Using OWL for Querying an XML/RDF Syntax

Rubén Tous
Universitat Pompeu Fabra (UPF)
Pg. Circumval.lació, 8. E-08003
Barcelona, Spain
ruben.tous@upf.edu

Jaime Delgado
Universitat Pompeu Fabra (UPF)
Pg. Circumval.lació. E-08003
Barcelona, Spain
jaime.delgado@upf.edu

ABSTRACT

Some recent initiatives try to take profit from RDF to make XML documents interoperate at the semantic level. Ontologies are used to establish semantic connections among XML languages, and some mechanisms have been defined to query them with natural XML query languages like XPath and XML Query. Generally *structure-mapping* approaches define a simple translation between trivial XPath expressions and some RDF query language like RDQL; however some XPath constructs cannot be covered in a *structure-mapping* strategy. In contrast, our work takes the *model-mapping* approach, respectful with node order, that allows mapping all XPath axis. The obtained XPath implementation has the properties of schema-awareness and IDREF-awareness, so it can be used to exploit inheritance hierarchies defined in one or more XML schemas.

Categories and Subject Descriptors

D.2 [**Software**]: Software Engineering; D.2.12 [**Software Engineering**]: Interoperability —*Data mapping*

General Terms

Design, algorithms

Keywords

XML, RDF, XPath, interoperability, semantic integration, ontologies, schema-awareness, idref-awareness

1. INTRODUCTION

1.1 Motivation

This work aims to contribute to a recent research trend that defines an XML to RDF mapping allowing XML documents interoperate at the semantic level. We improve other approaches demonstrating that an XPath processor, respectful with the node order, can be implemented on top of RDF, feeding an inference engine with an XML/RDF Syntax ontology defined with OWL. The resulting processor has some interesting properties, not present in conventional implementations, like schema-awareness and IDREF-awareness.

WWW 2005, May 10–14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

1.2 Related work

In 2001, [7] defined the terms *structure-mapping* and *model-mapping* to differentiate between works that map the structure of some XML schema to a set of relational tables, and works that map the XML model to a general relational schema respectively.

More recently, some works face the problem to map XML and RDF from different points of view. [4] takes a *structure-mapping* approach and defines a direct way to map XML documents to RDF triples ([2] classifies this approach as *Direct Translation*). [3], [2], and [1] take also a *structure-mapping* approach but focusing on defining semantic mappings between different XML schemas ([2] classifies their own approach as *High-level Mediator*). They also describe some simple mapping mechanisms to cover just a subset of XPath constructs. Other authors like [5] or [6] take a slightly different strategy (though within the *structure-mapping* trend) and focus on integrating XML and RDF to incorporate to XML the inferencing rules of RDF (strategies classified by [2] as *Encoding Semantics*). Finally it's worth mention the RPath initiative [1], that tries to define an analogous language to XPath but for natural (not derived from XML) RDF data (this last work doesn't pursue interoperability between models or schemas).

2. AN OWL ONTOLOGY FOR THE XML MODEL (XML/RDF SYNTAX)

Our approach takes a strategy more similar to the *model-mapping* approach. We tried to represent the XML Infoset using an OWL ontology. This allows us to represent any XML document without any restriction and without losing information about node-order. The description of the ontology in Desctiption Logics syntax ($\mathcal{SHIQ}$-like style [8]) would be:

$$\begin{aligned} Document &\sqsubseteq Node \\ Element &\sqsubseteq Node \\ TextNode &\sqsubseteq Node \\ childOf &\sqsubseteq descendant \\ parentOf &\sqsubseteq ancestor \\ childOf &= parentOf^{-} \\ &\quad \mathcal{T}rans(ancestor) \\ ancestor &\sqsubseteq ancestorOrSelf \\ self &\sqsubseteq descendantOrSelf \end{aligned}$$

[1] http://web.sfc.keio.ac.jp/~km/rpath-eng/rpath.html

$$self \sqsubseteq ancestorOrSelf$$
$$self = sameAs$$
$$immediatePrecedingSibling \sqsubseteq precedingSiblinng$$
$$immediateFollowingSibling \sqsubseteq followingSibling$$
$$immediatePrecedingSibling = immediateFollowingSibling^{-}$$
$$\mathcal{T}rans(followingSibling)$$

3. XPATH TRANSLATION TO RDQL

3.1 XPath to RDQL translation algorithm

RDQL [2] is the popular RDF query language from HP Labs Bristol. Each XPath *axis* can be mapped into one or more triple patterns of the target RDQL query. Analogously each *nodetest* and *predicate* can be mapped also with just one or more triple patterns. The output RDQL query always takes the form:

```
SELECT *
WHERE
  (?v1, <rdf:type>, <xmloverrdf:document>)
  [triple pattern 2]
  [triple pattern 3]
  ...
  [triple pattern N]
```

The translation can be deduced from the XPath formal semantics. For example, the *following* axis is described as:

$$A_{following}(x) = \{x_1 \mid x_1 \in A_{descendant-or-self}(x_2) \wedge x_2 \in A_{following-sibling}(x_3)\} \wedge x_3 \in A_{ancestor-or-self}(x)\}$$

So the *following* axis must be translated to:

```
(?vi-2, <xmloverrdf:ancestor-or-self>, ?vi-3)
(?vi-1, <xmloverrdf:following-sibling>, ?vi-2)
(?vi, <xmloverrdf:descendant-or-self>, ?vi-1)
```

3.2 Example results

An example query could be:

```
/child::movies/child::movie/child::title
(in abbreviated form /movies/movie/title)
```

That is translated to:

```
SELECT *
WHERE
        (?v1, <rdf:type>, <xmloverrdf:document>)
        , (?v2, <xmloverrdf:childOf>, ?v1)
        , (?v2, <xmloverrdf:hasName>, "movies")
        , (?v3, <xmloverrdf:childOf>, ?v2)
        , (?v3, <xmloverrdf:hasName>, "movie")
        , (?result, <xmloverrdf:childOf>, ?v3)
        , (?result, <xmloverrdf:hasName>, "title")
```

3.3 Incorporating schema-awareness

To allow the inference-engine to consider the information contained in one or more XML Schemas, we have translated the XML schema language to RDF (XSD/RDF Syntax). We have also defined some axioms expressing the semantics of the different resources and properties of the XSD/RDF Syntax, and also their relationship with elements and attributes.

$$\mathcal{T}rans(fromSubstitutionGroup)$$
$$fromSubstitutionGroup \sqsubseteq hasName$$
$$\mathcal{T}rans(fromType)$$
$$fromType \sqsubseteq hasName$$
$$fromType \sqsubseteq subTypeOf$$

4. IMPLEMENTATION

The work has been materialised in the form of a Java API. We have used the Jena 2 API [3] for RDQL computation and OWL reasoning. To process XPath expressions we have modified and recompiled the Jaxen XPath Processor [4]. An on-line demo can be found at *http://theron.upf.es/contorsion*.

5. CONCLUSIONS

The work contributes to the recent research trend that defines an XML to RDF mapping allowing XML documents interoperate at the semantic level. We solve the node-order limitation with a *model-mapping* approach, allowings a complete mapping of all XPath axis to RDQL. We have developed a Java API that implements this functionality. The obtained XPath processor has some interesting properties like schema-awareness and idref-awareness.

6. REFERENCES

[1] B.Amann et al. Ontology-Based Integration of XML Web Resources. In Proceedings of the 1st International Semantic Web Conference (ISWC 2002),pages 117-131,2002.

[2] Cruz, I., Xiao H., Hsu F. An Ontology-based Framework for XML Semantic Integration. University of Illinois at Chicago. Eighth International Database Engineering and Applications Symposium. IDEAS'04. July 7-9, 2004 Coimbra, Portugal.

[3] A. Y. Halevy et al. Piazza: Data Management Infrastructure for Semantic Web Applications, 12th International World Wide Web Conference, 2003

[4] M.C.A.Klein. Interpreting XML Documents via an RDF Schema Ontology. Proceedings of DEXA 2002,pages 889-894.

[5] L.V.Lakshmanan and F.Sadri.Interoperability on XML Data.In Proceedings of the 2nd International Semantic Web Conference (ICSW 03).

[6] P.F.Patel-Schneider and J.Simeon.The Yin/Yang web:XML syntax and RDF semantics.In Proceedings of the 11th International World Wide Web Conference (WWW2002),pages 443-453.

[7] M. Yoshikawa et al. XRel: A Path-Based Approach to Storage and Retrieval of XML Documents using Relational Databases, ACM Transactions on Internet Technology, Vol. 1, No. 1, June 2001.

[8] Ian Horrocks et al. Practical reasoning for expressive description logics. In Proc. of the 6th Int. Conf. on Logic for Programming and Automated Reasoning (LPAR99), number 1705 in LNAI pages 161180. Springer, 1999.

[2] http://www.w3.org/Submission/RDQL/

[3] http://www.hpl.hp.com/semweb/jena.htm

[4] http://jaxen.org/