Search Space Reduction for Post-Matching Correspondence Provisioning

Thomas Kowark and Hasso Plattner

Hasso Plattner Institute, Potsdam, Germany
{firstname.lastname}@hpi.de,
http://www.hpi.de

If users participate in ontology matching, the goal always is to minimize the amount of necessary interactions while maximizing the gains in alignment quality [2]. Interaction can either happen pre-matching (selection of matching systems or parameter tuning), during the matching process (judging intermediate results or providing sample correspondences), or post-matching (detecting incorrect correspondences and providing missing ones). In this paper, we evaluate an approach that aims to reduce post matching interactions by exploiting concept proximity within ontologies. An initial analysis of reference alignments available for OAEI revealed that, if a correspondence for one element (class or property) of an ontology exists, the probability that a correspondence also exists for a closely connected element is higher than for unconnected elements. Based on this finding, we extracted the closeness criteria depicted in Figure 1. For evaluation, we applied the criteria on candidate alignments that were created by top-performing systems of OAEI 2014 for the anatomy, library, and conference tracks. For each criterion, we determined which elements it would add to the task set, i.e., the selection of ontology elements a user should provide correspondences for. Based on these task sets (UT) we calculated the expected number of interactions (IE) it would on average take to provide all included correspondences (IC), if elements were presented to the user at random. To assess whether our selection technique is viable, we further compared this value to the amount of interactions it would take users on average to provide the same amount of missing correspondences, if tasks were randomly selected from the entirety of elements that are not included in correspondences after initial, automatic matching.

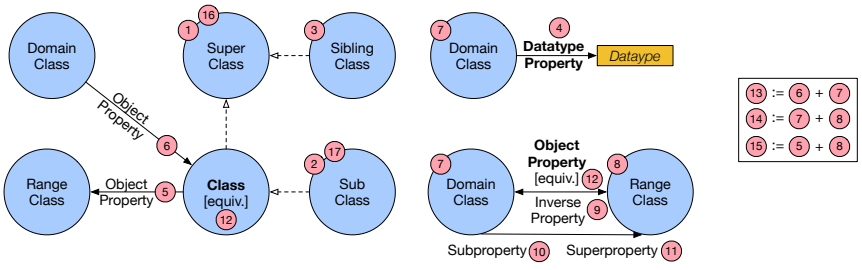

Figure 1. Connections considered for element proximity. Matched elements are depicted bold, ignored entities in italic. Visual Notation for OWL Ontologies (VOWL)[1]

The ratio between the two values is called task set compression. Minimal criteria sets denote the closeness criteria, which yield the corresponding task sets. Since we strive for minimization of user tasks, only the smaller ontology in terms of concept count was considered. As shown in Table 1, an average task set reduction of 60% could be achieved for the conference ontologies of OAEI, while increasing the recall from 0,62 to 0,956. For taxonomy-like ontologies, such as the ones used in the library and anatomy tracks, only marginal compression or even increase in interaction expectancy was achieved. Future work will therefore focus on such cases by finding other, more suitable task selection criteria and adapting existing ones, e.g., by limiting the depth of hierarchy traversal for class relationships. Furthermore, correspondences generated through different matcher settings (high precision vs. high recall) could be explored in addition to criteria based solely on ontology structures in order to yield smaller task sets with an increased potential success ratio for user interactions.

ontologies	$\|UT\|$	IC	IE	Compression	Minimal Criteria Sets	R_{cand}	R_{comp}
cmt-conference	12	4	10	0,2	[9, 17]	0,6	0,87
confof-conference	11	3	9	0,19	[13]	0,73	0,93
conference-edas	36	5	31	0,39	[1, 2]	0,65	0,94
ekaw-conference	38	8	35	0,51	[3, 6, 16]	0,6	0,92
conference-iasted	51	7	46	0,57	[1, 2, 5]	0,36	0,86
sigkdd-conference	13	4	11	0,25	[4, 9, 12]	0,6	0,87
confof-cmt	31	9	29	0,48	[2, 4, 6, {7, 8, 14}]	0,38	1
cmt-edas	27	4	22	0,35	[6]	0,69	1
cmt-ekaw	38	5	33	0,49	[5, 9, 17]	0,55	1
cmt-iasted	36	0	36	0,44	[]	1	1
sigkdd-cmt	7	1	4	0,13	[2]	0,92	1
confof-edas	26	8	24	0,43	[1, 3, 5, 9]	0,58	1
confof-ekaw	18	4	15	0,33	[1, 13, 15]	0,8	1
confof-iasted	30	5	26	0,45	[1, 3, 15]	0,44	1
confof-sigkdd	16	4	13	0,25	[4, 17]	0,57	1
ekaw-edas	63	11	59	0,71	[9, 13, 17]	0,52	1
edas-iasted	35	8	32	0,28	[2]	0,53	0,95
sigkdd-edas	20	5	18	0,37	[2, 4]	0,6	0,93
ekaw-iasted	73	3	56	0,76	[2, 15], [2, 13]	0,7	1
sigkdd-ekaw	22	4	18	0,32	[3, 15]	0,64	1
sigkdd-iasted	36	0	36	0,58	[]	0,87	0,87
avg(conference)	30,4	4,8	26,8	0,404		0,635	0,956
mouse-human	683	57	672	1,09	[2, 3, 16]	0,9	0,94
stw-thesoz	3604	169	3584	0,97	[2, 3]	0,78	0,84
fma-nci	1011	216	1007	1,08	[2, 3, 16]	0,85	0,93
fma-snomed	3485	1997	3484	0,99	[2, 3]	0,71	0,95
nci-snomed	10008	2281	10005	0,94	[1, 2, 3, 12]	0,67	0,82

Table 1. Overview about the maximal task reduction that could be achieved using minimal criteria sets. The used criteria are numbered according to Figure 1. R_{cand} is the recall achieved by the automatic matcher, R_{comp} the recall after user interaction.

References

1. Lohmann, S., Negru, S., Haag, F., Ertl, T.: VOWL 2: User-Oriented Visualization of Ontologies. In: Proceedings of the 19th International Conference on Knowledge Engineering and Knowledge Management. pp. 266–281. EKAW '14 (2014)
2. Shvaiko, P., Euzenat, J.: Ontology matching: State of the art and future challenges. IEEE Trans. on Knowl. and Data Eng. 25(1), 158–176 (Jan 2013)