Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in: https://oatao.univ-toulouse.fr/22205

Official URL
http://ceur-ws.org/Vol-2032/om2017_poster6.pdf

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Towards a complex alignment evaluation dataset

Élodie Thiéblin, Ollivier Haemmerlé, Nathalie Hernandez, Cassia Trojahn

IRIT & Université de Toulouse 2 Jean Jaurès, Toulouse, France
{firstname.lastname}@irit.fr

1 Motivation and background

Simple ontology alignments, largely studied, link one entity from a source ontology to one entity of a target ontology. One of the limitations of these alignments is, however, their lack of expressiveness which can be overcome by complex alignments. Different approaches for generating complex alignments have emerged in the literature [4,5,6]. However, there is a lack of datasets on which they can be evaluated.

Ontology matching is the process of generating an alignment. An alignment A between a source $o1$ and a target $o2$ ontologies is a set of correspondences [2]. Each correspondence is a triple $\langle e_{o1}, e_{o2}, r \rangle$. e_{o1} and e_{o2} are the members of the correspondence: they can be single ontology entities or constructions of these entities using constructors or transformation functions. r is a relation (e.g., $\equiv$, $\leq$, $\geq$) between e_{o1} and e_{o2}. We consider two types of correspondences:

- **simple** correspondence when both e_{o1} and e_{o2} are single entities: e.g. $\forall x$, $o1{:}Person(x) \equiv o2{:}Human(x)$ is a simple correspondence.
- **complex** correspondence when at least one of e_{o1} or e_{o2} is a construction of entities, i.e. involving at least a constructor or a transformation function. For example, $\forall x,y,\ o1{:}priceInDollars(x,y) \equiv \exists y1,\ o2{:}priceInEuro(x,conversion(y))$ is a complex correspondence with a transformation function (*conversion* that states that $y1 = changeRate \times y$). $\forall x,\ o1{:}AcceptedPaper(x) \equiv \exists y$, $o2{:}Paper(x) \wedge o2{:}acceptedBy(x,y)$ is a complex correspondence with constructors.

A complex alignment contains at least one complex correspondence.

2 The evaluation dataset

The proposed dataset is based on the OntoFarm dataset [9] composed of 16 ontologies on the conference organisation domain and simple reference alignments between 7 of these ontologies. This dataset has been widely used in the ontology alignment evaluation domain [8]. The dataset proposed here is a first version of an extension of the OntoFarm dataset including complex correspondences. 3 out of the 7 ontologies of the reference alignments have been manually aligned (*cmt*, *conference* and *edas*), resulting in 3 alignments: *cmt-conference*, *cmt-edas* and *conference-edas*. The methodology applied to create the complex dataset consists

in manually finding an equivalent construction of target entities for each source entity. All correspondences have a single entity member and an other member that is either a single entity (simple correspondence) or a construction (complex correspondence). The correspondences are diverse for they can be classified with 8 different correspondence patterns or compositions of them [7]. In the 3 alignments, the dataset contains 51 complex correspondences. The alignments are expressed in First Order Logic and in EDOAL[1]. The resulting alignments were translated into OWL axioms as an ontology merging process. The HermiT reasoner [3] was used to check the consistency of the merged ontology. The dataset is available online at `http://doi.org/10.6084/m9.figshare.4986368.v4` under a CC-BY License.

3   Conclusion and future work

We have proposed a complex coherent dataset with complex correspondences between 3 ontologies of the OntoFarm dataset. As perspectives, the dataset will be extended with other ontologies of this dataset. The confidence of a correspondence (a value associated with a correspondence to express its confidence degree) could be added to the dataset. This could express, as in [1], the consensus level of experts on each correspondence. Finally, we aim at using this dataset for the purpose of evaluating complex matchers.

References

1. Cheatham, M., Hitzler, P.: Conference v2. 0: An uncertain version of the OAEI Conference benchmark. In: ISWC. pp. 33–48. Springer (2014)
2. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer Berlin Heidelberg (2013)
3. Glimm, B., Horrocks, I., Motik, B., Stoilos, G., Wang, Z.: HermiT: An OWL 2 reasoner. Journal of Automated Reasoning 53(3), 245–269 (2014)
4. Jiang, S., Lowd, D., Kafle, S., Dou, D.: Ontology matching with knowledge rules. In: Transactions on Large-Scale Data-and Knowledge-Centered Systems XXVIII, pp. 75–95. Springer (2016)
5. Parundekar, R., Knoblock, C.A., Ambite, J.L.: Linking and building ontologies of linked data. In: ISWC. pp. 598–614. Springer (2010)
6. Ritze, D., Meilicke, C., Šváb Zamazal, O., Stuckenschmidt, H.: A pattern-based ontology matching approach for detecting complex correspondences. In: 4th ISWC workshop on ontology matching. pp. 25–36 (2009)
7. Scharffe, F.: Correspondence Patterns Representation. Ph.D. thesis, Faculty of Mathematics, Computer Science and University of Innsbruck (2009)
8. Zamazal, O., Svátek, V.: The Ten-Year OntoFarm and its Fertilization within the Onto-Sphere. Web Semantics: Science, Services and Agents on the World Wide Web 43, 46–53 (Mar 2017)
9. Šváb, O., Svátek, V., Berka, P., Rak, D., Tomášek, P.: Ontofarm: Towards an experimental collection of parallel ontologies. Poster Track of ISWC 2005 (2005)

---

[1]`http://alignapi.gforge.inria.fr/edoal.html`