# Supporting Multilingual Ontology Matching With MoKi

Mauro Dragoni and Giulio Petrucci

FBK–IRST, Trento, Italy
`dragoni|petrucci@fbk.eu`

**Abstract.** Multilingual ontology matching is a recent research topic concerning the application of the traditional schema matching algorithms in conjunction with the use of multilingual resources. In this demo, we present an architecture, based on the use of Information Retrieval (IR) techniques, developed for addressing the task of suggesting alignments between two or more multilingual ontologies. The validation performed on domain-specific use cases demonstrated the suitability of the presented system.

## 1 Introduction

Ontology matching [5][1] is a task that has attracted considerable attention in recent years. However, with very few exceptions, research in this field has primarily focused on the development of monolingual matching algorithms. As more and more artifacts, especially in the Linked Open Data realm, become available in a multilingual fashion, novel matching algorithms are required. With the growth of interest in creating multilingual semantic artifacts for representing knowledge, the possibility of defining effective links between different artifacts allows to spread knowledge not only for enriching the representation of the knowledge itself, but also for breaking language barriers in accessing contents.

In the case of a multilingual environment, there are some peculiarities that can be exploited in order to relax the classic alignment task:

– the use of multilinguality allows to reduce the problems raised when two different concepts have the same label;
– multilingual resources provide term translations that have already been adapted to the represented domains ([6]);

In this demo, we present a work exploiting the two aspects described above in order to build a multilingual term-based approach for defining mappings between multilingual ontologies. Such an approach has been evaluated on domain-specific (agriculture and medical domains) use cases in order to measure both the efficiency of the tool and the effectiveness of the algorithm in a real-world environment.

## 2 The IR-based Approach For Multilingual Ontology Matching

Below, we described the procedure followed for implementing the system and the strategy that we used for defining new matches given two ontologies.

*Representation of Information* The approach is based on the exploitation of textual information (labels) associated with each concept described in an ontology. Where, with the term "label", we mean a string identifying the concept associated with its language tag (i.e. "concept_label@lang_code")

This way, we are able to exploit, for each entity modeled in the ontology, all labels in all languages by making string-based solutions both effective and semantically sensible. Indeed, when a particular label (independently by the language) is chosen by experts during the creation of an ontology, they implicitly inject in this choice their knowledge about the equivalence of meanings between different translations of each label.

*Index Construction* For each entity defined in the ontology, we have a set of pairs "label-language", and, in case of synonymy, we may have more pairs for the same language. Such labels are tokenized, stemmed , and normalized by using natural language process libraries. After this process, each pair is stored in an inverted index [7].

For instance, by considering a concept "Activity" having three labels in three different languages, its representation will be like it is shown below:

```
[prefLabel] "activity"@en   -->   label-en:activity
[prefLabel] "attivit"@it    -->   label-it:attivit
[prefLabel] "actividad"@es  -->   label-es:actividad
```

*Matches definition* Once the index representing and ontology (called "target") is created, the definition of new matches is done by performing queries using multilingual information extracted from the entities defined in the other ontology (called "source"). Such a query is created by building a structured representation of information defined in the concept contained in the source ontology that needs to be mapped. Therefore, similarly to the creation of the indexed records, by considering the example shown above, the corresponding query is built as follows:

```
[prefLabel] "executable activity"@en; "attivit eseguibile"@it; "actividad ejecutable"@es
```

the created query will be:

```
label-en:"executable activity" OR label-it:"attivit eseguibile" OR label-es:"actividad ejecutable")
```

When the system receives the request for trying to create a new match, it performs a search operation through the records built during the indexing phase. As a result, a rank ordered by the retrieval status value (RSV) is produced and returned by the system.

Further details about the algorithms used in this approach may be found in [3,2]

## 3  MoKi Implemented Facilities

The back-end component described in the previous Section can be accessed through the user facilities that have been integrated as extensions of the MoKi [4] tool and it can be used through the set of user facilities briefly described below.

For managing the mappings, a dedicated section in the concept modeling page as been integrated as shown in Figure 1. Here, the expert is able to see which are the concepts that have been already mapped with the current one and to decide if to maintain such mappings or to remove them. For creating a new mapping, the expert has to choose which ontology to use for requesting mapping suggestions, and then to click on the "Add New Mapping" button for invoking the suggestion service.

Fig. 1: User facility for invoking the suggestion service.

When the request is sent, on the background the structured representation of the current concept is converted into a query (as described in Section 2) which is performed on the index containing the concepts of the ontology specified by the expert. When the rank of the suggestions is composed, it is proposed to the expert as shown in Figure 2. For each suggestion, ordered by confidence score, the expert is able to open the concept description page (if available) by clicking on the concept URI, and to eventually define a new mapping by clicking on the "Create Mapping" button.



Fig. 2: Example of rank produced by the IR system containing the five suggestions for mapping the concept "Learning" defined in the Organic.Lingua ontology with a concept coming from the Agrovoc one.

Besides this, in a separated module, experts are able to upload new ontologies and creating the related indexes. Figure 3 shows the interface used by the experts for uploading a new ontology in the repository. This facility allows to convert the uploaded ontology in the structured representation described in Section 2 and to store it in a dedicated index. From this interface, experts are able to manage the ontologies already stored in the repository by viewing some basic information about them and, eventually, to delete one or more ontologies.

For adding a new ontology to the repository, experts have to select the file containing the ontology, write a description, decide an acronym for referring the ontology in the other sections of the tool, and, finally, press the "Save" button.

## 4   Concluding Remarks

The presented system has been tested on a set of domain-specific multilingual ontologies (Eurovoc, Agrovoc, and Gemet for the agriculture domain, and MDR, MeSH, and SNOMED for the medical domain). Table 1 presents the obtained results. The reported precision has been measured after the analysis of the first $X$ elements contained in the

## Ontology Manager

Number of loaded ontologies: 2

| ID | Ontology Namespace | Ontology Acronym | Description | | |
|----|--------------------|------------------|-------------|--------|--------|
| 1 | http://aims.fao.org/aos/agrovoc/ | Agrovoc | Agrovoc ontology. | Explore | Delete |
| 2 | http://eurovoc.europa.eu/ | Eurovoc | Eurovoc ontology. | Explore | Delete |

Add a new ontology to the repository

| Namespace | Acronym | Description | File | |
|-----------|---------|-------------|------|--|
| | | | Browse... No file selected. | Save |

Fig. 3: Interface used for loading a new ontology within the system.

rank, respectively 1, 3, and 5, in order to show how the effectiveness of the system changes by varying the number of suggestions considered. The obtained results demonstrated the suitability of the proposed system for supporting the multilingual ontology matching task.

| Mapping Set | # of Mappings | Prec@1 | Prec@3 | Prec@5 | Recall |
|-------------|---------------|--------|--------|--------|--------|
| Eurovoc → Agrovoc | 1297 | 0.816 | 0.931 | 0.967 | 0.874 |
| Agrovoc → Eurovoc | 1297 | 0.906 | 0.969 | 0.988 | 0.695 |
| Gemet → Agrovoc | 1181 | 0.909 | 0.964 | 0.983 | 0.546 |
| Agrovoc → Gemet | 1181 | 0.943 | 0.981 | 0.994 | 0.740 |
| MDR → MeSH | 6061 | 0.776 | 0.914 | 0.956 | 0.807 |
| MeSH → MDR | 6061 | 0.716 | 0.888 | 0.939 | 0.789 |
| MDR → SNOMED | 19971 | 0.621 | 0.826 | 0.908 | 0.559 |
| SNOMED → MDR | 19971 | 0.556 | 0.760 | 0.855 | 0.519 |
| MeSH → SNOMED | 26634 | 0.690 | 0.871 | 0.931 | 0.660 |
| SNOMED → MeSH | 26634 | 0.657 | 0.835 | 0.908 | 0.564 |

Table 1: Results obtained on a set of multilingual ontologies.

## References

1. Z. Bellahsene, A. Bonifati, and E. Rahm, editors. *Schema Matching and Mapping*. Springer, 2011.
2. M. Dragoni. Multilingual ontology mapping in practice: a support system for domain experts. In *The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, Pennsylvania, USA, October 11-15, 2015. To appear.*
3. M. Dragoni. Exploiting multilinguality for creating mappings between thesauri. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, SAC 2015, pages 382–387. ACM, 2015.
4. M. Dragoni, A. Bosca, M. Casu, and A. Rexha. Modeling, managing, exposing, and linking ontologies with a wiki-based tool. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland, May 26-31, 2014.*, pages 1668–1675. European Language Resources Association (ELRA), 2014.
5. J. Euzenat and P. Shvaiko. *Ontology matching*. Springer, 2007.
6. F. Lin and A. Krizhanovsky. Multilingual ontology matching based on wiktionary data accessible via sparql endpoint. *CoRR*, abs/1109.0732, 2011.
7. C. J. van Rijsbergen. *Information Retrieval*. Butterworth, 1979.