LabelTranslator - A Tool to Automatically Localize an Ontology

Mauricio Espinoza[1], Asunción Gómez Pérez[1], and Eduardo Mena[2]

[1] UPM, Laboratorio de Inteligencia Artificial, 28660 Boadilla del Monte, Spain
`asun@fi.upm.es, mespinoza@delicias.dia.fi.upm.es`
[2] IIS Department, Univ. of Zaragoza, María de Luna 1, 50018 Zaragoza, Spain
`emena@unizar.es`

Abstract. This demo proposal briefly presents LabelTranslator, a system that suggests translations of ontology labels, with the purpose of localizing ontologies. LabelTranslator takes as input an ontology whose labels are described in a source natural language and obtains the most probable translation of each ontology label into a target natural language. Our main contribution is the automatization of this process, which reduces human efforts to localize manually the ontology.

Keywords: Ontology localization, Multilingual ontologies.

1 Introduction

Typically, the names assigned to ontology terms appear in a specific natural language. Thus, in order to achieve a more generally applicable ontology, it is necessary to guarantee that the same knowledge be recognizable in different natural languages. Moreover, some organizations working in a multilingual environment demand multilingual ontologies.

To cope with these problems we present LabelTranslator, a system that automatically localizes ontologies in English, Spanish, and German. The Ontology Localization Activity (OLA) consists in adapting an ontology to a concrete language and culture community, as defined in [3]. The technological background of the system comprises: 1) the extraction of possible single-lemma translations from semantic and translation resources, 2) the disambiguation of the translation senses, and 3) the ranking of the translations.

Some details on the aspects above enumerated are given in the following section. For a more comprehensive description of our system, we refer to [2].

2 System Description

The current version of LabelTranslator has been implemented as a Neon[3] plugin, but it can easily become an independent module, for example a web service. Figure 1 illustrates the main steps given by our system and their executing

---

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

sequence to localize an ontology among different natural languages. Notice that, the steps two and three are executed sequentially, then, the system enforces that the steps four and five be finished to continue its execution.
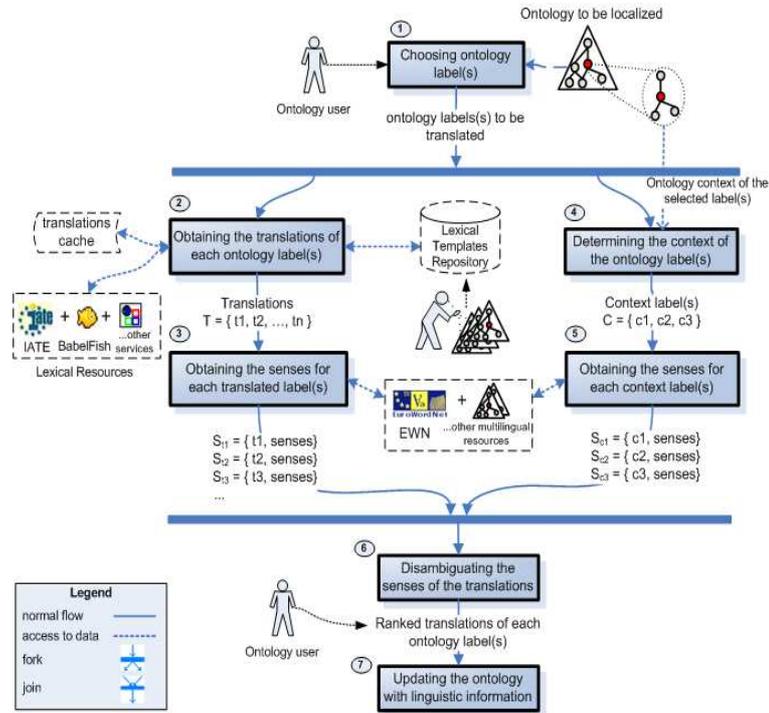

Fig. 1. Main steps of LabelTranslator.

1. *Choosing ontology terms(s) to localize.* The system starts with an ontology (described) in OWL-DL, F-Logic, or RDF provided by the ontology user. The system uses some views of the Neon ToolKit to load the ontology and store the multilingual results, respectively. In Figure 2, we show a screenshot of both the *Ontology Navigator* and the *Entity Properties* view with information related to our sample ontology[4].

   When the system imports a new ontology, it offers a perspective over ontological data using the NeOn ToolKit-style. Then, the user chooses the label of the ontology term(s) to be translated. In our example, the concept *planta* (plant). By right clicking on a frame (concept, attribute, or relation), the *Translate* action performs the translation of an ontology label (see Figure 2).

---

[4] The sample ontology used here belongs to biosphere domain and it needs to be localized from Spanish into English.
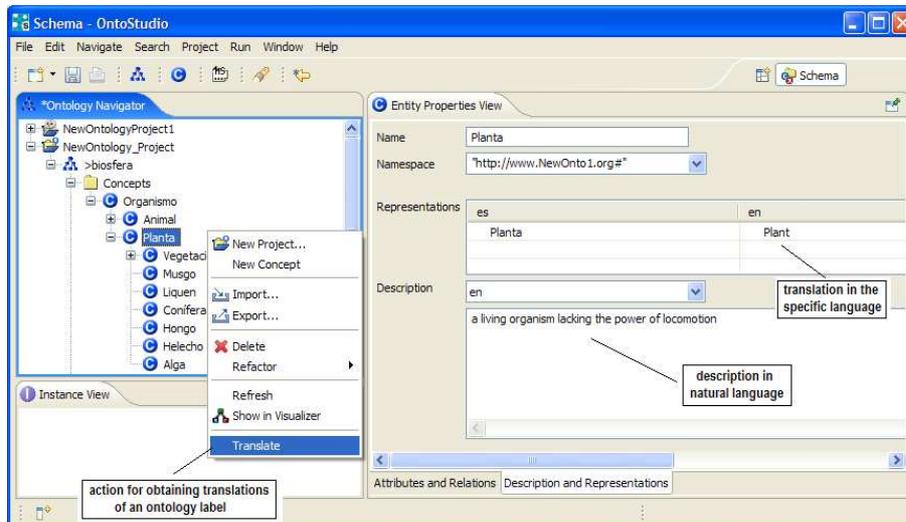
Fig. 2. A screenshot of the NeOn Toolkit views used by the LabelTranslator plug-in.

2. *Obtaining the translations of each ontology label(s).* To obtain the translations of the selected ontology label(s), the system relies on different linguistic resources: 1) multilingual translation web services such as Babelfish[5], GoogleTranslate[6], etc., and 2) other lexical resources as IATE[7]. The current prototype supports translations from/into English, Spanish, and German. A *cache* stores previously translations to avoid accessing the same remote data twice. In our example the system retrieves 12 translations for the selected label "planta". Thus, $T = \{$ sole, flora, plant, mill, story, level, plant life, ..., factory, manufactory $\}$.

   In the case of compound labels where no entry is present in the translation resources used, the label is split into its components; the individual components are translated and then combined into a compound label the target language. The original order of each component is considered to combine components respecting the word order of the target language using lexical templates[8]. The learning of the lexical templates was semi-automatically derived from the patterns found in the labels used in ontologies of different domain and described in English, Spanish, and German. A more detailed description of this process can be found in [2].

3. *Obtaining the senses for each translated label(s).* For each translated label the system retrieves a list of semantic senses using the approach proposed in

---

[5] http://babelfish.altavista.com/

[6] http://www.google.com/translate_t

[7] http://iate.europa.eu/iatediff/SearchByQueryLoad.do?method=load

[8] The notion of lexical template proposed in this paper refers to text correlations found between a pair of languages.

a previous work [6]. Our system takes as input a list of words (each translated label), discovers their semantics in run-time and obtains a list of senses extracted from different ontology pools; it deals with the possible semantic overlapping among senses. These senses are used by the *disambiguation method* (see step 6) to sort the different translations of an ontology label according to similarity with its lexical and semantic context.

4. *Determining the context of the ontology label(s).* In order to determine the context of an ontology label, the system retrieves the set of labels associated with the label under consideration. The list of context labels comprises a set of names which can be direct label names and/or attributes label names, depending on the type of term that is being translated. The number of context labels is limited to those labels with the higher values of similarity according to the Normalized Google Distance [1] (NGD). NGD measures the semantic relatedness between any two terms, considering the relative frequency in which two terms appear in the Web within the same documents. Those labels that have the higher values of similarity with the label under consideration are chosen. In our example the system selects only three context labels, out of eight previously found. Thus, $C = \{$ *organismo(organism), liquen(lichen), conífera(conifer)* $\}$.

5. *Obtaining the senses for each context label(s).* The senses of each context label are discovered with the same process used to discover the senses of each translated label (see step 3).

6. *Disambiguating the senses of the translations.* A disambiguation process is needed to sort the list of translations according to similarity with the context of the label to translate. This method takes as input the set of senses of each translated label (step 3) and the senses of their context labels (step 5). From this set of senses, the method relies on a relatedness measure [4] based on glosses to disambiguate the translations. Candidate translations are then ranked according to the similarity with their context in the ontology, and the ranked list is used to either present the user the best candidates first, or to use the highest-scoring candidate to translate the label automatically.
Figure 3 shows a screenshot of the translation dialog with the translations of the sample ontology label "planta". In our example, the system correctly suggests "plant" in the sense of "a living organism lacking of power of locomotion" as first translation of the selected ontology label.

7. *Updating the ontology with linguistic information.* Once the right sense has been selected, the system updates the linguistic information of the ontological term. Our system supports the linguistic model [5] designed for the representation and structuration of multilingual information in ontologies. In the current version, the system fills in runtime the fields *representations* and *description* of the Entity Properties view (shown in Figure 2), according to the most probable translations proposed by the disambiguation method. These fields represent the link between the conceptual knowledge and the linguistic information discovered. In a future release (current under development), we will add support to the new linguistic model which captures all the relevant linguistic/terminological information associated with concepts.
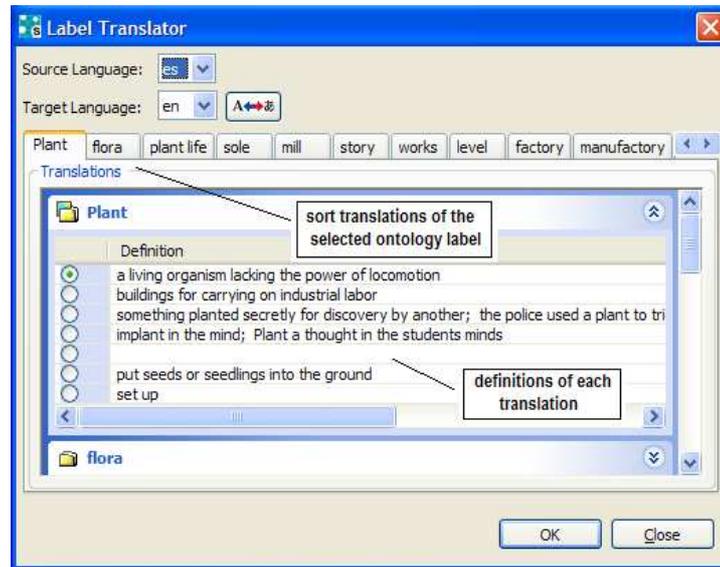
Fig. 3. Translations into English of the spanish ontology label "planta".

References

1. R. L. Cilibrasi and P. M. Vitnyi. The Google Similarity Distance. *IEEE Transactions on Knowledge and Data Engineering*, 19(3):370–383, march 2007.
2. 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.*
3. M. C. Suarez-Figueroa, (coordinator). NeOn Development Process and Ontology Life Cycle. NeOn Project Deliverable 5.3.1 (2007).
4. T. Pedersen, S. Banerjee, and S. Patwardhan. Maximizing Semantic Relatedness to Perform Word Sense Disambiguation. Research Report UMSI 2005/25, University of Minnesota Supercomputing Institute, March 2005.
5. W. Peters, E. Montiel-Ponsoda, and G. Aguado de Cea. Localizing ontologies in owl. In *OntoLex2007*, 2007.
6. R. Trillo, J. Gracia, M. Espinoza, and E. Mena. Discovering the semantics of user keywords. *Journal on Universal Computer Science. Special Issue: Ontologies and their Applications, ISSN 0948-695X*, 2007.