MEDINFO 2013
C.U. Lehmann et al. (Eds.)

doi:10.3233/978-1-61499-289-9-961

Integrating the Human Phenotype Ontology into HeTOP Terminology-Ontology Server

Julien Grosjean[a], Tayeb Merabti[a], Lina F. Soualmia[a,b], Catherine Letord[a], Jean Charlet[b], Peter N. Robinson[c,d], Stéfan J. Darmoni[a,b]

[a] *CISMeF & TIBS, LITIS EA 4108, Rouen University Hospital, Rouen, France*
[b] *INSERM, Unité Mixte de Recherche en Santé (UMR_S) 872, équipe 20, Paris, France*
[c] *Institute for Medical Genetics and Human Genetics, Charité Universitätsmedizin Berlin, Augustenburger Platz 1, Berlin, Germany*
[d] *Berlin-Brandenburg Center for Regenerative Therapies, Germany*

Abstract and objective

The Human Phenotype Ontology (HPO) is a controlled vocabulary which provides phenotype data related to genes or diseases. The Health Terminology/Ontology Portal (HeTOP) is a tool dedicated to both human beings and computers to access and browse biomedical terminologies or ontologies (T/O). The objective of this work was to integrate the HPO into HeTOP in order to enhance both works. This integration is a success and allows users to search and browse the HPO with a dedicated interface. Furthermore, the HPO has been enhanced with the addition of content such as new synonyms, translations, mappings. Integrating T/O such as the HPO into HeTOP is a benefit to vocabularies because it allows enrichment of them and it is also a benefit for HeTOP which provides a better service to both humans and machines.

Keywords: Terminology; Phenotype; Distributed Systems; Rare Diseases.

Introduction

The Human Phenotype Ontology (HPO) is a standardized, controlled vocabulary, which allows phenotypic information to be described in an unambiguous fashion in medical publications and databases [1].

The more recent developments of terminology and ontology (T/O) browsers or portals (e.g. UMLS browser [2] and BioPortal [3]) led our team to develop the Health multi-Terminology and Ontology Portal (HeTOP) [4] (http://www.hetop.eu/). It integrates both UMLS T/O and non-UMLS T/O.

The objective of this work is to describe the integration of the HPO into the HeTOP and how such integration can contribute to enhance the quality of the information.

Methods

Natural Language Processing (NLP) techniques were used in this work to propose candidate terms for the translation/mapping of HPO terms. Since HPO is an ontology and has been defined very cleanly, an OBO (Open Biological and Biomedical Ontologies) parser has been specified for the HPO OBO file, downloaded directly from the HPO web site to integrate it into the database.

Results

From the 10,206 HPO concepts, for 18% of them at least one potential candidate term was proposed from the different T/O included in the HeTOP. Each of them was manually reviewed and other terms have been manually translated. Currently, 7,801 HPO concepts are translated into French (76.4%), including 2,984 synonyms. A total of 2,983 semantic mappings were also manually reviewed. Those results are available online, directly from HeTOP: http://pts.chu-rouen.fr/connexion.html?lang=en (login=hpodemo, password=demo11).

Discussion/Conclusion

Since HeTOP is originally based on T/O developed in French, it contains much less knowledge resources than ULMS (n=150) and BioPortal (n=303) but it contains one major functionality that does not exists in other portals: it is crosslingual, allowing navigation among 23 languages. Another good point is that the web site has been designed for humans; the quality of data matters so as its representation. The HeTOP is currently used by 500 unique machines per day, mainly by librarians, translators, students and physicians. The translations of terms and the interoperability between T/O such HPO are also a major leverage of the quality of the data and terminologists and ontologists could find in HeTOP a great opportunity to deal with the lexicons quality.

References

[1] Robinson PN, Mundlos S. The Human Phenotype Ontology. Clin Genet 2010;77:525–534.

[2] Bodenreider O. The Unified Medical Language System (UMLS): Integrating biomedical terminology. Nucleic Acids Res 2004;32:267–270.

[3] Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, Jonquet C, Rubin DL, Storey MA, Chute CG, Musen MA. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Research 2009; Jul;37(Web Server issue):W170-3.

[4] Grosjean J, Merabti T, Dahamna B, Kergourlay I, Thirion B, Soualmia LF & Darmoni SJ. Health Multi-Terminology Portal: a semantics added-value for patient safety. Stud Health Technol Inform. 2011;166: 129-138.