

Using BioPortal as a Repository for Mediating Ontologies in Ontology Alignment

Weiguo Xia¹, Ernesto Jimenez-Ruiz², Valerie Cross¹

¹Computer Science and Software Engineering, Miami University, Oxford, OH USA 45056
{xiaw,crossv}@miamioh.edu

²Computer Science, University of Oxford, United Kingdom
ernesto.jimenez.ruiz@gmail.com

Leading ontology alignment (OA) systems add background knowledge sources to find mappings that string-based matchers are unable to find. Mediating ontologies (MOs) are typically pre-selected before the OA process begins. BioPortal Mediating Finder was developed to use BioPortal as a repository to dynamically select and access MOs for OA. BioPortal [1] provides access to over 370 biomedical ontologies and their mappings. The software development and results on the OAEI anatomy and large biomedical tracks with LogMap and several OA systems are reported. In [2] the fast-selection approach to find MOs was presented. Experiments produced the top five: Mouse Anatomy, SYN, Uberon, CL and EHDAA2 for the anatomy track. The following algorithm is used with each of these mediating ontologies.

Algorithm: Use of a mediating ontology with LogMap

Input: O1, O2: input ontologies; MO: mediating ontology; Output:M: output mappings;

1: M1 := LogMap(O1;MO)

2: M2 := LogMap(MO;O2)

3: MC := ComposeMappings(M1;M2)

4: M:= LogMap(O1;O2;MC)

5: return M

Uberon produced the highest f-measure of 0.8792; SYN followed with 0.8648. The dynamic selection of Uberon by the BioPortal Mediating Finder provides evidence this approach can find qualified mediating ontologies since Uberon has been pre-selected for several OA systems in the anatomy track. Two experiments determined a suitable minimum number of mediating ontologies required to produce a mapping for it to be included and a suitable confidence factor to capitalize on the high recall with only one mediating ontology. The resulting condition is shown in Table 1. LogMap using BioPortal Mediating Finder produces a higher F-measure than LogMap by itself.

Table 1. Original LogMap vs. LogMap+BioPortal Mediating Finder on Anatomy Track

	Prec	Recall	F-measure
Original LogMap	0.9125	0.8463	0.8782
LogMap+ Bioportal Mediating Finder (2 MOs AND Cf > 0.7) OR (1 MO AND Cf > 0.8)	0.8793	0.9037	0.8913

Experimental results on the large biomedical track with the two versions of each ontology are in Table 2. The top five mediating ontologies were determined using the fast-selection approach. The top four MOs for the anatomy track were in the top four MOs for the alignment task with the FMA. Only SYN was in the top five for SNOMED-NCI tasks. Table 2 shows results first for the small versions and then for the complete ontologies. LogMap is compared to that of Logmap + BMF.

Table 2. Alignment Results for Large Biomedical Tracs, FMA, NCI, and SNOMED

	Comparing LogMap to LogMap + BioPortal Mediating Finder (BMF)					
	FMA-NCI		FMA-SNOMED		SNOMED-NCI	
<i>SMALL</i>	LogMap	+ BMF	LogMap	+ BMF	LogMap	+BMF
Precision	0.9541	<i>0.9281</i>	0.9643	0.9521	0.8677	0.8950
Recall	0.8598	0.8793	0.6680	0.6792	0.5602	0.5854
F-measure	0.9045	0.9030	0.7892	0.7982	0.6809	0.6963
<i>WHOLE</i>	LogMap	+ BMF	LogMap	+ BMF	LogMap	+BMF
Precision	0.8689	<i>0.7443</i>	0.8753	<i>0.8220</i>	0.8677	0.8590
Recall	0.7939	0.8570	0.5967	0.6189	0.5602	0.5854
F-measure	0.8297	0.7967	0.7096	0.7061	0.6809	0.6963

Three (in bold) have improved f-measures. The large number of concepts in FMA and NCI result in many new mappings, significantly reducing precision (in italics) although recall increases. More sophisticated filtering is needed as in the anatomy track. BMF with other recent OAEI systems, two top (AML and GOMMA) and two poorer (Hertuda and AOT) is in Table 3. F-measures (bold) increases though slightly for Hertuda.

Table 3. OA Systems using BioPortal Mediating Finder

	Prec	Recall	F-meas		Prec	Recall	F-meas
AML	0.9548	0.8219	0.8834	AOT	0.4352	0.7751	0.5574
+ BMF	0.9319	0.8938	0.9125	+ BMF	0.4325	0.7777	0.5559
Hertuda	0.6892	0.6728	0.6809	GOMMA	0.9505	0.7975	0.8673
+ BMF	0.7024	0.6735	0.6811	+ BMF	0.9395	0.7995	0.8639

References

1. Ghazvinian, A., Noy, N.F., Jonquet, C., Shah, N.H., Musen, M.A.: What four million mappings can tell you about two hundred ontologies. In: Int'l Sem. Web Conf. (ISWC) (2009)
2. Chen, X., Xia, Weiguo, Jiménez-Ruiz, E., Cross, V.V.: Extending an Ontology Alignment System with BioPortal: A Preliminary Analysis. In: Ontology Matching Workshop, International Semantic Web Conference (ISWC) (2014)