Hypernym relation extraction for establishing subsumptions: preliminary results on matching foundational ontologies

Mouna Kamel[1]*, Daniela Schmidt[2], Cassia Trojahn[1], and Renata Vieira[2]

[1] Institut de Recherche en Informatique de Toulouse, Toulouse, France
{mouna.kamel,cassia.trojahn}@irit.fr
[2] Pontificia Universidade Catolica do Rio Grande do Sul, Porto Alegre
daniela.schmidt@acad.pucrs.br, renata.vieira@pucrs.br

Abstract. This paper presents an approach for matching foundational ontologies involving subsumption relations. The approach relies on extracting hypernym relations from ontology annotations for establishing such kind of correspondences. We report preliminary results on exploiting lexico-syntactic patterns and definitions layout. Experiments were run on DOLCE and SUMO and the generated alignment was evaluated on a manually generated subsumption reference.

1 Introduction

Foundational ontologies describe general concepts (e.g., physical object) and relations (e.g., parthood), which are independent of a particular domain. The clarity in semantics and the rich formalization of these ontologies are fundamental requirements for ontology development [5] improving ontology quality. They may also act as semantic bridges supporting interoperability between ontologies [8, 10]. However, the development of different foundational ontologies re-introduces the interoperability problem, as stated in [6]. This paper addresses the problem of matching foundational ontologies.

Early works addressed this problem on different perspectives e.g., discussing their different points of view [14, 16, 9] or providing concept alignments between them [13, 7]. Few works have addressed the automatic matching of this kind of ontologies, such as in [7] where alignments between BFO, DOLCE and GFO were built both with automatic tools and manually, with substantially fewer alignments found by the tools. In fact, current tools fail on correctly capturing the semantics behind the ontological foundational concepts, what requires deeper contextualization of the concepts. Besides that, the task requires the identification of other relations than equivalences, such as subsumption and meronym. Few systems are able to discover other relations than equivalence (e.g., AML and BLOOM), with few propositions in the literature [19, 20]. We argue here that the knowledge encoded in the ontologies has to be further exploited. In that way, we propose to borrow approaches from relation extraction from text in NLP in order to establish subsumption relations between the ontologies to be matched.

While the approach is not completely new, as NLP techniques are often used to extract knowledge from text, their exploitation in ontology matching brings some novelty.

Relation extraction in ontology matching has been considered in few works. In [15], a supervised method learns patterns of subsumption evidences, while in [1] the approach relies on free-text parts of Wikipedia in order to help detecting different types of relations, even without clear evidence in the input ontologies themselves. Hearst patterns has been adopted in [17] and [18], with the former using them to eliminate noise in matching results. Here, we report preliminary results on exploiting lexico-syntactic patterns from Hearst [4] and evidences of hypernym relation carried out in definitions layout. Experiments were run on DOLCE and SUMO and the generated alignment has been evaluated on a manually generated subsumption reference. The novelty here is to exploit such methods for foundational ontology matching involving subsumption.

2 Proposed approach

Our approach relies on two main steps: (i) hypernym extraction from ontology annotations and (ii) subsumption generation between ontology concepts, as detailed below.

Hypernym extraction The hypernym relation extraction takes as input the ontology annotations as concept definitions (what are common in top-level ontologies). A *definition* attaches a meaning to a term denoting the concept. The term that is to be defined is called the *definiendum*, and the term or action that defines it is called the *definiens*. In the example below, the *definiendum* = “Product” and the *definiens*=“An Artifact that is produced by Manufacture and that is intended to be sold”. Many linguistic studies show that definitions mostly express one of the main lexical relations e.g., hypernymy, meronymy or synonymy, between *definiens* and *definiendum* [11].

```
<owl:Class rdf:ID= "Product">
  <rdfs:comment> An Artifact that is produced by Manufacture and
  that is intended to be sold.</rdfs:comment>
</owl:Class>
```

Different strategies are exploited for extracting the hypernym relations:

Hypernym relations expressed using definitions layout We focus on cases where the *definiens* starts by expressing an entity (denoted by a term and different from the *definiendum*) which have some properties. In the above example, the entity in the *definiens* is “Artifact” and the property is “that is produced by Manufacture and that is intended to be sold”. Thus the *definiendum* (Product) is an *hyponym* of the definiens (Artifact). When no property is expressed, it is usually a synonym relation, as below:

```
<owl:Class rdf:about="#Quale">
  <rdfs:comment> An atomic region. </rdfs:comment>
</owl:Class>
```

Hypernym relations lexically expressed in text annotations OWL class definitions may also be more fine grained exploited, as comment paragraphs may contain well-written text. We then exploit this text using a set of lexico-syntactic patterns from Hearst [4]:

[`NP such as` $\{$`NP ,`$\}*$ $\{$`or`$|$`and`$\}$ `NP`], [`NP like` $\{$`NP ,`$\}*$ $\{$`or`$|$`and`$\}$ `NP`], [`NP which is an example of NP`], [`NP including` $\{$`NP ,`$\}*$ $\{$`or`$|$`and`$\}$ `NP`], [`NP is called NP if`], [`NP is an NP that`].

For instance, the pattern [`NP like` $\{$`NP ,`$\}*$ $\{$`or`$|$`and`$\}$ `NP`] means that a noun phrase (`NP`) must be followed by the word "like", which must be followed by an NP or by a list of NPs separated by comma, having before the last NP "or" or "and". When applied on the definition below, the hypernym relations (Self Connected Object, planet), (Self Connected Object, star) and (Self Connected Object, asteroid) can be identified.

```
<owl:Class rdf:about="#AstronomicalBody">
  <rdfs:comment> The Class of all astronomical objects of
     significant size. It includes Self Connected Objects
     like planets, stars, and asteroids ...
 </rdfs:comment>
</owl:Class>
```

Hypernym relations carried out by the concept identifier Hypernym relations may also be identified from modifiers of a head of a compound noun denoting the identifier of the OWL class. In the example above, the hypernym relation (astronomical body, body) can be identified thanks to this strategy.

Subsumption generation Having extracted all the hypernym relations from both ontologies to be matched, we verify if the terms appearing as hyponyms and hypernyms denote concepts in the ontologies. In the example above, as the alignment is directional, "Product" denotes a concept in the source ontology and "Artifact" in the target ontology, hence this hypernym pair is kept.

3 Experiments

Material and methods We used the foundational ontologies DOLCE [3][1], an ontology of *particulars* which aims at capturing the ontological categories underlying human commonsense; and SUMO [12][2], an ontology of particulars and universals. The reference alignment involving 41 subsumption correspondences comes from [13]. The approach has been implemented with GATE: to extract concepts and their associated comments from the ontology OWL file and restructuring them according to an XML format; to identify terms using first the TermoStat term extractor, and then expanding the recognition of terms using JAPE rules (for instance, the sequence made of a TermoStat term preceded or followed by adjectives, constitutes a new term); to annotate the XML corpus with different NLP tools (ANNIE Tokenizer, Stanford POS, Stanford parser, Gazeteer of identified terms); and to identify hypernym relations.

[1] `http://www.loa.istc.cnr.it/old/DOLCE.html`
[2] `https://github.com/ontologyportal/sumo`

Results and discussion Table 1 shows the results of each strategy and their combination. As somehow expected, patterns are very precise while head modifier provides good results in terms of recall with respect to the other strategies. Comparing the approach to the OAEI 2018 matchers[3] (Table 2), besides the fact that we do not distinguish subsumption and equivalence relations when computing precision and recall, no matcher were able to find the correspondences. From the 41 reference correspondences, only one correspondence refers to similar terms (`dolce:geographical-object` and `sumo:GeographicArea`) and 5 of them could be found via a head modifier method (e.g., `dolce:organization` and `sumo:PoliticalOrganization`). In order to see how close the generated alignments were to the reference, we have calculated the relaxed precision and recall [2], that measure the closeness of the results to the reference. While the results of our approach are not that close to the reference, in terms of recall we obtain results similar than the relaxed recall for all matchers.

Combination			Layout			Patterns			Head modifier			Layout+patterns		
P	F	R	P	F	R	P	F	R	P	F	R	P	F	R
.27	.23	.20	.18	.13	.10	1,00	.05	.03	.32	.20	.15	.22	.16	.13

Table 1. Results of the different relation extraction strategies.

System	**Classical**			**Relaxed**		
	P	F	R	P	F	R
M1	.00	.00	.00	.00	.00	.00
M2	.00	.00	.00	.33	.18	.15
M3	.00	.00	.00	.39	.27	.21
M4	.00	.00	.00	.77	.34	.21
M5	.00	.00	.00	.32	.25	.17
M6	.00	.00	.00	.28	.14	.12
M7	.00	.00	.00	.57	.31	.21
M8	.00	.00	.00	.50	**.42**	.21
Proposed approach	.27	**.23**	.20	.28	.28	.29

Table 2. Classical and relaxed precision (P), recall (R) and F-measure (F) of the proposed approach and matchers.

4 Conclusions

We have reported here preliminary results on exploiting symbolic hypernym relation extraction approaches for generating subsumption correspondences between foundational ontologies. This task is still a gap in the field and the initial results presented here can be improved in different ways. First of all, we plan to improving the relation extraction by (i) extending the list of lexico-syntactic patterns, (ii) exploiting syntactic analysis of the text and treating anaphores, and (iii) using background resources such as DBpedia, BabelNet (in particular top level layers of these resources). We also plan to combine relation extraction strategies with matching strategies (structural) and word embeddings, as well as to work on other lexical relations like meronymy. Finally, we plan to apply the approach on domain ontologies.

[3] The aim here is not to evaluate the matching systems themselves, for that reason, their names have been anonymized.

Acknowledgments We warmly thank D. Oberle for sending us all the generated alignments between SUMO and DOLCE-Lite.

References

1. E. Beisswanger. Exploiting relation extraction for ontology alignment. In *Proceedings of the International Semantic Web Conference*, pages 289–296, 2010.
2. M. Ehrig and J. Euzenat. Relaxed precision and recall for ontology matching. In *Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies*, 2005.
3. A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening Ontologies with DOLCE. In *Proceedings of the 13th Conference on Knowledge Engineering and Knowledge Management*, pages 166–181, 2002.
4. M. A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th Conference on Computational Linguistics*, pages 539–545, 1992.
5. C. Keet. The use of foundational ontologies in ontology development: An empirical assessment. In *Proceedings of the Extended Semantic Web Conference*, pages 321–335, 2011.
6. Z. Khan and C. Keet. Addressing issues in foundational ontology mediation. In *Proceedings of the Conference on Knowledge Engineering and Ontology Development*, pages 5–16, 2013.
7. Z. Khan and C. Keet. The Foundational Ontology Library ROMULUS. In *Proceedings of the 3rd International Conference on Model and Data Engineering*, pages 200–211, 2013.
8. V. Mascardi, A. Locoro, and P. Rosso. Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation. *Knowledge and Data Engineering*, 22(5):609–623, 2010.
9. L. Muñoz and M. Grüninger. Verifying and mapping the mereotopology of upper-level ontologies. In *Proceedings of the International Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pages 31–42, 2016.
10. J. C. Nardi, R. de Almeida Falbo, and J. P. A. Almeida. Foundational ontologies for semantic integration in EAI: A systematic literature review. In *Proceedings of the 12th IFIP WG Conference on e-Business, e-Services, and e-Society, I3E*, pages 238–249, 2013.
11. R. Navigli, P. Velardi, and J. M. Ruiz-Martínez. An annotated dataset for extracting definitions and hypernyms from the web. In *Proceedings of the International Conference on Language Resources and Evaluation*, 2010.
12. I. Niles and A. Pease. Towards a Standard Upper Ontology. In *Proceedings of the Conference on Formal Ontology in Information Systems*, pages 2–9, 2001.
13. D. Oberle, A. Ankolekar, P. Hitzler, P. Cimiano, M. Sintek, M. Kiesel, B. Mougouie, S. Baumann, S. Vembu, M. Romanelli, and Buitelaar. DOLCE Ergo SUMO: On Foundational and Domain Models in the SmartWeb Integrated Ontology. *Web Semantics*, 5(3):156–174, 2007.
14. A. Seyed. BFO/DOLCE Primitive Relation Comparison. In *Nature Proceedings*, 2009.
15. V. Spiliopoulos, G. A. Vouros, and V. Karkaletsis. On the discovery of subsumption relations for the alignment of ontologies. *Journal of Web Semantics*, 8(1):69 – 88, 2010.
16. L. Temal, A. Rosier, O. Dameron, and A. Burgun. Mapping BFO and DOLCE. In *Proceedings of the World Congress on Medical Informatics*, pages 1065–1069, 2010.
17. W. R. van Hage, S. Katrenko, and G. Schreiber. A method to combine linguistic ontology-mapping techniques. In *International Semantic Web Conference*, pages 732–744, 2005.
18. R. Vazquez and N. Swoboda. Combining the semantic web with the web as background knowledge for ontology mapping. In *Meaningful Internet Systems*, pages 814–831, 2007.
19. A. Vennesland. Matcher composition for identification of subsumption relations in ontology matching. In *Proceedings of the Conference on Web Intelligence*, pages 154–161, 2017.
20. N. Zong, S. Nam, J.-H. Eom, J. Ahn, H. Joe, and H.-G. Kim. Aligning ontologies with subsumption and equivalence relations in linked data. *Knowledge Based Systems*, 76(1):30–41, 2015.