Ontology Mapping to support multilingual ontology-based question answering

Maria Teresa Pazienza(*), Armando Stellato(*)
Lina Henriksen(+), Patrizia Paggio(+), Fabio Massimo Zanzotto(#)

(*) AI Research Group, Dept. of Computer Science, Systems and Production
University of Rome, Tor Vergata
Via del Politecnico 1, 00133 Rome, Italy
{pazienza, stellato }@info.uniroma2.it

(+) Center for Sprogteknologi,
University of Copenhagen
Njalsgade 80
2300-DK Copenhagen S,
{lina,patrizia}@cst.dk

(#) DiSCo
University of Milano Bicocca
Via Bicocca degli Arcimboldi, 8
20126 Milano (Italy)
zanzotto@disco.unimib.it

Abstract

In this paper we report results obtained by using the XeOML mapping language to harmonize two ontologies developed as a support for a multilingual, distributed question answering system for two university domains. The test bed has allowed us to test the various possibilities offered by the language to express simple and complex mappings that reflect different levels of similarities between classes and properties of the two ontologies. Inference rules were also created to expand the coverage of the initial manually defined mapping document. The non-trivial nature of the study is backed up by overall quantitative information on the nature of the mapping resource produced.

2 The MOSES Scenario

Our approach to ontology mapping has being developed in the context of the EU project MOSES IST-2001-37244. The project's overall objective was to develop an ontology-based methodology to create, maintain, search and adapt semantically structured Web contents in a federation of sites (Atzeni et al, 2004) (Paggio et al. 2004).

As a test bed the project has produced an agent-based knowledge management system and an ontology-based search engine that accept questions and produce answers in natural language for the Web sites of the two European universities of Roma Tre and Copenhagen. The specific purpose of ontology mapping is to support question answering across the two nodes, each of which is equipped with its own domain model.

In this scenario, a user submits questions to the system from one of the two sites in the language pertaining to this site, and indicates whether the question is to be interpreted as a single-node one – to be processed locally – or a federated one. In the case of a federated question, access to all the nodes in the federation is supported. After having analyzed the input by means of a local linguistic analyzer, the local content agent issues a mapping request to the agents responsible for the remote nodes to get the query mapped onto the remote ontologies and processed by the remote content agents. The content retrieved from the remote nodes is then combined with the relevant local content and presented to the user, e.g. as shown – in English for readability's sake – below.

> Q: Who teaches history?
> The following answers are found on the KU site:
> **"Lars Hansen underviser i historie"**
> The following answers are found on the ROMA III site:
> **"Paolo Rossi insegna storia dell'arte"**

2.1 MOSES Ontologies

The ontologies underlying the Danish and Italian nodes of the MOSES test bed build on the "university ontology" from the DAML-OIL ontology library. Several changes and extensions were necessary to cover the two university sub-systems and to adapt the ontologies to the OWL implementation of the Topic Maps formalism (Garshol, 2003) used by the MOSES content management system. Instances have been created by the project's user groups by downloading them from existing databases or manually extracting data from the Web pages.

Having to deal with two separate ontologies for the same domain poses several challenges. Firstly, classes and relations in the two ontologies are labeled in different languages. Secondly, there are structural differences: not all the nodes in one ontology are represented also in the other and vice-versa; moreover, domain relations are treated differently. In the Italian ontology, all relations are binary in keeping with the original DAML-OIL model, whereas the Danish ontology makes use of n-ary relations in the spirit of the Topic Maps formalism. Finally, even when a mapping can reasonably be defined between classes and properties of the two ontologies, this may be at the cost of hiding differences intrinsic to the cultural and social aspects of the two academic worlds. Thus to map instances that appear only superficially similar, the set-oriented semantics of OWL mapping primitives had to be extended.

	Italian Ontology						Danish Ontology				
	Source	MD	D-Imp	A-Imp	Imp		Source	MD	D-Imp	A-Imp	Imp
Classes (*Subject*)	94	50	55	86	89		152	54	67	146	147
Classes (*Document*)	48	17	19	27	29		22	17	17	20	20
Classes (*Relations*)	47	24	X	X	X		36	18	X	X	X
ObjectProperties	95	52	X	X	X		58	37	X	X	X
DatatypeProperties	18	4	X	X	X		10	3	X	X	X

Table 1: Statistics of the MOSES mapping

3 Mapping the ontologies

The mapping between the two ontologies has been expressed using XeOML (Pazienza et al., 2004), a language for ontology mappings which has been developed at the University of Rome, Tor Vergata.

The principle beyond XeOML is to define a layered view in which the complex logical-algebraic aspects of the ontology mapping task coexist (but are neatly separated) with the semantics describing the nature and quality of the identified correspondences. In XeOML a neutral ontology structure is adopted to achieve interoperability between different representation languages all of which share an object oriented approach to knowledge modeling. This approach allows for a mediation activity between independent agents, since each agent needs only know about the knowledge model adopted to express the ontological resource without necessarily understanding the model owned by the interlocutor.

Mappings are conceptually divided into two categories: *Simple Mappings*, which define one-to-one relations between ontology elements of the same type, and *Complex Mappings*, which link heterogeneous objects. Complex mappings vary in the nature of the concepts involved and in the operations they apply over them, e.g.:

- *restrictions* on classes/associations applied to the range of their attributes
- *aggregations* (on an extensional basis) of multiple classes/associations
- *transformations* between heterogeneous object structures: properties into relations (reified through use of classes and properties), classes into instances.
- *join* of relations upon common roles

In addition, it is possible to express similarity between different mapped conceptual structures. For instance, there is no direct equivalent of the Italian class `Professore Associato` in the Danish ontology. However to answer a federated question involving this notion, the system should resort to the Danish class `Lektor`, which describes a similar type of position in the Danish academic career. The mapping cannot be expressed through an OWL same-as primitive, as this is not a problem of translation between two identical objects. Furthermore, the fact that a similar rather than equivalent concept has been used should be reported to the user.

1.2 XeOML reasoning

Our approach to ontology mapping aims at reducing the cost related to manual production of the mapping relations. Inference rules for deriving relations which are not explicitly mentioned in an MD (mapping document) have thus been defined. This reasoning is limited to the information asserted in the MD, while resolution of complex expressions related to the two domains is delegated to local reasoners, in the spirit of a modular and distributed environment.

Table 1 expresses some quantitative information on the resulting mapping experiment, reporting – for each category of ontological elements – the number of objects which are present in each of the two ontologies (*Source* column), the number of elements which needed to be made explicit in the mapping document (*MD*), the number of objects which can be mapped by retrieving bags of more/less specific concepts from the other ontology (D-Imp, A-Imp), and the total number of concepts retrieved through reasoning on documented mappings (Imp).

References

[Atzeni *et al.*, 2004] P. Atzeni, R. Basili, D. H. Hansen, P. Missier, P. Paggio, M. T. Pazienza, F. M. Zanzotto. Ontology-based question answering in a federation of university sites: the MOSES case study. *9th International Conference on Applications of Natural Language to Information Systems* (NLDB'04) Manchester (United Kingdom).

[Garshol, 2003] L.M. Garshol. Living with Topic Maps and RDF. Technical report, 2003. http://www.ontopia.net/topicmaps/materials/tmrdf.html.

[Paggio *et al.*, 2004] P. Paggio, D. H. Hansen, R. Basili, M. T. Pazienza, F. M. Zanzotto. 2004. Ontology-based question analysis in a multilingual environment: the MOSES case study. *Workshop OntoLex 2004, held jointly with LREC 2004* LISBON, Portugal.

[Pazienza et al., 2004] M. T. Pazienza, A. Stellato, M. Vindigni, F. M. Zanzotto. XeOML: An XML-based extensible Ontology Mapping Language. Workshop on Meaning Coordination and Negotiation, held in conjunction with 3rd International Semantic Web Conference Hiroshima, Japan. 2004