View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by City Research Online

City Research Online

City, University of London Institutional Repository

Citation: Jimenez-Ruiz, E., Hovland, D., Slaughter, L., Håndstad, T. and Waaler, A. (2018). On adopting Ontology Alignment techniques within the Phenotype Acquisition Process. CEUR Workshop Proceedings, 2042,

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: http://openaccess.city.ac.uk/id/eprint/22950/

Link to published version:

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

On adopting Ontology Alignment techniques within the Phenotype Acquisition Process*

E. Jiménez-Ruiz[1], D. Hovland[1], L. Slaughter[1,2], T. Håndstad[2], A. Waaler[1]

[1] Department of Informatics, University of Oslo, Norway [2] Oslo University Hospital, Norway

1 Introduction

The work presented in this paper is framed within the context of the BigMed project, a project funded by the Norwegian Research Council. One of the objectives of BigMed is to enhance the phenotype acquisition process in newborns with a monogenetic disorder, one of the four patient groups studied in the project. The use of the Human Phenotype Ontology (HPO) [1] to tag phenotypes and systems like PhenoTips have substantially contributed to the overall phenotype acquisition workflow. PhenoTips [2] is a system for the acquisition of phenotypic information in patients with a genetic disease. PhenoTips also suggests, given a selected set of HPO terms, candidate diagnoses using OMIM (Online Mendelian Inheritance in Man) codes, and related genes for a subsequent genetic test. Although PhenoTips represents a fantastic effort, we believe it could be extended with suitable Semantic Web solutions. In this paper, we present the first steps to adopt ontology alignment techniques to contribute to the diagnostic process.

2 Preliminary evaluation

We have conducted an evaluation about the coverage of links between phenotypes (HPO terms) and diseases (OMIM terms) obtained using ontology alignment systems.

Reference datasets. We use as reference the *phenotype_annotation_hpoteam.tab* file (we refer to this file as hpo-annotations) which contains annotations linking HPO terms with OMIM (disease) terms.[1] We also use the annotation file *new_phenotype* (we refer to this file as phenotips-annotations) used in PhenoTips as part of the Bayesian Ontology Query Algorithm[2] (BOQA) module to suggest diagnosis. From the *hpo-annotations* we extracted 81,154 links (pairs hpo term-omim term), while 98,359 links were obtained from the *phenotips-annotations* file.

Evaluation set-up. We have used the mappings computed by participating systems in the *Disease and Phenotype* track [3] of the Ontology Alignment Evaluation Initiative [4] (OAEI). The 2017 edition of the *Disease and Phenotype* track consists of four tasks.[3] In this paper we focus on the results in the task involving the matching of the HPO and OMIM ontologies. The first row in Table 1 shows the number of mappings computed by the systems AML, LogMap, LogMapBio, DiSMatch, and mappings extracted from BioPortal. *All-Syst* represents the union of all computed/extracted mappings.

The mappings computed in the *Disease and Phenotype* track represent phenotype-to-phenotype correspondences (mostly equivalence) between HPO and OMIM terms.

* Funded by the BIGMED and HealthInsight projects, and the SIRIUS Centre (no.: 237889).

[1] `http://compbio.charite.de/jenkins/job/hpo.annotations/`

[2] PhenoTips fork of BOQA: `https://github.com/phenotips/boqa`

[3] `http://oaei.ontologymatching.org/2017/phenotype/`

Table 1. Computed mappings or links between HPO and OMIM terms.

Link Type	AML	LogMap	LogMapBio	DiSMatch	BioPortal	All-Syst
Phenotype-to-Phenotype	6,344	7,202	7,726	7,680	3,768	11,510
Phenotype-to-Disease	42,354	35,825	39,157	32,990	26,984	52,413

Table 2. Coverage of the links driven by the automatic ontology alignment systems.

System	hpo-annotations		phenotips-annotations	
	Precision	Recall	Precision	Recall
AML	0.880	0.459	0.848	0.349
LogMap	0.897	0.396	0.867	0.316
LogMapBio	0.889	0.429	0.860	0.342
DiSMatch	0.832	0.338	0.816	0.274
BioPortal	0.938	0.312	0.920	0.252
All-Syst	0.804	0.519	0.777	0.414

In order to obtain phenotype-to-disease links between HPO and OMIM terms, as in the reference datasets presented above, we exploited the semantic information in OMIM and the computed mappings. We extracted the associated OMIM disease terms for a given OMIM phenotype term using the ontology properties *manifestation_of* and *inheritance_type_of*. The second row in Table 1 shows the number of phenotype-to-disease links derived from the phenotype-to-phenotype mappings.

Results and Discussion. Table 2 shows the coverage of the links derived from the automatically computed alignments.[4] The results in terms of Recall are not impressive and only when considering all alignment sets we get near to 50% Recall. Giving a closer look at the results, we noticed that a few OMIM ids present in the annotation files were missing in the OMIM ontology.[5] However, the missing OMIM ids are not significant enough to explain the low Recall values. We also observed that a large number (i.e., $>$1,000) of HPO terms occurring in the links extracted from the annotation files were not aligned to OMIM terms by any of the ontology alignment systems. There may be three possible explanations: *(i)* the alignment systems fail to discover relevant correspondences, *(ii)* the OMIM ontology does not cover all phenotype terms from the HPO ontology, and/or *(iii)* the description of related diseases to a phenotype in the OMIM ontology is limited. Future work aims at clarifying these possible limitations.

Although the results regarding Recall are not encouraging, the results with respect to Precision may be considered promising since there are links derived by the automatically computed correspondences that are not present in the annotation files. These new links, however, require manual curation to assess their validity.

References

1. Köhler, S., et al.: The human phenotype ontology in 2017. Nucleic Acids Res. **45**(D1) (2017)
2. Girdea, M., et al.: PhenoTips: Patient Phenotyping Software for Clinical and Research Use. Human Mutation **34**(8) (2013) 1057–1065
3. Harrow, I., Jimenez-Ruiz, E., et al.: Matching disease and phenotype ontologies in the ontology alignment evaluation initiative. Journal of Biomedical Semantics (in press) (2017)
4. Achichi, M., et al.: Results of the Ontology Alignment Evaluation Initiative 2017. In: 12th International Workshop on Ontology Matching (OM). (2017) http://oaei.ontologymatching.org/.

[4] The coverage has been calculated in terms of Precision and Recall with respect to the (reference) link sets extracted from the *hpo-annotations* and *phenotips-annotations* files.

[5] OMIM v.2016AB: `https://bioportal.bioontology.org/ontologies/OMIM`