fire, have to be planned and carried out. Therefore, both planned measures as well as measures that are actually performed have to be stored by an emergency management system. Fig. 1 shows a multi purpose class, a typical deviation between a pragmatic class model and a reference ontology: The class Measure contains a flag type indicating whether an instance is an actual or a planned measure. In the ontology, both belong in different categories which do not even have a meaningful common super category.

Information exchange in the SoKNOS prototype requires IT systems to represent and share information according to the reference ontology but residing in the corresponding object. Therefore, a mapping from objects to the ontology has to be put in place and vice versa. However, objects of a class cannot be mapped statically at design time. In our example, mapping an object of a class with either Measure or PLANNED MEASURE can only be performed at run-time by inspecting the Measure object and dynamically assigning a mapping, based on the value of the type flag.

Besides multi-purpose classes, there are a number of other deviations that can occur due to the differences of class models and reference ontologies:

- **Multi purpose relations**, just as multi purpose classes, are relations between objects that are used for expressing different relations in the ontology, distinguished by a flag or by implicit background knowledge, such as naming conventions of the related object. A typical example would be a String attribute contactData, which, if it contains an @ symbol, is interpreted as an e-mail address, otherwise as a telephone number.
- **Artificial classes** are classes that do not have a corresponding category in the ontology, but hold some meaningful information, such as a class CustomerData containing both address and payment information.
- **Relation classes** are classes that do not correspond to a category, but to a relation in the ontology, e.g. a Marriage class corresponding to a MARRIED TO relation in the ontology.
- **Shortcuts** are direct relations between objects that only exist as indirect relations via intermediate categories in the ontology, e.g., a Customer object having a ZIP code attribute, while the connection between the CUSTOMER and the ZIP CODE category involves the intermediate category CITY in the ontology.
- **Conditional objects** are objects whose existence is determined by a flag. Most commonly, this occurs as a deleted flag indicating that the object has been deleted.
- **Object counting** occurs if a set of relations to other objects is represented by the number of those objects, e.g., using an attribute numberOfDamages instead of a set of damage objects.
- **Non atomic attributes** are attributes that contain information belonging to several categories in the ontology. A typical example are attributes such as address, which contains both the street name and the street number. Those are often mapped to two different categories in a reference ontology.

III. Mapping Specification

Simply mapping one element from the class model to a category in the ontology is insufficient as we have learned in the previous section. Instead, arbitrary mappings between class models and ontologies are required which in turn necessitate an expressive representation formalism to specify the mappings. Therefore, some more versatile solution is needed, which allows for declaratively expressing dynamic mappings and interpreting those mappings at run-time.

We use rules for expressing dynamic mappings between class models and ontologies (and vice versa). The rules can
be evaluated at run-time on objects to create the desired RDF graph describing the object, and on an RDF graph to create the corresponding set of objects. Fig. 2 shows how the mapping rules are used to transform a Java object into an RDF graph and back, using the example depicted in Fig. 1. The next sections explain the different rule syntaxes in detail.

A. Mapping Class Models to Ontologies

Our mapping approach uses tests on the objects to be mapped as rule bodies, and a set of RDF triples to be produced as rule heads. For each class, one rule set is defined. In cases where objects have relations to other objects, the rule sets of the corresponding related classes are executed when processing a related object. Rule sets defined for super classes are inherited to sub classes, however, the developer may also override inherited rules explicitly.

As the example depicted in Fig. 2 shows, the tests on the objects are defined using XPath expressions. The result of the expression can be used as a variable in the RDF triple to be generated (referred to with the % sign), as well as the current object (referred to with the . sign). The uri function used in a triple generates a unique URI for an object. The triples may also contain blank nodes, which are needed, e.g., to cope with shortcuts.

The XPath expressions may also contain regular expressions to deal with implicit background knowledge and non-atomic data types. Those can be used for conditions (e.g., does the contactData attribute contain an @ symbol?) or for splitting data values, e.g., separating street names from numbers. A repeat function can be used to cover object counting deviations (e.g., produce a set of n blank nodes for an attribute value of n).

B. Mapping Ontologies to Class Models

The mapping rules from ontologies to class models are similar. Again, rule sets are defined per ontology category, are inherited to sub categories, and rules can be explicitly overridden by the developer.

For rule bodies, SPARQL expressions are used. In the rule heads, objects are created by using createObject, and attribute values for created objects are set (by using getObject and an XPath expression identifying the attribute to be set). The execution order of rules is defined such that all createObject statements are executed first, assuring that all objects are created before attempting to set attribute values.

For determining which categories an RDF instance belongs to, and for executing SPARQL statements, reasoning on the RDF graph and the domain ontology can be used. To cover non-atomic data types, the rules may use a concat function for concatenating different results of a SPARQL query. For coping with object counting, a count function can be used to produce the corresponding attribute values (once SPARQL version 1.1, which supports counting, becomes a standard, this will be obsolete).

Although the rules for both directions look similar, there is one subtle difference. The XPath expressions used on the object model are executed with closed world semantics, while the SPARQL expressions used on the RDF model are executed with open world semantics. The rationale is that the set of objects in an information system is completely known (and therefore forms a closed world), whereas an RDF graph representing a set of objects will typically only represent a subset of the original information.

IV. ARCHITECTURE

Fig. 3 shows the architecture of our implemented prototype with two IT systems involved in an information exchange. As a first step, Java→RDF and RDF→Java mapping engines have to be put in place. As a second step, Java→RDF rules for mapping each object and RDF→Java rules for mapping each ontology category have to be specified, as discussed in the previous section. The rules have to registered in the corresponding mapping engine. At runtime, the IT systems are handled as black boxes with an API, following the paradigm of non-intrusiveness. When an object from an IT system is retrieved via its API and needs to be exchanged, an annotation in RDF is generated which conforms to reference ontology. The exchange of objects between IT systems is controlled by object exchange components. They retrieve objects from and hand over objects to the IT systems’ APIs, invoke the
corresponding mapping engines, and control the transfer to other object exchange components.

Let us assume IT System 1 needs to exchange information with IT System 2. For generating the RDF representation, rules are processed by a Rule Engine. For executing rules, the Object Inspector evaluates conditions, and a URI Factory creates unique URIs for each object. An RDF Writer, unifies the outcomes of all relevant rules and creates the RDF annotation.

On the side of IT System 2, an object has to be created from the received RDF annotation. The RDF Reader first reads the received annotation, and another Rule Engine processes the mapping rules. A SPARQL Processor evaluates conditions, and an Object Factory eventually creates objects. The objects can then be used for calling IT System 2’s API.

When exchanging objects, it may be desirable not to transfer all RDF statements that can be generated for a particular object, but only a subset. Since connected object graphs can be fairly large, the resulting RDF annotation can grow considerably in size, thus slowing down the transmission as well as the transformation back to objects. Therefore, an RDF Filter can be applied for reducing the set of RDF statements to be transmitted. The filter uses RDF templates which consist of blank nodes and labeled relations, according to the reference ontology. When applying a template to an RDF graph, only those resources are kept which are contained in the template.

Although the filtering could also be done by defining a subset of rules, we have chosen to keep both issues separated. The rules define all possible annotations that can be generated for a particular object and are therefore universal. The templates define the subset that is needed for a particular use case, e.g., transmission to another IT system. It is possible to use the same set of rules with a different set of templates for different use cases. For example, information exchange between systems owned by the same company might include more information, while the information exchanged with systems outside of the company is restricted.

We have implemented the solution sketched above in a Java-based prototype, using JXPath² for evaluating the XPath expression on Java objects and Jena [11] for RDF processing. The parsers for the rules were generated with JavaCC.³ By using Java’s reflection API [12] and relying on the Java Beans specification [13] (a naming convention for object constructors and for methods for accessing property values), the implementation is non-intrusive and does not require changes to the underlying class model.

V. EVALUATION

We have evaluated our mapping prototype with the class model used in the SoKNOS prototype and the reference ontology of emergency management described in [9]. The SoKNOS class model consists of 58 classes with 30 attributes and 25 relations. The ontology consists of 156 categories and 136 relations. The number of occurrences of the deviations of the mapping between both are depicted in Table I. Thus, about 16% of all model elements (classes, attributes, and relations)

²http://commons.apache.org/jxpath/
³https://javacc.dev.java.net/
TABLE I

DEVIATIONS IN THE SOKNOS PROTOTYPE.

<table>
<thead>
<tr>
<th>Deviation type</th>
<th>No. of Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi purpose class</td>
<td>3</td>
</tr>
<tr>
<td>Multi purpose relation</td>
<td>2</td>
</tr>
<tr>
<td>Artificial class</td>
<td>0</td>
</tr>
<tr>
<td>Relation class</td>
<td>1</td>
</tr>
<tr>
<td>Shortcuts</td>
<td>3</td>
</tr>
<tr>
<td>Conditional objects</td>
<td>2</td>
</tr>
<tr>
<td>Object counting</td>
<td>3</td>
</tr>
<tr>
<td>Non atomic attributes</td>
<td>4</td>
</tr>
</tbody>
</table>

and 6% of all ontology elements (categories and relations) are involved in some of the deviations.

Our mapping prototype was able to handle most of the deviations. The only problematic cases are non-atomic data types. In cases where a convention exists that can be formalized, such as “a street name never contains numbers, and the street number always starts with a digit between 1 and 9,” our approach can handle them by using regular expressions. Cases where such a formalization cannot be found, such as for separating first and last names, cannot be handled automatically without human intervention.

We processed different, artificially created sets of objects from 10 to 10,000 connected objects in order to test the performance and scalability of our approach. Fig. 4 shows the processing times for producing RDF annotations for Java objects. The processing times scale linearly, the average time for processing one object is below one millisecond.

Fig. 5 shows the corresponding processing times for creating Java objects from RDF annotations. For processing SPARQL statements, different reasoning options can be used. We tested our approach with three built-in reasoners of Jena [11]: the transitive reasoner, which only uses rdfs:subClassOf and rdfs:subPropertyOf statements, the RDFS reasoner, and the OWL reasoner, which covers OWL Lite reasoning. Except for the latter, which creates some overhead for larger sets of objects, creating an object from RDF takes less than ten milliseconds, and the processing times scale linearly.

VI. RELATED WORK

In this paper, we have discussed an approach for creating annotations for Java objects, and for re-constructing Java objects from such annotations. Such approaches can be roughly categorized into generative, intrusive, and non-intrusive approaches. The latter are the only approaches that can be applied when the developer cannot or must not change the class model, be it for technical or for legal reasons.

Generative approaches create a class model from an ontology. They can thus be seen as a special case of model driven architecture (MDA) [14]. Examples for generative approaches are RDFReactor [15], OWL2Java [16], OntoJava [17], and agogo [18]. As a special case, the class model may also be generated on-the-fly at run-time when using dynamically typed scripting languages, as shown with ActiveRDF [19] for Ruby and Tramp for Python. For the developer, such approaches are more comfortable, since they do not require additional tools such as a code generator, but work directly out of the box by invoking an API. In general, generative approaches usually create one Java class per ontology category, therefore, a 1:1 mapping exists by definition. Generative approaches are a suitable solution for developing new IT systems, but not for integrating existing ones.

Intrusive approaches do not generate new Java classes from an ontology, but modify (i.e., intrude into) an existing class model by adding additional code fragments, such as Java annotations or additional methods and/or attributes. Examples for intrusive approaches are otm-j [20], Sommer,7 Texai KB [21], RDFBeans,8 Jenabean,9 and P-API [22]. All those approaches require a 1:1 mapping between Java classes and categories in the ontology.

Non-intrusive approaches can produce annotations without altering the class model. The only examples we found are ELMO10 and the EMF-RDF transformations introduced in [23]. Both approaches require that the “domain model should be rather close to the ontology” [23] and cannot cope with most of the deviations discussed above.

While some of the approaches also cover both directions of conversion – from Java to RDF and from RDF to Java – and can cope with some of the deviations discussed above, to the best of our knowledge, there are no solutions that can cope with all the deviations and can therefore support arbitrary mappings between class models and reference ontologies.

Fig. 4. Run time for creating the RDF annotation for a set of Java objects

Fig. 5. Runtime for creating Java objects from a set of annotations, using different types of reasoning

6http://www.aaronsw.com/2002/tramp/
7http://code.google.com/p/jenabean/
8http://code.google.com/p/rdfbeans/
9http://code.google.com/p/jenabean/
10http://www.openrdf.org/doc/elmo/1.5/

The tests were run on a Windows PC with a Pentium 3.6GHz dual core processor and 2GB of RAM.
VII. CONCLUSION AND OUTLOOK

We have discussed a number of typical deviations between class models and reference ontologies. In a use case from the emergency management domain, we have shown that the deviations do occur in actual IT systems and ontologies. We argue that also other domains are affected by such deviations, e.g., the Oil and Gas domain.

We have introduced an approach for mapping class models to ontologies. Our approach uses rules both for creating RDF representations of objects as well as vice versa, thus supporting information exchange between Java objects as RDF between potentially heterogeneous class models.

A prominent area of using ontologies is facilitating interoperability between heterogeneous information systems by allowing for semantically unambiguous information exchange. As this may involve IT systems that cannot or must not be altered for technical or legal reasons, our approach has been implemented in a non-intrusive way. It supports arbitrary mappings between class models and ontologies. We have further evaluated the performance and scalability of our approach, showing that the transformations are performed within milliseconds, growing linearly with the number of objects.

The paper has focused on the use case of information integration and exchange. However, the mechanism of mapping different class models to one common ontology may also be used to make the data in different applications available as a unified linked data set, allowing for reasoning and for unified visualization [24].

Currently, the developer has to develop the mapping rules by hand. In the future, techniques developed, e.g., in the field of ontology matching [25] may be employed to assist the user, however, the state of the art in ontology matching is most often focused on discovering 1:1 mappings.

REFERENCES


