Creating a Multilingual Terminological Resource using Linked Data: the case of Archaeological Domain in the Italian language

Giulia Speranza[1], Carola Carlino[1], Sina Ahmadi[2]

[1]UNIOR NLP Research Group, University of Naples "L'Orientale", Naples, Italy [gsperanza,ccarlino]@unior.it
[2]Insight Centre for Data Analytics, National University of Ireland, Galway, Irland [sina.ahmadi@insight-centre.org]

Introduction

The lack of multilingual terminological resources in specialized domains constitutes an obstacle to the access and reuse of information. In the technical domain of Cultural Heritage (CH) and, in particular, archaeology, such an obstacle still exists for Italian language. Regarding the field of CH, multilingualism is still a challenge due to the tendency of experts to store terminologies monolingually [1]. In our experiment we propose to enrich with translations (en-fr-de-es) and lexicographic information (i.e. number, gender) retrieved from the web (Wikidata and Wiktionary [2]) a monolingual specialized thesaurus of archaeology and presenting this new resource using Ontolex-Lemon [3].

Methodology

Given a list of terms in the source dataset, we first retrieve those concepts to which the term is associated on Wikidata. The retrieved terms are then enriched by linguistic information from Wiktionary. The enriched information are evaluated by annotators, and then converted into the Ontolex-Lemon model in the Resource Description Framework (RDF).

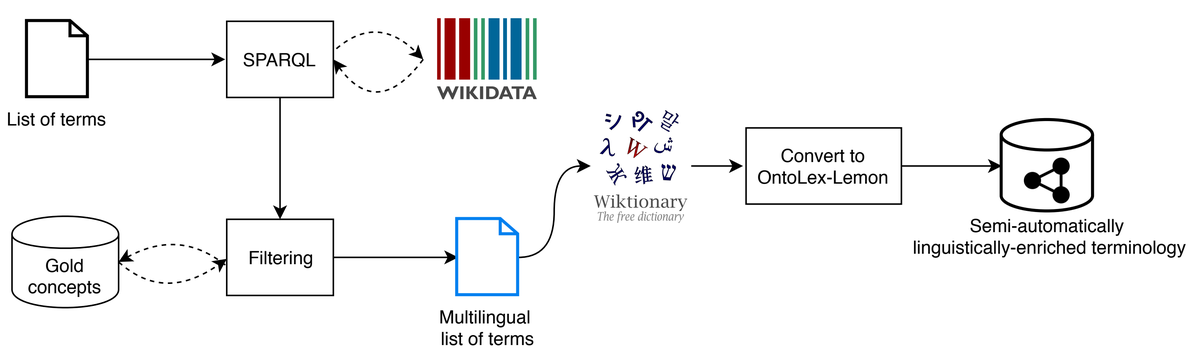

Figure 1:Terminological Enrichment Process

Conclusion

Results show that leveraging LOD resources is a valid option for enriching ontologies linguistically. Nonetheless, since CCRs are created by a community effort, a manual verification is always needed. Finally, the effort of this study can be framed within the more general context of contributing to the implementation and advancement of the multilingual Web of Data and the LLOD movement.

Case Study

Our source dataset was the "RA Thesaurus per la descrizione dei reperti archeologici" an open monolingual Italian vocabulary created by the ICCD (Istituto Centrale per il Catalogo e la Documentazione) in collaboration with the Italian Ministry of Cultural Heritage and Activities (MiBAC) in order to regulate the terminology to be used to identify archaeological finds in Italy.

Objectives

- Retrieving lexicographic information
- Enriching specialized thesaurus
- Represent info with Ontolex-Lemon

Ontolex-Lemon

```
:lexicon a lime:Lexicon;
 lime:entry :ascia ;
 lime:language
 <http://www.lexvo.org/page/iso639-1/it>.

:ascia a ontolex:LexicalEntry,
    ontolex:Word ;
   ontolex:canonicalForm :form_ascia ;
   rdfs:label "ascia"@it ;
   lexinfo:partOfSpeech lexinfo:noun ;
   lexinfo:gender lexinfo:feminine .

:form_ascia a ontolex:Form ;
   dct:language
       <www.lexvo.org/page/iso639-1/it>;
   ontolex:writtenRep "ascia"@it ;
   lexinfo:number lexinfo:singular ;
   ontolex:sense :ascia_n_sense ;
   ontolex:denotes wd:Q2517447;
  <https://it.wikipedia.org/wiki/Ascia>;
   dct:subject wd:Q382995 ;
   owl:sameAs dati:009000000004 .

:trans a vartrans:Translation ;
   vartrans:source :ascia_n_sense ;
   vartrans:target
       frl:fr_herminette_sense .
```

Figure 2:Ontolex-Lemon term representation

References

[1] Konstantinos N. Vavliakis, Georgios T. Karagiannis, and Pericles A. Mitkas.
Semantic Web in cultural heritage after 2020.
In *Proceedings of the 11th International Semantic Web Conference (ISWC), Boston, MA, USA*, pages 11–15, 2012.

[2] Christian M. Meyer and Iryna Gurevych.
Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography.
In Sylviane Granger and Magali Paquot (eds.), Electronic Lexicography. Oxford: Oxford University Press, 259-29, 2012.

[3] John P. McCrae, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano.
The Ontolex-Lemon model: development and applications.
In *Proceedings of eLex 2017 conference*, pages 19–21, 2017.