

Progress Towards an Ontology Mapping Service

Ian Harrow¹, Martin Romacker², Andrea Splendiani³, Peter Woollard⁴, Tom Plasterer⁵, Rainer Winnenburger⁶, Scott Markel⁷, Siddharth Mehta⁸, Chris Piddington⁹, Yasmin Alam-Faruque¹⁰, Rama Balakrishnan¹¹, Jane Lomax¹², Jane Reed¹³, Christian Senger¹⁴, Jabe Wilson¹⁵, Filip Pattyn¹⁶, Ernesto Jimenez-Ruiz^{17,18}, Simon Jupp¹⁹ and Thomas Liener¹⁹

¹Ontologies Mapping Project, Pistoia Alliance, USA, ²Roche Innovation Center Basel, Switzerland, ³Novartis, Switzerland, ⁴GlaxoSmithKline, UK, ⁵AstraZeneca, USA, ⁶AbbVie, USA, ⁷Dassault Systèmes BIOVIA, USA, ⁸Accenture, USA, ⁹Amgen, USA, ¹⁰Eagle Genomics, UK, ¹¹Genentech, USA, ¹²SciBite, UK, ¹³Linguamatics, UK, ¹⁴OSTHUS, Germany, ¹⁵Elsevier RELX, UK, ¹⁶Ontoforce, Belgium, ¹⁷The Alan Turing Institute, UK, ¹⁸University of Oslo, Norway, ¹⁹European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), UK

Keywords: Ontology Mapping Service: Ontology Matching Algorithm: Ontology Mapping Evaluation

1. Introduction

The Pistoia Alliance was established nearly 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. The project is now focussed on the development of an ontology mapping service (OMS).

2. Motivation for an Ontology Mapping Service

Ontologies are dynamic resources which evolve over time. Common changes include: class addition, class modification, combination of classes and hierarchical relationships. Hence, ontology mappings cannot be static and they need to evolve constantly synchronised with their source ontologies. Therefore, any ontology mapping needs to be provided not only as a one-off process, but as an ongoing service.

3. Service Requirements

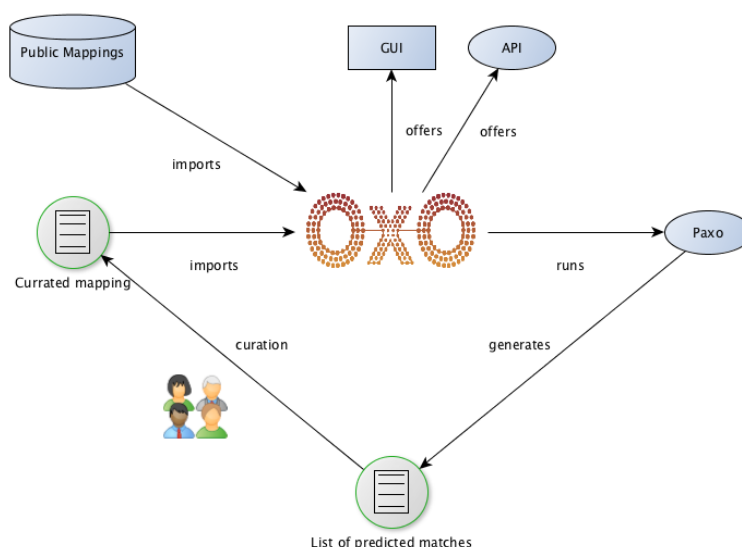
Three major requirements or deliverables for an OMS were identified: 1) Development of an algorithm for mapping between ontologies hosted by the Ontology Lookup Service (OLS) and the Ontology Mapping Repository (OxO) at EMBL-EBI; 2) Prediction of ten ontology mappings between 5 public ontologies in the phenotype and disease domain 3) Evaluation of the 10 predicted ontology mappings by comparison to a silver standard and via random sampling and assessment by experts.

The 5 public ontologies mapped by this service are: 1) Human Phenotype Ontology (HPO) 2) Human Disease Ontology (DOID) 3) Orphanet Rare Diseases Ontology (ORDO), Mammalian Phenotype Ontology (MP) and the widely used taxonomy, Medical Subject Headings (MeSH).

4. Service Implementation and Optimisation

The architectural relationship between the OxO mapping repository and the new ontology mapping prediction algorithm, Paxo, is shown in Figure 1. OxO imports cross references from public sources and runs the Paxo algorithm to create a mapping of predicted matches. Paxo combines evidence from existing mappings held in OxO with standard lexical similarity measures to generate a set of predicted matches. These matches can be reviewed and validated by curators using a simple spreadsheet format, and made accessible to the community by re-loading them back into OxO.

Figure 1: The relationship between the OxO mapping repository and the new mapping algorithm Paxo.



The parameters of Paxo were optimised for the selected ontologies to find a balance between recall (matches missing from the silver standards) and precision (correct matches from random sampling). By the end of the prototype service period, the algorithm parameters had recall that ranged 63-82% while precision ranged from 32-69% for the 6 mappings between the 4 public ontologies. Recall ranged from 81-100% for the 4 mappings between MeSH and the 4 public ontologies with precision greater than 90%. The predicted mappings generated by this service will become accessible openly via the OxO repository at EMBL-EBI [7].

As no hand-curated gold standard mappings exist to measure recall, we used a panel of numerous algorithms to generate a set of silver standard mappings from a minimum of three consensus votes. The panel of algorithms are participants in the annual challenge for Ontology Alignment Evaluation Initiative (OAEI) [2, 3] which included the top performing LogMap [4] and AML [5], in addition to the purely lexical algorithm, LOOM [6] which served as a baseline.

5. Future Plans

After achieving the encouraging results of the prototype service for mapping ontologies and taxonomies in the phenotype and disease domain, we are now planning to optimise the parameters of Paxo to predict mappings between ontologies and terminologies in the biological and/or chemical laboratory domain. This will test the generic applicability of the new ontology matching algorithm.

Acknowledgements

We would like to express gratitude for funding of the Ontologies Mapping project by some paying members of the Pistoia Alliance Inc. We wish to thank Helen Parkinson for helpful discussions which have influenced this work. SJ benefited from funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 654248 (CORBEL). EJ was supported by the AIDA project, funded by the UK Government's Defence & Security Programme in support of 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. <http://oaei.ontologymatching.org>
3. Harrow I *et al* (2017) Matching disease and phenotype ontologies in the ontology alignment evaluation initiative. *J Biomed Semantics*. 8(1), 55
4. Jimenez-Ruiz E and Cuenca Grau B (2011) LogMap: Logic-based and Scalable Ontology Matching. *International Semantic Web Conference* (1) 2011, 273-288
5. Faria D *et al* (2018) Tackling the challenges of matching biomedical ontologies. *J Biomed Semantics* 9(1), 4
6. Ghazvinian A *et al* (2009). Creating mappings for ontologies in biomedicine: simple methods work. *AMIA. Annual Symposium (AMIA 2009) San Francisco, CA*
7. <https://www.ebi.ac.uk/spot/oxo>