ONTMAT1: Results for OAEI 2019*

Saida Gherbi[1] and M[ed]Tarek Khadir[2]

[1]LabGed, ESTI, Annaba 23000, Algeria
Saida_gharbi23@yahoo.fr
[2]LabGed, University Badji Mokhtar Annaba, 23000, Algeria
Khadir@labged.net

Abstract: This paper describes an ontology matching system named ONTMAT1, and presents the results obtained for the Ontology Alignment Evaluation Initiative (OAEI) 2019. ONTMAT1 compares entities of ontologies to align by structural and terminological methods which use a reasoner along with wordnet dictionnary. Thus, based on similarities of individual, datatype properties and the semantic of property restriction, the weight that estimates the performance of structural and linguistic similarities is calculated.

Keywords: Ontology, Alignment, OWL.

1 Presentation of the system

ONTMAT1 (ONTology MATching) is an ontology alignment tool, aiming to align OWL entities (classes, object properties), participating for the first time in OAEI (Conference track). The specificities of ONTOMAT1 are presented below:

1.1 State, purpose, general statement

ONTMAT1 uses terminological methods based on n-gram measure and WordNet dictionary [1] that is exploited as background knowledge along with pellet reasoner [2], to provide synonyms of names of individuals, concepts, and properties, of ontologies source ($O1$) and target($O2$). The results obtained are saved in: individual matrix (M_{ind}), concepts matrix (M_{cpt}), and properties matrix (M_p), for individuals, concepts and properties, respectively.

Furthermore, a new weight that evaluates the impact of restriction property (object properties [3] and data type properties) on the structural similarity of concept is calculated. Thus, the impact of terminological similarity is 1 minus this weight. Then, the final result of concepts alignment is the sum of these similarities.

1.2 Approach description

The suggested algorithm is composed of 3 levels as explain in the following:

1. In level 1, normalization techniques such as lemmatization [4], are applied on each entities name of matrices (M_{cpt}, M_{ind}, M_p). Then, the n-gram measure is used to assess the similarity among these entities. This measure is opted because it permits the control of the lexicon size and keeping at the same time a reasonable threshold for every composed term (names).The obtained value is assigned to the intersection between entities into every matrix. Since, the metric measures used to align entities may suffer of several drawbacks, such as: the existence of synonyms that expresses the same entity using different words. Entities names are also compared to WordNet synsets using n-gram and the relation among synsets are inferred by Pellet reasoner. Then, the relations among these entities are deduced from relations inferred by the reasoner.
   If *synonym* relation is inferred, then the value of intersection among these entities in their matrix becomes the average between 1.0 and the value calculated by the n-gram measure, else the existent value is preserved.
2. In level 2, every property restriction defines the class allocated by a weight w_i that evaluates the influence of its semantic on this class.
   The sublanguage OWL-DL of OWL (Web Ontology Language) certified by the World Wide Web Consortium (W3C)[i] is adopted in this paper to define the offered ontology matching algorithm. This language distinguishes two types of property restrictions: value constraints and cardinality constraints, which give a semantic sense to the assessed weight. A value constraint applies constraints on the range of the property. These constraints put on the class C or an object *o* can be:
   - *allValuesFrom(C)*, is the same to the universal (for-all: ∀) quantifier of Predicate logic that for each instance of C, every value for Property must satisfies the constraint. Therefore, the algorithm can assert that this property has a robust impact on the class. Consequently, from its semantic, the influence of this restriction on the class is considered "strong" and suggested 1.0 as weights w_{1i}, w_{2j} in O_1, O_2, respectively, affected by ONTMAT1 to *allValuesFrom*.
   - *someValuesFrom(C)*, is similar to the existential quantifier of Predicate logic that for each instance of C, there exists at least one value for Property that satisfies the constraint. Therefore, the influence of this constraint on a given class can be valued as average and the value 0.75 is affected to w_{1i} in O_1 and w_{2j} in O_2.
   - *hasValue(o)*, joins a restriction class to a value *o*, which may be an individual or a data value. This restriction designates a class of all individuals for which the concerned property has at least one value *semantically* equivalent to *o* (it can, also, have supplementary values). The effect of this restriction can be considered as weak and the

assigned weights (w_{1i}, w_{2j} in O_1, O_2, respectively) are evaluated to 0.25.

- A cardinality constraint is defined by *maxCardinality(n)* and *minCardinality(n)*, where (*n*) is the number of values that a property can take. *Owl:maxCardinality(n)* describes a class of all individuals that have *at most n* diverse values (individuals or data values) for the concerned property. The influence of this constraint is only on *n* value, for this reason, it is estimated as a weak constraint and ONTMAT1 affects 0.25 to weights w_{1i}, w_{2j} in O_1, O_2, respectively. The same for *minCardinality(n)* that describes a class of all individuals that have *at least n* various values for the concerned property.

3. Level 3 assesses structural similarity between concepts established upon properties restrictions. Property restrictions can be either *datatype properties* (data literal is the value of properties), or *object properties* (individual is the value of properties)[ii]. Firstly, restriction names of concepts (C_{1i}, C_{2j}) to be matched are compared using terminological methods.
   Secondly, same terminological methods are used to measure similarities among *datatype properties* names of both concepts to align, as well as the average of these similarities is calculated to determine *data similarities*.
   Finally, similarities among individuals of concepts to match are extracted from M_{ind} to compute their average *data similarities*.
   Afterwards, weights w_i and w_j evaluated influences of property on concepts are multiplied by *data similarities* and *data similarities*. Furthermore, values affected to M_p will be replaced by those deduced in this level.
4. The last level consists on aggregation of above similarities of concepts. Consequently, the final similarity is the sum of structural similarity and 1 minus the average of structural weights multiplied by terminological similarity.

1.3 Adaptations made for the evaluation

The alignment format adapted by the results, is the "=" sign for equivalence relation with confidence of 1.

However our system provides other relation called fuzzy relation symbolized by *-1*, proposed to resolve the problem of domination of structural similarity. This relation designates that the suggested system cannot decide about the relation that can be among the entities to match. This relation is assigned to concepts in which the difference between its $TermSim(C_{1i}, C_{2j})$, $StructSim(C_{1i}, C_{2j})$, has a value that exceeds a certain threshold considered according to the expertise of the application in OAEI.

2 Results

In this version we wish to test the techniques used by ONTMAT1, for instance: the inferences mechanisms applied upon WordNet, and the deduction of the matching among entities using weight based on restriction properties. The track used to perform these tests is the conference track. Conference track comprises 16 ontologies from the domain of conference organization.

The results of the evaluation based on crisp reference alignments that contains only classes (*M1-rar2; M1-ra1; M1-ra2*) are considered in this study because the objective of this version is to show the influence of the weight and the reasoner on the classes alignment and properties will be treated in the next version

As depicted in Table 1, ONTMAT1 provides fairly stable alignments when matching conference ontologies. Table 2 illustrates that ONTMAT1's performance in discrete and continuous cases increases 16 percent in terms of F-measure over the sharp reference alignment from 0.55 to 0.64, driven, principally, by increased recall.

Table 1. Results based on the crisp reference alignments.

	Precision	F-Measure 1	Recall
M1-ra1	0.82	*0.61*	0.49
M1-ra2	0.77	0.56	0.44
M1-rar2	0.77	0.58	0.46

Table 2. Results based on the uncertain version of the reference alignment.

Precision	F-measure1	Recall
Uncertain reference alignments (Sharp)		
0.82	0.55	0.41
Uncertain reference alignments (Discrete)		
0.82	0.64	0.52
Uncertain reference alignments (Continuous)		
0.82	0.64	0.53

Finally, ONTMAT1 have generated only one incoherent alignment in the evaluation based on logical reasoning.

2.1 Discussions on the way to improve the proposed system

To improve the proposed application, properties of ontologies ($O1, O2$) will also be aligned. Then, adapt it to read all files type, and integrate the translator to test our tool under other tracks as: Instance Matching, MultiFarm.

3 Conclusion and future work

We have briefly described the mechanisms exploited by our proposition ONTMAT1, and presented the results obtained under the conference track of OAEI 2019.

This is our firs participation in OAEI with ONTMAT1, the results are satisfying, and the system presents some limitations in term of recall. In the future, a greater effort will be made to improve ONTMAT1 results, and participate in more tracks.

References

1. Fellbaum, C. :WordNet: An Electronic Lexical Database, MIT Press, Cambridge, MA (1998).
2. Sirin E. et Parsia B. 2004 . “Pellet : An owl dl reasoned”, DansHaarslev, V. et Möller, R. (éditeurs), Proceedings of the International Workshop on Description Logics (DL2004).
3. Gherbi S., M. T. Khadir, Inferred ontology concepts alignment using and an external dictionary, Procedia Computer Science 83 (2016) 648- 652, the 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Aliated Workshops.doi:http://dx.doi.org/10.1016/j.procs.2016.04.145.URL http://www.sciencedirect.com/science/article/pii/ S1877050916301752.
4. Euzenat, J., Shvaiko. P, (2007). “Ontology Matching” , Springer-Verlag Berlin Heidelberg

[i] https://www.w3.org/

[ii] https://www.w3+.org/TR/owl-ref/