Ontologies Guidelines for Best Practice and a Process to Evaluate Existing Ontologies Mapping Tools and Algorithms

Ian Harrow[1], Martin Romacker[1], Andrea Splendiani[1], Stefan Negru[1], Peter Woollard[1], Scott Markel[1], Yasmin Alam-Faruque[1], Martin Koch[1], Erfan Younesi[1], James Malone[1] and Ernesto Jimenez-Ruiz[2]

[1] Pistoia Alliance Ontologies Mapping Project, Pistoia Alliance Inc. USA
ian.harrow@pistoiaalliance.org
[2] Department of Computer Science, University of Oxford, UK
ernesto@cs.ox.ac.uk

Abstract. This extended abstract presents an ongoing work by the Pistoia Alliance Ontologies Mapping project to develop user requirements for an ontologies mapping service.

Keywords: Ontologies, Mapping, Guidelines, Tool, Algorithm, Evaluation

1   Introduction

The Pistoia Alliance Ontologies Mapping project[1] was set up to find or create better tools or services for mapping between ontologies in the same domain and to establish best practices for ontology management in the Life Sciences. It was proposed through the Pistoia Alliance Ideas Portfolio Platform[2] which was selected by the Pistoia Alliance Operations Team for development of a formal business case.

The project has delivered a set of guidelines for best practice which build on existing standards. We show how these guidelines can be used as a "checklist" to support the application and mapping of source ontologies in the disease and phenotype domain. These guidelines are accessible on a public wiki.[3]. The project has contributed a process to evaluate existing ontologies mapping tools and a new phenotype track at OAEI (OM) 2016 to support evaluation of ontology matching (OM) algorithms.
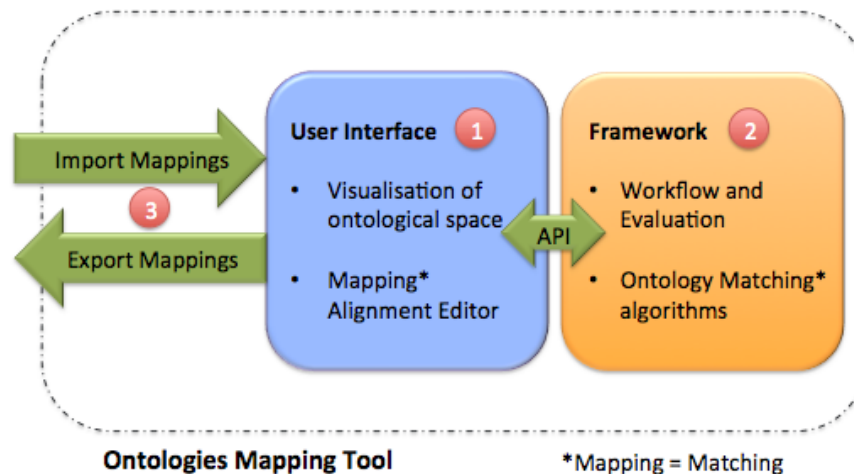
Fig. 1. The three major aspects of the Tool requirements.

2 User Requirements and Evaluation of existing Ontologies Mapping Tools

Another important output of this project was to specify the requirements for an Ontologies Mapping Tool. The three major aspects of the user requirements are illustrated in Figure 1. This shows the specification of the 1) User interface; 2) Framework, including ontology matching algorithms and 3) Import and export of mappings.

These requirements were used in a preliminary survey that established that such tools already exist which substantially meet them. Therefore, we have developed a formal process to define and submit a request for information (RFI) from existing ontologies mapping tool providers. We will summarise our findings from this evaluation of seven ontologies mapping tools from academic and commercial providers on the poster. The guidelines and RFI materials, including the specific details of the requirements are accessible on the same public wiki, mentioned previously.

3 Evaluation of Ontology Matching Algorithms for Disease and Phenotype

A critical component of any Ontologies Mapping tool is the embedded ontology matching algorithm. Therefore, the Pistoia Alliance Ontologies Mapping

---

[1] http://www.pistoiaalliance.org/projects/ontologies-mapping

[2] IP3:https://www.qmarkets.org/live/pistoia/home

[3] https://pistoiaalliance.atlassian.net/wiki/display/PUB/Ontologies+Mapping+Resources

Project is supporting development and evaluation of ontology matching algorithms though sponsorship and organisation of the new Disease and Phenotype track for OAEI (OM) 2016,[4] which is also be summarised in this section. This new track has been organised because currently, mappings between ontologies in a given data domain are mostly curated by bioinformatics and disease experts in academia or industry, who would benefit from automation of their procedures. This could be accomplished through implementation of ontology matching algorithms into their existing workflow environment or investment in an ontologies mapping tool for management of the ontologies mapping life cycle.

3.1 OAEI Disease and Phenotype Dataset

There will be two tasks for this novel OAEI track which will comprise of pairwise alignment of:[5]

- Human Phenotype Ontology[6] (HPO) to Mammalian Phenotype Ontology[7] (MP), and
- Human Disease Ontology[8] (DOID) to the Orphanet and Rare Diseases Ontology[9] (ORDO).

The first task is important for translational science where human phenotype includes inherited diseases and mammalian phenotype ontology which is based on rodents as a model mammalian organism for many laboratory studies, including gene knock out. The second task includes representation of rare human diseases in both ontologies which are of fundamental importance for understanding how genetic variation can cause disease.

We have extracted a "baseline" reference alignments for the track based on the available BioPortal mappings [2]. Most of the BioPortal mappings are automatically generated by the LOOM system;[10] therefore this BioPortal-based reference alignment will only be considered as a baseline since it is incomplete or may contain erroneous mappings.

3.2 OAEI Evaluation Process

The evaluation of the Disease and Phenotype Track will be run with support of the SEALS infrastructure.[11] Systems will be evaluated and ranked according to the following criteria:

---

[4] The OAEI is an annual campaign for the evaluation of ontology matching systems [1]: `http://oaei.ontologymatching.org/`

[5] `http://oaei.ontologymatching.org/2016/phenotype/index.html`

[6] `http://bioportal.bioontology.org/ontologies/HP`

[7] `http://bioportal.bioontology.org/ontologies/MP`

[8] `http://bioportal.bioontology.org/ontologies/DOID`

[9] `http://bioportal.bioontology.org/ontologies/ORDO`

[10] `http://www.bioontology.org/wiki/index.php/BioPortal_Mappings`

[11] `http://oaei.ontologymatching.org/2016/seals-eval.html`

- Precision and Recall with respect to a voted reference alignment that will be built automatically to generate consensus voting for the outputs of the participating systems using the standard evaluation process.
- Recall with respect to manually generated mappings for three areas (carbohydrate, obesity and breast cancer). This will be a novel aspect of the evaluation.
- Manual assessment of a subset of the generated mappings, especially the ones that are not suggested by other systems. This will be a novel aspect of the evaluation.
- Performance in other tracks will also be taken into account (especially the interactive session at the OAEI 2016 campaign [3]).

4 Summary and Plans

This poster describes access to ontologies guidelines and a process to evaluate existing tools and algorithms for ontology matching. The Ontologies Mapping project is currently defining the user requirements for a sustainable service, which will use such OM tools or algorithms. We will conduct a survey of Pistoia Alliance members to understand the need for such a service and whether it should be implemented in future.

Acknowledgements

References

1. Cheatham, M., Dragisic, Z., Euzenat, J., Faria, D., Ferrara, A., Flouris, G., Fundulaki, I., Granada, R., Ivanova, V., Jiménez-Ruiz, E., Lambrix, P., Montanelli, S., Pesquita, C., Saveta, T., Shvaiko, P., Solimando, A., dos Santos, C.T., Zamazal, O.: Results of the ontology alignment evaluation initiative 2015. In: Proceedings of the 10th International Workshop on Ontology Matching. (2015) 60–115
2. Fridman Noy, N., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A.D., Chute, C.G., Musen, M.A.: BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Research **37**(Web-Server-Issue) (2009)
3. Dragisic, Z., Ivanova, V., Lambrix, P., Faria, D., Jimenez-Ruiz, E., Pesquita, C.: User validation in ontology alignment. In: International Semantic Web Conference. (2016)