Multilingual and Localization Support for Ontologies

Mauricio Espinoza, Asunción Gómez-Pérez and Elena Montiel-Ponsoda

UPM, Laboratorio de Inteligencia Artificial, 28660 Boadilla del Monte, Spain
{jespinoza, asun, emontiel}@fi.upm.es

Abstract. This demo proposal aims at providing support for the localization of ontologies, and as a result at obtaining multilingual ontologies. We briefly present an advanced version of LabelTranslator, our system to localize ontology terms in different natural languages. The current version of the system differs from previous works reported in [1, 2] in that it relies on a modular approach to store the linguistic information associated to ontology terms. Additionally, it uses a synchronization method to maintain the conceptual and linguistic information updated.

Keywords: Ontology localization, Multilingual ontologies.

1 Introduction

Multilinguality in ontologies has become an impending need for institutions worldwide with valuable 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 defined in [5, 6]. For this end, we designed and implemented LabelTranslator [1, 2], a system that suggests translations of ontology labels among English, Spanish and German, with the purpose of building a multilingual ontology.

The current trend in the integration of multilinguality in ontologies suggests the suitability of keeping ontology knowledge and linguistic (multilingual) knowledge separated and independent. However, the first version of LabelTranslator followed a non-modular approach, in which multilingual information was embedded in the ontology by means of the RDFS/OWL predicates. This approach has important limitations related to the restricted amount of linguistic information that can be attached to ontology concepts. Furthermore, multilingual information is limited to strings without information about senses in their respective languages, nor about provenance of the information, which makes concept localization to different natural languages quite difficult.

In this paper we describe the features and design aspects of the second version of the LabelTranslator system. The current version of our tool supports the Linguistic Information Repository (LIR) [3, 4] model, which follows the current trend in the integration of linguistic information in ontologies. The LIR model consists of a set of linguistic classes, whose nature accounts for the localization of ontology terms in a certain language.

The rest of this paper is structured as follows. In Section 2 we briefly describe the LIR model designed for the representation of multilingual information in ontologies. Then, a comprehensive solution to the problems of managing the conceptual knowledge and the linguistic knowledge by means of synchronization techniques can be found in Section 3.

2 Ontology Enrichment by means of the LIR model

The first version of our tool used some fields of the NeOn ToolKit[1] for storing the multilingual information generated by the system. The link we established between the terms in the ontology and its associated translations was characterized by simple references between concepts and labels (as offered by the standard owl:comment and rdfs:label properties). However, for the second version of LabelTranslator we decided to go for a modular approach, in which the conceptualization is kept apart from the multilingual information.

With this purpose in mind, an ontology of linguistic and terminological concepts has been developed within the NeOn project in order to capture relevant linguistic information that linked to domain ontologies enables their localization. The LIR is a portable model that can be associated to any element of an OWL ontology by means of an OWL meta-ontology. The main classes that compose the LIR (Lexicalization, Sense, Definition, UsageContext, Note and Source) are organized around the LexicalEntry class, which is related to any ontology element (by means of the hasLexicalEntry relation). The set of LIR concepts enables a complete description in natural language of the ontology element they are associated to. Additionally, by means of typical lexical relations within the same language (e.g. hasSynonym) and across languages (hasTranslation), the LIR organizes linguistic information within and across languages with the aim of enabling ontology localization. The LIR provides a selection of ISO standard classes for describing linguistic and terminological knowledge, ensuring in this way its eventual extension to cover other types of linguistic data (e.g. morphological or syntactic description), if required.

When the ontology user creates/imports a new OWL ontology in Neon, the LabelTranslator system automatically builds an empty linguistic model associated to the ontology under consideration. This model is used by our system to store the linguistic information associated with each ontology term. Figure 1 shows the *Linguistic Information* page associated to the sample ontology term *FAO*. LabelTranslator fills in runtime the majority of fields of the linguistic page, according to the information obtained by the system in the translation process. A more detailed description of the translation process can be found in [1].

Initially, the linguistic information page shows five sections that correspond to the lexical entries of the selected ontology element (*FAO* in our example) and the associated information of each lexical entry: *lexicalizations*, *lexical entry senses*, *usage contexts*, *sources*, and *notes*. In the example given in figure1 we

[1] http://www.neon-toolkit.org/

illustrate also part of the information supported by the LIR model. Thus, three lexical entries (LexicalEntry-1,LexicalEntry-2, and LexicalEntry-3) are associated with the same concept (FAO:Class), which means that they are all terms that identify one and the same concept. Two lexical entries (LexicalEntry-1 and LexicalEntry-2) belong to the same Language (English), whereas the third lexical entry (LexicalEntry-3) belongs to Spanish. The two English lexical entries are considered synonyms, and translations of the Spanish lexical entry. This information is shown in the field *Lexical Entry Relationships*.

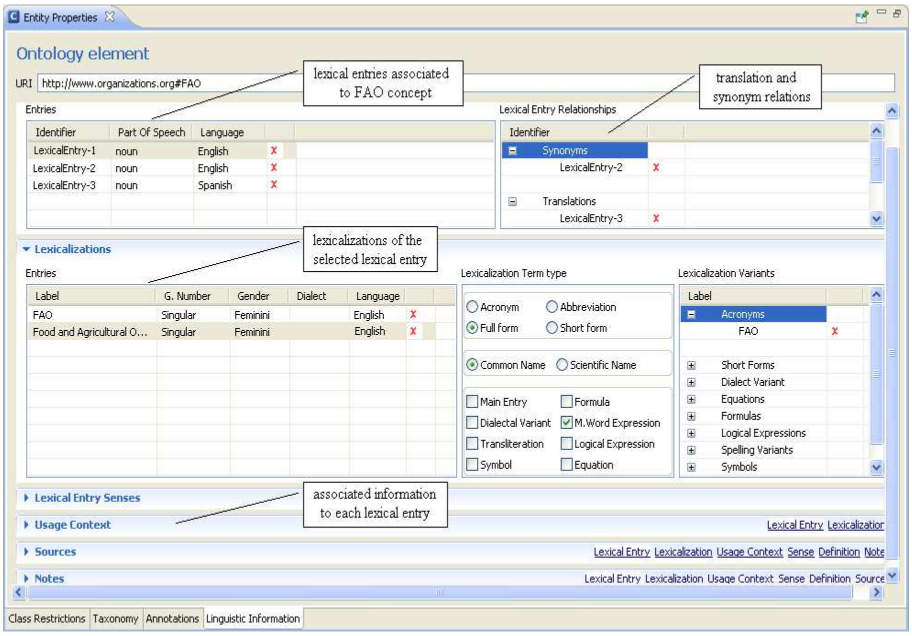

Fig. 1. Linguistic Information page with data of the concept FAO.

Of course, every time that the user chooses a new entry, the interface automatically displays the information correlated in the different sections. Thus, in our example LexicalEntry-1 includes two lexicalizations whose labels are FAO and Food and Agriculture Organization, respectively. Food and Agriculture Organization has like acronym FAO, and, moreover, it is considered a *common name* (in opposition to scientific name) and a *multi-word expression*. This information is shown in the fields *Lexicalization Term Type* and *Lexicalization Variants*.

3 Localization Management

In this section we present a comprehensive solution to the problems of managing changing on the ontology model (OM) and propagating these changes to linguistic model (LM). In order to keep both models synchronized we first need to find out exactly what has been changed in the ontology model, then find the equivalent places in the linguistic model, and only then start the updating.

Addition of new terms in the ontology, or deletion of an existing term can be controlled by some mechanism of change synchronization. In the NeOn ToolKit, an advanced change tracking, based on Resource Delta[2], is able to capture changes even when ontological terms have changed their position within the ontology model. By adopting this feature, our system can accurately identify the minimal set of changes needed to adjust the structure of the linguistic model, a critical first step to ensure that a matching change is made in the localized ontology.

Simplifying localization management using synchronization. LabelTranslator provides a model where sets of ontology terms and linguistic information associated (in different languages) are separately stored. Therefore, it would be very difficult for a person to update/review all the linguistic definitions and information associated with a particular concept. We believe that this process will be done by different people at different times and in different countries. Thus, the maintenance cycle for each language will often be separated. Figure 2 illustrates the localization management used in our system to synchronize the conceptual and linguistic information. In the following we analyze the process in more detail, describing the actions performed by each actor of our scenario.

- ***Ontology expert.*** (S)he is responsible for editing the changes in the ontology model. All the changes executed in each user session are stored in a repository as a new version. The types of changes that our system can manage are the following: changes of the label content (e.g., ontology label rename) and ontology structure changes (e.g., delete or add operations). For each case, LabelTranslator stores the type of operation executed and its additional information (e.g., the name of the renamed label). This information is used in our system to synchronize the conceptual and linguistic information.
- ***Linguist expert(s).*** The linguist expert in a specific target language is responsible for performing the localization process. Notice that this process always uses the last version of an ontology. When the linguist needs to update the linguistic model (LM), our system tries to synchronize both models, performing the following actions: (1) obtaining the current version of the LM to be updated, (2) extracting the last version of the changes in the ontology model (OM) from which the last localization was taken (normally the one with the same number as the LM), (3) performing all the actions

[2] A resource delta represents changes in the state of a resource tree between two discrete points in time

of the file of changes in the LM, and (4) updating the LM version in the repository.

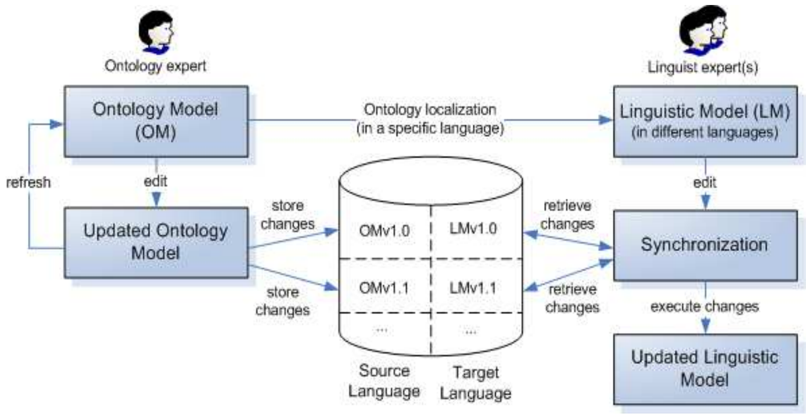

Fig. 2. Synchronization of ontology and linguistic model.

Acknowledgements

This work is supported by the European Commission's Sixth Framework Program under the project name: Lyfecycle support for networked ontologies (NeOn) (FP6-027595), and the National Project "GeoBuddies" (TSI2007-65677C02).

References

1. M. Espinoza, A. Gómez-Pérez, and E. Mena. Enriching an ontology with multilingual information. In *Proc. of 5th European Semantic Web Conference (ESWC'08), Tenerife, (Spain), June 2008.*
2. M. Espinoza, A. Gómez-Pérez, and E. Mena. Labeltranslator - automatically localizing an ontology. In *Proc. of 5th European Semantic Web Conference (ESWC'08), Tenerife, (Spain), June 2008.*
3. E. Montiel-Ponsoda, G. Aguado, A. Gómez-Pérez, and W. Peters. Modelling multilinguality in ontologies. In *Coling 2008: Companion volume - Posters and Demonstrations, Manchester, UK*, 2008.
4. E. Montiel-Ponsoda, W. Peters, G. Aguado, M. Espinoza, A. Gómez-Pérez, and M. Sini. Multilingual and localization support for ontologies. In *NeOn Project Deliverable 2.4.2 (August 2008).*
5. M. C. Suárez-Figueroa and A. Gómez-Pérez. First attempt towards a standard glossary of ontology engineering terminology. In *Proc. of 8th International Conference on Terminology and Knowledge Engineering (TKE'08)*, 2008.
6. M. C. Suárez-Figueroa and A. Gómez-Pérez. Towards a glossary of activities in the ontology engineering field. In *Proc. of 6th International Conference on Language Resources and Evaluation (LREC'08)*, 2008.