

Ontology Mapping for the Laboratory Analytics Domain

Ian Harrow¹, Thomas Liener¹, and Ernesto Jimenez-Ruiz^{2,3}.

¹Ontologies Mapping Project, Pistoia Alliance, USA ²City, University of London, UK and ³SIRIUS, University of Oslo, Norway.

1. Introduction

The Pistoia Alliance was established ten years ago to promote innovation by industry through pre-competitive collaboration to reduce the barriers to innovation. The Ontologies Mapping Project [1] was established in 2016 to enable better tools and services for mapping between ontologies and to establish best practices for ontology management in the Life Sciences.

2. Extendibility of the Ontology Mapping algorithm

We have reported already on the development of the algorithm, Paxo for mapping between public ontologies hosted by the Ontology Lookup Service (OLS) and the Ontology Mapping Repository (OxO) at EMBL-EBI [2, 3]. Paxo was used previously to map between public ontologies in the phenotype and disease domain, while here we report on mapping in the laboratory analytics domain.

3. Selected public Ontologies for Mapping

Eleven public ontologies were selected from the laboratory analytics domain for mapping with Paxo as listed below:

Lab analytics domain	Ontology name	Short name
Chemistry	Chemical Information Ontology	CHEMINF
Chemistry	Physico-Chemical Methods and Properties Ontology	FIX
Chemistry	Allotrope Merged Ontology Suite	AFO
Chemistry	Chemical Methods Ontology	CHMO
Biology	Ontology for Biomedical Investigations	OBI
Biology	Eagle-I Research Resource Ontology	ERO
Biology	Mass Spectrometry Ontology	MS
Biology	BioAssay Ontology	BAO
Biology	Experimental Factors Ontology	EFO
General	National Cancer Institute Thesaurus	NCIT
General	Medical Subject Headings	MESH

4. Perceived value of Ontology Mappings

Each ontology was scored for perceived value (PV) by the 9 members of the project team, from numerous pharmaceutical and biotechnology companies. Each ontology was assigned a score of 3 for high PV, 2 for medium PV and 1 for low PV and 0 for no PV by each of the 9 team members. This gave the total PV score (a simple summation of scores) for each of the 54 mappings predicted by Paxo, which informed our priorities for evaluation:

Ontologies	PVO1	OBI	1+2	ERO	1+2	MS	1+2	BAO	1+2	EFO	1+2
PVO2		23		15		19		26		26	
ERO	15	OBI - ERO	38								
MS	19	OBI - MS	42	ERO - MS	34						
BAO	26	OBI - BAO	49	ERO - BAO	41	MS - BAO	45				
EFO	26	OBI - EFO	49	ERO - EFO	41	MS - EFO	45	BAO - EFO	52		
MESH	24	OBI - MESH	47	ERO - MESH	39	MS - MESH	43	BAO - MESH	50	EFO - MESH	50
NCIT	25	OBI - NCIT	48	ERO - NCIT	40	MS - NCIT	44	BAO - NCIT	51	EFO - NCIT	51

n = 20

Ontologies	PVO1	CHEMINF	1+2	FIX	1+2	AFO	1+2	CHMO	1+2
PVO2		13		11		24		22	
FIX	11	CHEMINF - FIX	24						
AFO	24	CHEMINF - AFO	37	FIX - AFO	35				
CHMO	22	CHEMINF - CHMO	35	FIX - CHMO	33	AFO - CHMO	46		
MESH	24	CHEMINF - MESH	37	FIX - MESH	35	AFO - MESH	48	CHMO - MESH	46
NCIT	25	CHEMINF - NCIT	38	FIX - NCIT	36	AFO - NCIT	49	CHMO - NCIT	47

n = 14

Ontologies	PVO1	OBI	1+2	ERO	1+2	MS	1+2	BAO	1+2	EFO	1+2
PVO2		23		15		19		26		26	
CHEMINF	13	CHEMINF - OBI	36	CHEMINF - ERO	28	CHEMINF - MS	32	CHEMINF - BAO	39	CHEMINF - EFO	39
FIX	11	FIX - OBI	34	FIX - ERO	26	FIX - MS	30	FIX - BAO	37	FIX - EFO	37
AFO	24	AFO - OBI	47	AFO - ERO	39	AFO - MS	43	AFO - BAO	50	AFO - EFO	50
CHMO	22	CHMO - OBI	45	CHMO - ERO	37	CHMO - MS	41	CHMO - BAO	48	CHMO - EFO	48

n = 20

5. Evaluation of selected Ontology Mapping sets

Thirteen mappings with high total PV scores and unique matches were selected for evaluation of recall and precision:

Mapping	LOOM standard	Paxo matches to LOOM	Recall % (c.f. LOOM)	Paxo uniques	Precision %
EFO - NCIT	1,303	1,249	96%	3,048	95%
EFO - MESH	776	736	95%	3,002	73%
AFO - NCIT	NA			528	68%
BAO - NCIT	70	68	97%	496	62%
CHMO - NCIT	122	113	93%	492	93%
OBI - NCIT	128	120	94%	384	75%
CHMO - MESH	68	61	90%	374	95%
ERO - NCIT	165	149	94%	353	88%
BAO - MESH	29	19	66%	254	70%
AFO - CHMO	NA			239	45%
ERO - MESH	74	66	89%	211	88%
MS - NCIT	68	60	88%	209	50%
AFO - MESH	NA			149	65%

The parameters of Paxo were selected to balance recall (matches missing from the LOOM baseline standard) and precision (correct matches from random sampling from unique matches where n=60). Recall ranged from 66% to 97% while precision for unique matches ranged from 45% to 95% for each mapping. These predicted mapping sets will be made accessible openly via the project web page [4].

6. Summary and Future Plans

Fifty-four ontology mappings were predicted using the Paxo algorithm which demonstrates how it can be applied to any pair of ontologies hosted by OLS and OXO at EMBL-EBI, within a single domain where overlap of class concepts is likely to be found.

As no hand-curated gold standard mappings exist to measure recall, in the near future we will use a panel of numerous algorithms to generate a set of silver standard mappings from a minimum of three consensus votes as we have published previously [6]. The panel of algorithms are participants in the annual challenge for Ontology Alignment Evaluation Initiative (OAEI) [5, 6] which included the top performing LogMap [7] and AML [8], in addition to the purely lexical algorithm, LOOM [9] which served as a baseline standard [6].

Future work may include crowd validation of predicted mappings and further mapping between ontologies in the clinical domain.

Acknowledgements

We would like to express gratitude to all the Pistoia Alliance Ontology Mapping project team members and their parent organisations who contributed expertise, time and funding. EJR was supported by the AIDA project, funded by the Alan Turing Institute, and the SIRIUS Centre for Scalable Data Access (Research Council of Norway, project no.: 237889).

References

1. <http://www.pistoiaalliance.org/projects/ontologies-mapping>
2. <https://doi.org/10.6084/m9.figshare.7346057.v1>
3. <https://www.ebi.ac.uk/spot/ontology>
4. <https://www.pistoiaalliance.org/projects/current-projects/ontologies-mapping>
5. <http://oaei.ontologymatching.org>
6. Harrow I *et al* (2017) Matching disease and phenotype ontologies in the ontology alignment evaluation initiative. *J Biomed Semantics*. 8(1), 55
7. Jimenez-Ruiz E and Cuenca Grau B (2011) LogMap: Logic-based and Scalable Ontology Matching. *International Semantic Web Conference* (1) 2011, 273-288
8. Faria D *et al* (2018) Tackling the challenges of matching biomedical ontologies. *J Biomed Semantics* 9(1), 4
9. Ghazvinian A *et al* (2009). Creating mappings for ontologies in biomedicine: simple methods work. *AMIA. Annual Symposium (AMIA 2009) San Francisco, CA*