Pistoia Alliance
ONTOLOGIES MAPPING

Ontologies Mapping for the Laboratory Analytics Domain

Key Information

Author Names:

Ian Harrow (Project Manager), Thomas Liener (Paxo author & mapping work) and the Ontologies Mapping Project Team (Funders & Partners below)

Project contact details:

Ian.harrow@pistoiaalliance.org

Project website:

https://www.pistoiaalliance.org/projects/current-projects/ontologies-mapping

Funders (Phase)

BIOVIA 3DS (2,3), GSK (All), Roche (All), Amgen (3), AstraZeneca (3,4), Accenture (3), Bayer (3), Merck & Co (1,2), Novartis (1,2), and Bristol Myers-Squibb (4)

Partners (Phase)

OAEI Organiser, Dr Ernesto Jiménez-Ruiz (2,3,4) EMBL-EBI (3,4), Allotrope Foundation (4), Osthus (All), Eagle Genomics (All), SciBite (All), Linguamatics (All), Novartis (3,4), AbbVie (3,4), Bayer (4) and Elsevier (All)

References

OAEI annual challenge

Ontology Alignment Evaluation Initiative

http://oaei.ontologymatching.org

The Phenotype and disease track has been organised by the Ontology Mapping project over the last four years.

Technical paper from OAEI 2016 authored by the OM project

Ian Harrow, Ernesto Jiménez-Ruiz, Andrea Splendiani, Martin Romacker, Peter Woollard, Scott Markel, Yasmin Alam-Faruque, Martin Koch, James Malone and Arild Waaler (2017)

"Matching disease and phenotype ontologies in the ontology alignment evaluation initiative". *Journal of Biomedical Semantics* 2017
DOI 10.1186/s13326-017-0162-9

Review article authored by the OM project

Ian Harrow, Rama Balakrishnan, Ernesto Jiménez-Ruiz, Simon Jupp, Jane Lomax, Jane Reed, Martin Romacker, Christian Senger, Andrea Splendiani, Jabe Wilson, Peter Woollard

"Ontology mapping for semantically enabled applications" *Drug Discovery Today* 2019
DOI: 10.1016/j.drudis.2019.05.020

Book Recommendation

Jerome Euzenat and Pavel Shvaiko (2013) 2nd ed. Ontology Matching Springer-Verlag
DOI: 10.1007/978-3-642-38721-0

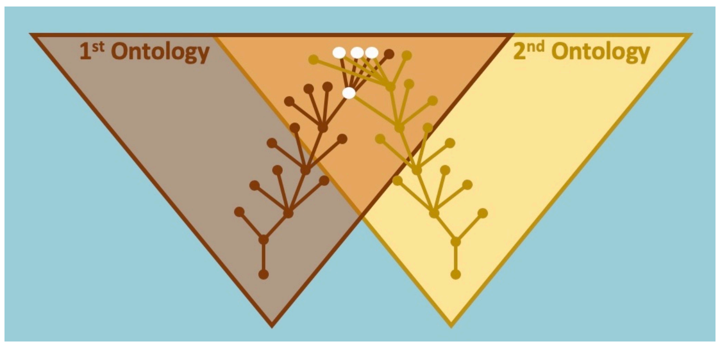
Business Challenge

A growing number of ontologies underpin numerous important applications such as semantic search, data integration, fact extraction and AI/machine learning.

Use of different ontologies within same data domain hampers interoperability and application. This can be solved by mapping between ontologies, as described below.

What are Ontology Mappings?

An Ontology Mapping comprises of pairwise matches between two ontologies. They provide a modular engineering solution to expand coverage at reduced cost of maintenance for building applications.



Summary

- Thomas Liener and Simon Jupp (SPOT group, led by Helen Parkinson at EMBL-EBI) developed the Paxo algorithm for ontology mapping which is openly available on GitHub.
- Mapping "standards" from NCBO BioPortal and the annual OAEI challenge help us to evaluate the quality of the predicted mappings.
- Publication of the review article "Ontology Mapping for Semantic Applications" in Drug Discovery Today.
- Our results demonstrate the successful application of the Paxo mapping algorithm to ontologies in two different domains.
    - Previously for the phenotype and disease domain
    - Here for the laboratory analytics domain (right panel)
    - We have demonstrated that the mapping algorithm can be applied to any pair of ontologies hosted by EMBL-EBI.

Plans

- Ideas being explored for 2020 include:
    - Crowd validation of predicted mappings in the OxO repository of ontology mappings, hosted by EMBL-EBI
    - Undertake Ontology Mapping for the clinical domain

Contact Ian Harrow for further info

Selected Ontologies

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

- Eleven public ontologies were selected for mapping.

Perceived value of Ontology Mappings

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

- The Paxo algorithm generated 54 ontology mappings.
- Each ontology was scored for perceived value by 9 members of the project team, representing different organisations
- Perceived Values (PVs) of each ontology in a mapping pair gave the total score which informed our priorities for evaluation of 13 mappings (below).

Evaluation of selected Ontology Mappings

Ontology Mapping	Recall % (Silver 3 votes)	Paxo uniques (Silver 3 votes)	Recall % (Silver 2 votes)	Paxo uniques (Silver 2 votes)	Precision % (Uniques, not exact)
AFO - CHMO	99%	40	87%	24	45%
BAO - NCIT	99%	206	94%	189	62%
EFO - NCIT	98%	1,385	94%	1,189	95%
OBI - NCIT	96%	178	90%	165	75%
ERO - NCIT	96%	190	94%	181	88%
EFO - MESH	94%	1,923	84%	712	73%
AFO - NCIT	93%	203	66%	137	68%
MS - NCIT	92%	111	78%	93	50%
CHMO - NCIT	89%	235	76%	221	93%
ERO - MESH	88%	113	68%	67	88%
CHMO - MESH	86%	165	51%	116	95%
AFO - MESH	81%	47	52%	27	65%
BAO - MESH	70%	100	47%	63	70%

- Thirteen mappings were selected for evaluation of:
    - Recall (c.f. silver standard, 3 & 2 vote consensus)
    - Precision (from sampling not exact unique mappings

(The evaluation method is detailed in Harrow *et al* 2017)

Pistoia Alliance: Lowering barriers to R&D innovation

The Pistoia Alliance is a global, not-for-profit alliance of life science companies, vendors, publishers, and academic groups that work together to lower barriers to innovation in R&D.

Our members collaborate as equals on open projects that generate significant value for the worldwide life sciences community.

www.pistoiaalliance.org
@pistoiaalliance