Towards the definition of guidelines for RDF and Topic Maps interoperability

Valentina Presutti, Lars Marius Garshol, Fabio Vitali
Steve Pepper and Nicola Gessa

{presutti| vitali| gessa}@cs.unibo.it
Dipartimento di Scienze dell'Informazione, University of Bologna
{larsga| pepper}@ontopia.net
Ontopia AS

Abstract. The Semantic Web relies on the presence of semantic annotations which describe information in a machine readable form. There are two standard formalisms which are suitable for this aim: RDF and Topic Maps.
This paper presents an analysis of the issues that need to be addressed in order to define a set of rules for performing automated and consistent translations between RDF and Topic Maps, i.e. semantic mapping issues. The analysis is based on existing approaches to the RDF and Topic Maps translation in the light of the new official formal models released for both RDF and Topic Maps.

1 Introduction

The Resource Description Framework (RDF) [1] is a model developed by the W3C for representing information about resources in the World Wide Web. Topic Maps [2] is a standard for knowledge integration developed by the ISO. While unification has to date not been possible , a number of attempts have been made to uncover the synergies between RDF and Topic Maps and to find ways of achieving interoperability at the data level. The main purpose of our work is that of defining guidelines for the RDF and Topic Maps interoperability. This paper represents a first step towards that main goal. In particular, through the analysis of the existing approaches for the RDF and Topic Maps interoperability we identify the issues we are going to deal with during the definition of the guidelines.

Topic Maps in a nutshell We assume the reader is already familiar with RDF, but probably many readers are not familiar with Topic Maps, and so we will provide a brief introduction. Topic Maps consist of *topics*, each of which represent some thing of interest, known as *subjects Associations* repesent relationships between two or more topics, have a type (which is a topic) and consist of a set of (role type, role player) pairs (where both elements are topics). *Occurrences* can be either simple property-value assignments for topics, or references to resources

which contain information pertinent to a topic. Occurrences also have a type (which is a topic). In addition, topics may have one or more names (which also may be typed). Further, any construct can be reified (associations, occurrences, roles, and names), and any construct (except roles) may have a *scope*, which is a set of topic representing the context in which the construct is valid. The technology stack consists of a data model [2], an XML interchange syntax [3], a constraint language [4] , and a query language [5].

2 Analysis of existing approaches

We decided to focus on five existing proposals. Each translation proposal has been evaluated against the general criteria of: **Completeness**, and **Naturalness**. All the existing approaches fall into two distinct categories that in [6] are called "object mappings" and "semantic mappings". The two approaches can be summed up as follows: (i) **object mappings** use the low-level building blocks of one language to describe the object model of the other; (ii) **semantic mappings** start from higher level concepts that carry the semantics of each model and attempt to find equivalences between them. The analysis of the options and solutions provided in literature clearly shows the advantages of semantic mapping.

For the sake of brevity a description of the existing approaches is summarized in a table shown in Table 1. [7] describes two proposals which differ in the approach taken, hence it is evaluated twice.

Proposal	**Approach**	**Basis**	**Direction**	**Complete?**	**Natural?**
Moore [7]	semantic mapping	informal	both	no	yes
Moore [7]	object mapping	informal	both	almost	no
Lacher and Decker [6]	object mapping	[8]	TM2RDF	almost	no
Ogievetsky [9]	object mapping	[8]	both	almost	no
Garshol [10]	semantic mapping	[2]	both	almost	almost
UniBo [11]	hybrid	[2]	both	almost	almost

Table 1. Evaluation of existing proposals

It should be noted that [6] and [9] are both based on PMTM4[8], an early data model for Topic Maps that has since been abandoned. [7] predates any published data model for Topic Maps, and so is based on an informal understanding of Topic Maps. The following section describes the result of this analysis.

2.1 Semantic mapping issues

This section gives an informal description of the mapping behaviour we currently think is correct, with notes about currently open issues.

– **Things and proxies**: there is a fundamental equivalence between subjects and resources. This equivalence may be refined as follows: topics are to subjects as RDF nodes (excluding literals) are to resources. Subjects and resources are the *things* (entities, concepts, documents, whatever) about which assertions are made. Topics and RDF nodes (excluding literals) are the corresponding *proxies* that represent subjects and resources within the Topic Maps and RDF models respectively.
– **Identity**: both topics and resources may use URI references (or URIrefs) as identifiers. However, in Topic Maps there are two ways in which a URIref can be used to identify a subject: (i) *directly*, as the actual address (or locator) of the subject, in which case it is called a *subject locator*; (ii) *indirectly*, as the locator of an information resource that provides some human-interpretable indication of the subject, in which case it is called a *subject identifier*. There are problems with all of the analyzed proposals. Equating URIs in RDF with subject locators is problematic in several ways. Firstly it leads to incorrect semantics (as the description of the Unibo proposal shows). Secondly, the result is less natural. Finally, the identifiers of occurrence types and association types (which are typically subject identifiers) could not be used as the URIs of RDF properties. Equating URIs with subject identifiers also yields unnatural results, since the identifier of an addressable subject (i.e., an information resource) will not become the URI of the corresponding resource, as would be most natural in RDF. However, this alternative does not exhibit the other problems that result from favouring subject locators. The ideal solution would be to allow either subject identifiers or subject locators to be regarded as URIs (and vice versa), but at the same time to retain sufficient information when going from Topic Maps to RDF to be able to perform round-tripping. The recognition in [12] tag of the distinction between resources in general and information resources, and the insights in [13], may provide the foundation for such a solution. The issue of multiple identifiers is treated explicitly by [10] only. His proposal to use equivalence properties defined in OWL (i.e., owl:sameAs, owl:equivalentClass, and owl:equivalentProperty) should clearly be investigated in more detail since such an approach is likely to lead to increased interoperability between RDF and Topic Maps.
– **Names**: in RDF the name of a resource is usually represented by a single statement. RDF Schema defines a property for this purpose (rdfs:label) but many vocabularies define their own properties (e.g., `dc:title`, `foaf:name`, etc.). An accurate semantic mapping from Topic Maps can be achieved by translating base names to such properties. Both [14] and [11] take this approach, differing only in that [11] always maps a base name to rdfs:label (and vice versa), while [14] allows base names (including scoped base names) to be mapped to other properties. There are two possible situations going from Topic Maps to RDF: typed names and untyped names. For typed names the solution is to map name types to properties and properties to name types. Instead, there are two basic, alternative approaches to untyped names: (i) always translate untyped base names to the same property, e.g., rdfs:label,

rdftm:name, tm:basename. This approach leads on one hand to simpler mapping, and on the other hand to less natural result; (ii) translate untyped base names to different properties depending on the type of the topic/resource. With this approach we can obtain more natural result, but the mapping is much more complex.

- **Binary relationships**: representations of binary relationships have somewhat different topographies in RDF and Topic Maps. RDF uses a single statement , in which the subject and object represent the two resources that participate in the relationship. In Topic Maps there is no concept of subject and object in a binary association because the association has no direction. The nature of the two participating topics' involvement in the relationship is stated explicitly through their role types. The challenge is complimentary to this when going from Topic Maps to RDF. Both [10] and [11] solve this by allowing additional information to be provided that allows the RDF subject and object to be connected with their respective role types.
- **Non-binary relationships**: one major difference between the models of RDF and Topic Maps is that the latter permits non-binary relationships to be expressed directly: an association may have one, two, or more role players. In RDF on the other hand the base model permits only binary relationships. Most of the existing proposals for translating associations with more than two role-players are unsatisfactory, since they result in a large number of RDF statements. [15] proposes patterns for representing n-ary relations in RDF in which the relation is "re-represented" as a class rather than a property. Each such pattern requires n statements in order to express the relationship. If such patterns are adopted in the RDF community it would seem to be advisable, in the interest of compatibility, to follow them as closely as possible when translating n-ary associations from Topic Maps to RDF.
- **Occurrences**: both [10] and [11] recognize that occurrences are most naturally represented as single RDF statements where the property corresponds to the occurrence type. Internal and external occurrences correspond to statements whose objects are literals and resources respectively. Going from Topic Maps to RDF presents no problems at all; going the other way seems to require additional information in order to distinguish an internal occurrence from a name, and an external occurrence from an association or identifier.
- **Types and subtypes**: [10] and [11] agree on the fundamental semantic equivalence between the concept of type-instance in [2] and rdf:type, on the one hand; and between supertype-subtype and `rdfs:subClassOf` on the other. In addition, association types and occurrence types are regarded as equivalent to RDF properties. Role types present particular problems, as discussed above, and name types, as already noted, did not exist at the time the proposals were written.
- **Reification**: only [10] and [11] mention reification and neither proposal regards it as being problematic. In actual fact, [11] only talks explicitly about the reification of associations, while [10] mentions reified names, occurrences,

and associations. Neither proposal covers the reification of topic maps and association roles.
- **Scope**: the concept of scope is peculiar to Topic Maps and has been regarded as one of the major stumbling blocks for RDF/Topic Maps interoperability. All the existing proposals discuss the issue in one form or another but only [10] and [11] do so in terms of its semantics, i.e., as a way to express the contextual validity of an assertion. [10] makes the point that scope is most properly regarded as a special kind of assertion made about another assertion. Since assertions about assertions are handled through reification in both paradigms, and reification translates rather easily, Garshol proposes to translate scope using reification together with a property that captures the semantics of contextual validity. [10] treats scoped base names as a special case, however, and allows a base name in a particular scope to be translated to a specific property. Given that the forthcoming revised Topic Maps standard will permit typed names, it would be possible to avoid treating scoped names as a special case and still obtain natural results.
- **Other issues**: none of the existing proposals discuss how to represent RDF containers and collections, language tags, XML literals or typed literals in Topic Maps. Of these issues, the latter two are addressed by recent datatyping extensions to the Topic Maps model. Language tagging can be seen as a kind of contextual information akin to scope and treated accordingly. Containers and collections may or may not require special treatment. Since they are expressed using the fundamental building blocks of RDF (nodes and arcs), they may be represented using associations in Topic Maps. The semantics would not be lost and could be recovered when round-tripping. However, they would not be "visible" in terms of some equivalent Topic Maps construct.

3 Conclusion and future work

The main result of this work has been to identify a number of outstanding issues that need to be considered when providing an approach for the interoperability between RDF and Topic Maps. The immediate next step is that of providing for each issue described in section 2.1 a set of mapping rules that allow to perform automated translation between RDF and Topic Maps, to translate data from one form to the other without unacceptable loss of information or corruption of the semantics. Furthermore, it should also be possible to query the results of a translation in terms of the target model and it should be possible to share vocabularies across the two paradigms. The overall approach we are following is based on the definition of a specific vocabulary for the mapping. The guidelines will provide rules for performing both guided and unguided translations.

References

1. Resource Description Framework (RDF) Family of Specifications. http://www.w3.org/RDF, 2004.

2. L. M. Garshol and G. Moore. ISO/IEC 13250-2: Topic Maps – Data Model, 2005. http://www.isotopicmaps.org/sam/sam-model/.
3. Members of the Topic Maps.Org Authoring Group. Xml topic maps (xtm 1.0). Technical report, ISO, 2001. http://www.topicmaps.org/xtm/1.0/.
4. Moore G., Bogachev D., and Nishikawa M. Topic maps constraint language. Technical report, TopicMaps.org, 2005. http://www.isotopicmaps.org/tmcl/tmcl-2005-02-12.html.
5. Garshol L. M., Barta R., and Nishikawa M. Topic maps query language. Technical report, ISO, 2005. http://www.isotopicmaps.org/tmql/spec.html.
6. Lacher M and Decker S. On the integration of Topic Map data and RDF data. In *Proccedings of the Extreme Markup Languages 2001 Conference*, Montreal, Canada, 2001.
7. Moore, G. RDF and Topic Maps: An Exercise of Convergence. In *Proceedings of the XML Europe 2001 Conference*, Berlin, Germany, 2001.
8. Biezunski M. and Newcomb S. R. Topicmaps.net's Processing Model for XTM. http://topicmaps.net/pmtm4.htm, 2001.
9. Ogievetsky N. XML Topic Maps through RDF glasses. In *Proceedings of the Extreme Markup Languages 2001 Conference*, Montreal, Canada, 2001.
10. Garshol L.M. Living with RDF and Topic Maps. http://www.ontopia.net/topicmaps/materials/tmrdf.html, 2003.
11. Ciancarini P., Gentilucci R., Pirruccio M., Presutti V., and VitaliF. Metadata on the web: On the integration of RDF and Topic Maps. In *Proc. Conf. on Extreme Markup Languages*, page (to appear), Montreal, Canada, August 2003.
12. Jacobs I. and Walsh N. Architecture of the World Wide Web. Technical report, W3C, 2004. http://www.w3.org/TR/webarch.
13. Pepper S. and Schwab S. Curing the Web's Identity Crisis: Subject Indicators for RDF. Technical report, Ontopia, 2003. http://www.ontopia.net/topicmaps/materials/identitycrisis.html (2003).
14. Garshol L. Topic maps, RDF, DAML, OIL: A Comparison. http://www.ontopia.net/topicmaps/materials/tmrdfoildaml.html, 2001.
15. Noy N. and Rector A. Defining N-ary Relations on the Semantic Web: Use With Individuals. Technical report, W3C, 2005. http://www.w3.org/TR/swbp-n-aryRelations/ (2004).