Enriching ontologies with multilingual information

Guadalupe Aguado de Cea[1], Asunción Gómez-Pérez[1], Elena Montiel-Ponsoda[1]

[1] Ontology Engineering Group, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, 28660 Boadilla del Monte, Madrid
{lupe, asun, emontiel}@fi.upm.es

Abstract. Multilinguality in ontologies has become an impending need for institutions worldwide that have to deal with data and linguistic resources in different natural languages. Since most ontologies are developed in one language, obtaining multilingual ontologies implies to *localize* or adapt them to a concrete language and culture community. As the adaptation of the ontology conceptualization demands considerable efforts, we propose to modify the ontology terminological layer by associating an external repository of linguistic data to the ontology. With this aim we provide a model called *Linguistic Information Repository* (LIR) that associated to the ontology meta-model allows terminological layer localization.

Keywords: multilingual ontologies, Ontology Localization, Linguistic Information Repository (LIR)

1 Introduction

Multilinguality in ontologies is nowadays demanded by institutions worldwide with a huge number of resources in different languages. One of these institutions is the FAO[1]. Within the NeOn project[2], the FAO is currently leading a case study on fishery stocks in order to improve the interoperability of its information systems. The FAO, as an international organization with five official languages -English, French, Spanish, Arabic and Chinese- deals with heterogeneous and multilingual linguistic resources with different granularity levels. This scenario is an illustrative example of the need for semantically organizing great amounts of multilingual data. When providing ontologies with multilingual data, one of the activities identified in the NeOn ontology network development process is the Ontology Localization Activity, that consists in *adapting an ontology to a concrete language and culture community*, as defined in [1]. In particular, our aim is to obtain multilingual ontologies by localizing its terminological layer (terms or labels that name ontology elements), rather than modifying its conceptualization. Thus, we propose to link ontologies with a linguistic model, called *Linguistic Information Repository* (LIR), whose main feature is that it provides (1) a complete and complementary amount of linguistic data

[1] http://www.fao.org/
[2] http://www.neon-project.org/

that allows localization of ontology elements to a specific linguistic and cultural universe, and, (2) a unified access to aggregated multilingual data.

2 Related Work

Regarding the activity of Ontology Localization, we have identified three ways of modelling multilinguality in ontologies: 1) inclusion of multilingual labels in the ontology by means of the `rdfs:label` and `rdfs:comment` properties (most widespread modality), 2) mapping of several conceptualizations in different natural languages through an interlingual set of common concepts (as in EWN[3]), and 3) association of an external linguistic model to the ontology (as in LingInfo [2]). The first modality option restricts the amount and type of linguistic information that can be associated to the ontology. The second option requires a huge effort at two stages: first, when a new language has to be integrated in the multilingual system, since a new conceptualization has to be developed, and second, by the establishment of alignments among conceptualizations or between the new conceptualization and the interlingua. In this way, our hypothesis is that the best solution lies on the third option, in which the type and quantity of linguistic information is not restricted, and the linguistic elements that compose the model can be related among them. Regarding this latter option, we argue that existing models have not been intended to cover localization needs, and do not include enough information in this sense, but rather focus on other linguistic information such as the morphosyntactic aspects of ontology labels.

3 Proposed Approach

With the aim of providing available ontologies in one natural language with multilingual information thus allowing their localization, we have designed the LIR [3,4,5] within the NeOn project. The LIR is an external linguistic model based on existing linguistic (LMF[4]) and terminological (TMF[5]) representation schemas. The LIR permits the association of a set of linguistic data to any element in the ontology. The main classes or data categories that compose the LIR are: *LexicalEntry*, *Lexicalization*, *Sense*, *Definition*, *Language*, *Source*, *Note*, and *Usage Context* (as can be seen in Figure 1). Thanks to the relations that can be established among the LIR classes, the LIR mainly accounts for: well-defined relations within lexicalizations in one language and across languages, and conceptualization mismatches among different cultures and languages. The main benefits of this approach against the modeling options presented in section 2 are the following: a) the association of an unrestricted quantity of linguistic information to ontology elements; b) the establishment of relations among the linguistic elements, as well as the performance

[3] http://www.illc.uva.nl/EuroWordNet/
[4] Lexical Markup Framework ISO/CD 24613
[5] Terminological Markup Framework ISO 16642

of complex operations (reasoning) with them; c) the access and manipulation of the linguistic data (terminological layer) without interfering with the conceptualization, with the resulting benefits for non-ontology engineers; and d) the reuse of the contained linguistic information for other applications.

Up to now, the LIR has been implemented as an ontology in OWL [4], and is currently supported by the LabelTranslator NeOn plug-in [6] for an automatic localization of ontologies. A first set of tests has been conducted within NeOn to asses the suitability of the LIR model for the linguistic needs of the FAO. The LIR has proved to satisfy the FAO needs for i) establishing relations among lexicalizations within and across languages, ii) specifying variants for dialects or local languages, and iii) explicitly expressing translation specificities.

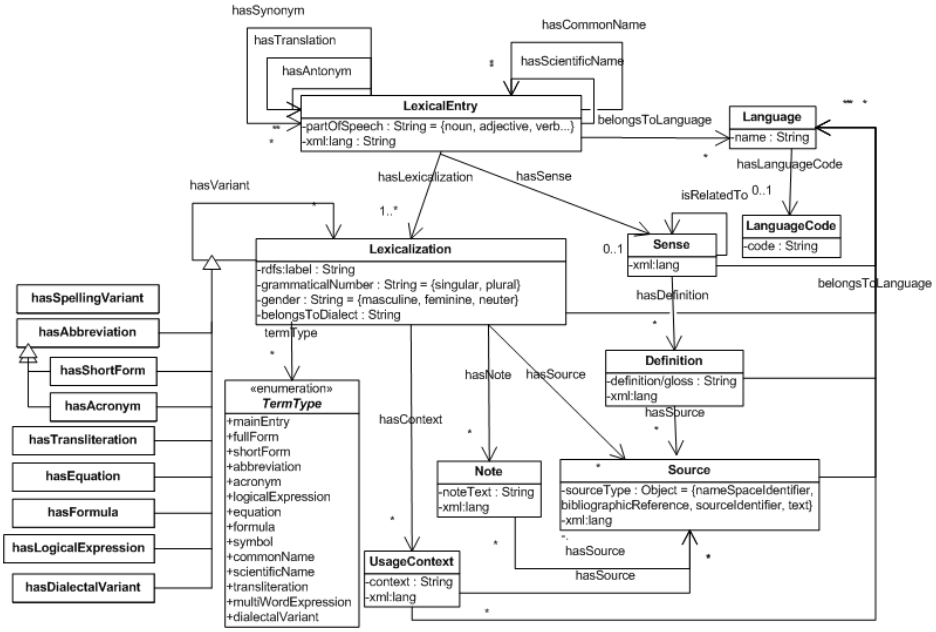

Figure 1. The LIR Model

Main References

1. Suárez-Figueroa, M.C. and Gómez-Pérez, A. *First Attempt towards a Standard Glossary of Ontology Engineering Terminology*. 8th International Conference on Terminology and Knowledge Engineering (TKE2008), Copenhagen (2008)
2. Buitelaar, P., M. Sintek, M. Kiesel. 2006. *A Multi-lingual/Multimedia Lexicon Model for Ontologies*. In Proc. ESWC'06, Budva, Montenegro (2006)
3. Montiel-Ponsoda, E., Aguado de Cea, G., Gómez-Pérez, A., and Peters, W. Modelling multilinguality in ontologies. In *Coling 2008: Companion volume - Posters and Demonstrations*, Manchester, UK (2008)

4. Montiel-Ponsoda, E. and Peters, W. (coordinators): Multilingual and Localization Support for Ontologies. NeOn Project Deliverable 2.4.2 (2008)

5. Peters, W., Montiel-Ponsoda, E., Aguado de Cea, G. *Localizing Ontologies in OWL*. In Proc. OntoLex, Busan, South Corea (2007)
6. Espinoza, M., Gómez-Pérez, A., Mena, E.: Enriching an Ontology with Multilingual Information. In Proc. of ESWC'08, Tenerife (Spain), LNCS Springer, ISBN 978-3-540-68233-2, ISSN-0302-9743, pp. 333--347 (2008)