Variations on Aligning Linked Open Data Ontologies

Valerie Cross, Chen Gu, Xi Chen, Weiguo Xia, Peter Simon

Computer Science and Software Engineering Department
Miami University, Oxford, OH 45056
crossv@muohio.edu

Traditional OA systems are not as suitable for aligning LOD ontology schemas; for example, equivalence relations are limited among LOD concepts so that OA systems for LOD ontology alignment also find subclass and superclass relations. Four recent approaches for LOD ontology alignment are BLOOMS (BL) [1] and BLOOMS+ [2], AgreementMaker (AM) [3], WikiMatch (WM) [4], and Holistic Concept Mapping (HCM) [5]. Table 1 briefly compares these systems for aligning LOD ontologies.

Table 1. Recent OA systems for aligning LOD ontology schemas

OA system	Mapping type	Knowledge Source	Data Structure	Algorithms	Experiment description
BL/BL+	Equivalence, subclass	Wikipedia category hierarchy	Concept category trees in a forest	Tree overlap using node depth, contextual similarity on superconcepts	LOD reference alignments, Proton mappings to DBpedia, Geonames, Freebase.
AM	Equivalence, subclass, superclass	WordNet, other LOD ontologies, i.e.,DBpedia or FOAF	Lexicon for a concept	Advanced Similarity Matcher, inferencing on import concept	LOD reference alignments
WM	Equivalence	Wikipedia articles	sets of articles	Jaccard index on article sets	OAEI 2011.5 conference track, multifarm dataset
HCM	Equivalence, similar to, disjoint	Wikipedia category hierarchy	Concept category trees in a forest.	IR tf-idf on comment, label keyword, topic sets ppjoin with Jaccard	Concepts from triples of Billion Triple Challenge dataset, expert evaluation

Unlike the Ontology Alignment Evaluation Initiative (OAEI), no standard reference alignment exists for LOD ontologies. Researchers [4] had experts develop a benchmark, the LOD reference alignments between ontology schema pairs taken from eight LOD ontologies: AKT Reference (A), BBC program(B), DBpedia (D), FOAF(F), Geonames(G), Music(M), SIOC (S), and the Semantic Web Conference (W) because of their substantial LOD coverage domain diversity, and publicly available schemas. Experts produced both subclass and equivalence mappings between the pairs listed in Table 2. BLOOMS and AM are compared in the last two columns [3] since WM or HCM produce only equivalence mappings and few of these exist in the LOD reference alignments. Both BLOOMS and AM use inferencing to produce some subclass mappings, BLOOMS using post-processing with the Jena reasoner and AM using its own inferencing techniques. To understand this influence, we performed an analysis on the LOD reference alignment for each pair to see the percentage of its mappings inferrable from its equivalence mappings, given in column 1. The * for M, B indicates an analysis was not possible since many BBC concepts could not be found directly in its file or even when opening the file using Protégé.

Table 2. LOD reference alignment pairs

Pair	% inferable	# mappings	BLOOMS		AM	
			Prec	Recall	Prec	Recall
F, D	87%	225	0.67	0.73	0.72	0.80
G, D	71%	41	0	0	0.26	0.68
M, D	20%	645	0.39	0.62	0.62	0.40
W, D	29%	519	0.70	0.40	0.58	0.35
M, B	*	528	0.63	0.78	0.48	0.16
S, F	27%	22	0.55	0.64	0.56	0.41
W, A	58%	366	0.42	0.59	0.48	0.43

BLOOMS has better recall except for F,D and G,D. F,D has 87% inferable mappings from its three equivalence relations. AM's use of other LOD ontologies and WordNet contributes to finding more correct mappings. For the G,D pair BLOOMS does not find the only equivalence relation SpatialThing = Place so that Jena cannot produce any of the inferable mappings. AM finds this mapping, likely from the comment field for SpatialThing including the word 'places.' AM finds 68% (recall) of the reference alignment mappings, very close to the 71% inferable mappings. Of the five remaining pairs, AM has better precision for M,D with the smallest percentage of inferable mappings. BLOOMS with Wikipedia finds more correct mappings since very few are from inferable relations. AM's lower recall corresponds with fewer inferable relations, but those that it does find are more likely correct with its 0.62 precision

References

1. Jain, P., Hitzler, P., Sheth, A. P., Verma, K., and Yeh, P. Z.: Ontology Alignment for Linked Open Data. In: Proceedings of the International Semantic Web Conference (ISWC) (2010)
2. Jain, P., Yeh, P. Z, Verma, K., Vasquez, G., Damova, M., Hitzler, P., and Sheth, A. P.: Contextual Ontology Alignment of LOD with an Upper Ontology:A Case Study with Proton. In Proceedings of the Extended Semantic Web Conference (ESWC) (2011)
3. Cruz, I., Palmonari, M., Caima, F., and Stroe, C.: Towards "On the Go" Matching of Linked Open Data Ontologies. In: Workshop on Discovering Meaning On the Go in Large Heterogeneous Data (LHD), The 22nd International Joint Conference on Artificial Intelligence (IJCAI-11) (2011)
4. Hertling, S. and Paulheim, H.: WikiMatch – Using Wikipedia for Ontology Matching, in: Seventh International Workshop on Ontology Matching (OM 2012) (2012)
5. Grutze, T., Bohm, C.. Naumann, F.: Holistic and Scalable Ontology Alignment for Linked Open Data. In: Workshop on Linked Data on the Web (LDOW) at WWW (2012)