Complex matching for multiple ontologies: an exploratory study

Madalena Pavão and Cátia Pesquita

LASIGE, Faculdade de Ciências da Universidade de Lisboa, Portugal
mpavao@lasige.di.fc.ul.pt, clpesquita@ciencias.ulisboa.pt

1   Introduction

Complex matching, i.e., finding correspondences that go beyond equivalence and are able to capture more complex relationships between entities or sets of entities, is a recognized challenge (for a more in-depth overview see [1]). However, in multi-domain areas, there is a need to perform 'holistic' matching - link multiple ontologies to address different perspectives of the underlying data, while maintaining the inherently distributed paradigm championed by the Semantic Web. This need motivates 'compound mappings' involving more than two ontologies. A specific case is the ternary compound mapping [2], for example, the HP class 'broad forehead' is equivalent to an axiom obtained by relating PATO ('increased width') and UBERON ('forehead') classes, via an intersection. In this study, we explore compound matching involving multiple concepts from multiple ontologies.

2   Methods and Results

A compound mapping is a tuple $<C_s,[C_{t0},...,C_{tn}],[P_{t0},...,P_{tn}],M>$, where C_s is a class from a source ontology, $[C_{t0},...,C_{tn}]$ and $[P_{t0},...,P_{tn}]$ are a set of target classes extracted from multiple target ontologies and the set of properties that stand between them, while M is a mapping relation established between the source class and the expression composed by the set of target classes and properties. For our purposes, we are restricting our approach to finding mappings where M is an equivalence, and we are simplifying our goal to just finding the set of target classes. Our compound matching algorithm is based on finding partial lexical matches between word sequences in the source class labels and full labels of target classes. The algorithms for compound alignment were developed within the AgreementMakerLight (AML) system [3] and use hash map-based data structures to improve scalability.

We evaluated our algorithm using four test cases: (1) HP as source, UBERON and PATO as targets; (2) HP as source, UBERON, PATO and GO as targets; (3) MP as source, UBERON and PATO as targets; (4) MP as source, UBERON, PATO and GO as targets[1]. The reference alignments for each task were generated

---

[1] HP:Human Phenotype Ontology; PATO: Phenotypic Quality Ontology; UBERON: Uber anatomy ontology; GO:Gene Ontology

by extracting all the Equivalent Classes Axioms of MP and HP OWL files with OWL API. For each ontology we created two references: (1) UB-PT: containing mappings that employ classes from the UBERON and/or PATO ontologies; (2) UB-PT-GO: containing all mappings that employ the UBERON and/or PATO and/or GO ontologies. Note that these are just partial alignments, since they only cover 39%of the classes in HP and 28.7% in MP.

In our evaluation (shown in Table 1) we have considered partial mappings as positive[2]. The MP tasks had a better performance than HP, and for MP there was a marked difference between the UB-PT and UB-PT-GO, with the former having an improvement of over 30% in f-measure.

Table 1. Performance metrics for partial mappings

	Precision	Recall	F-measure
MP-UB-PT	75.4%	98.8%	85.6%
MP-UB-PT-GO	46.4%	61.1%	52.8%
HP-UB-PT	26.0%	43.4%	32.5%
HP-UB-PT-GO	26.1%	44.9%	33.0%

3   Conclusions

This exploratory study highlights the difficulties in performing holistic matching, which range from scalability issues to the inherent increased complexity of the task. Recent evaluation of complex matching approaches revealed that all techniques produced f-measures below 20% [1]. Furthermore, building reference alignments for compound matching is a standing challenge, and the reference alignments we built can only be considered partial references. We envision several future work endeavors in this area, ranging from more complex lexical approaches to approaches richer in semantics that are able to capture property restrictions to reproduce the full equivalence axiom.

References

1. Thiéblin, É., Haemmerlé, O., Hernandez, N., Trojahn, C.: Task-oriented complex ontology alignment: Two alignment evaluation sets. In: European Semantic Web Conference, Springer (2018) 655–670
2. Oliveira, D., Pesquita, C.: Improving the interoperability of biomedical ontologies with compound alignments. Journal of biomedical semantics **9**(1) (2018) 1
3. Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I.F., Couto, F.M.: The agreementmakerlight ontology matching system. In: On the Move to Meaningful Internet Systems: OTM 2013 Conferences, Berlin, Heidelberg, Springer (2013) 527–541

---

[2] Mappings that have at least one class in common with the classes in the reference