Building reference alignments for compound matching of multiple ontologies using OBO cross-products

Catia Pesquita[1], Michelle Cheatham [2], Daniel Faria[1], Joana Barros [1], Emanuel Santos[1], and Francisco M. Couto[1]

[1]Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa, Portugal
[2] DaSe Laboratory, Wright State University, Dayton OH 45435, USA
cpesquita@di.fc.ul.pt

Abstract. Existing ontology matching techniques are limited to matching two ontologies, but we argue that producing 'compound' alignments, involving more than two ontologies, would be useful to support a next generation of semantic technologies. To foster the development of new techniques in this area, we have investigated the suitability of exploring OBO cross-products to derive ternary compound alignments that can be used as a benchmark. We were able to establish seven such reference alignments with over 100 mappings each, between ten biomedical ontologies. Preliminary experiments revealed that the increase in matching space and the inherently more difficult-to-compute ternary mapping pose interesting difficulties to compound ontology matching.

Introduction. Both the 'classical' and 'complex' (e.g., [1–3]) ontology matching approaches focus on discovering mappings between two ontologies. We argue that it would be useful for the developers of ontology alignment systems to develop new techniques and tools for identifying 'compound matches', i.e. matches between class or property expressions involving more than two ontologies. The simplest of these mappings would correspond to an equivalence mapping between a class A of one ontology and an expression relating classes B and C of two other ontologies, constituting a ternary relationship. We investigate the suitability of exploring OBO cross-products to create ternary compound alignments between ontologies which can function as a gold-standard to support the evaluation of novel matching methods for compound alignment.

Approach. We consider that a ternary compound alignment is a set of correspondences (mappings) between classes from a source ontology O_s and class expressions obtained by combining two other classes each belonging to a different target ontology O_{t1} and O_{t2}. We define a ternary compound mapping as a tuple $< X, Y, Z, R, M >$, where X, Y and Z are classes from three distinct ontologies, R is a relation established between Y and Z to generate a class expression that is mapped to X via a mapping relation M. Some of the logical definitions contained in OBO cross-products correspond to this type of mapping, for instance,

the class HP:0000337 labeled *broad forehead* is equivalent to an axiom obtained by relating the classes PATO:0000600 (*increased width*) and FMA:63864 (*forehead*) via an intersection qualified by an *inheres_in* relation. We analyzed the resources available at obofoundry.org [1] and identified seven cross-products collections each with at least 100 definitions corresponding to ternary compound mappings:

Source Ontology	Target Ontologies		Size
MP	PATO	UBERON	1725
HP	PATO	FMA	1519
MP	PATO	CL	407
WBPhenotype	PATO	GO	369
MP	PATO	GO	354
FYPO	PATO	GO	285
MP	PATO	NBO	100

To create the alignments based on the cross-products collections we used EDOAL [4], since it allows the construction of entities from other entities using algebraic operators. To represent *intersection_of* we employed a class expression with the *and* operator.

Experiments. In ternary ontology matching, the search space is cubic, so matching even relatively small ontologies can pose efficiency problems. In a preliminary experiment, we adapted the anchor-based strategy of the AgreementMakerLight system[5] as well as its WordMatcher algorithm to use a modified Jaccard index that penalizes words shared by both target classes. We tested it in the MP-PATO-CL and MP-PATO-NBO alignments, obtaining recall values of 30 and 11% respectively, but precision values below 1%. These results highlight some of the complexity behind compound alignments, even between ontologies that strive to follow the same naming conventions. We posit that to solve these issues, background knowledge or instances would be needed to be able to discriminate between the candidate mappings.

Acknowledgements. This study was funded by the Portuguese FCT through the SOMER project (PTDC/EIA-EIA/119119/2010) and the LASIGE Strategic Project (PEst-OE/EEI/UI0408/2014) and by the National Science Foundation under award 1017225 "III:Small: TROn—Tractable Reasoning with Ontologies."

References

1. Ritze, D., Völker, J., Meilicke, C., Šváb-Zamazal, O.: Linguistic analysis for complex ontology matching. Ontology Matching **1** (2010)
2. Šváb-Zamazal, O., Svátek, V., Iannone, L.: Pattern-based ontology transformation service exploiting oppl and owl-api. In: Knowledge Engineering and Management by the Masses. Springer (2010) 105–119
3. Meilicke, C., Noessner, J., Stuckenschmidt, H.: Towards joint inference for complex ontology matching. AAAI (Late-Breaking Developments) (2013)
4. David, J., Euzenat, J., Scharffe, F., Trojahn dos Santos, C.: The alignment api 4.0. Semantic web **2**(1) (2011) 3–10
5. 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, Springer (2013) 527–541

[1] http://obofoundry.org/index.cgi?show=mappings