Semantic Integration and Exploitation of Orthology Information and Genetic Disorders

J. A. Miñarro-Giménez[1], Marisa Madrid[2], J. T. Fernández-Breis[1]

[1]Departamento de Informática y Sistemas, Universidad de Murcia, Spain
[2]Cell Division Group, Paterson Institute for Cancer Research, CR-UK, University of Manchester, UK
[1]{jose.minyarro,jfernand}@um.es
[2]Mmadrid@picr.man.ac.uk

Abstract. Translational bioinformatics includes research on the development of novel techniques for the integration of biological and clinical data and the evolution of clinical informatics methodology to encompass biological observations. In this way, the integration of information about gene-related diseases with information about gene orthology would be very helpful for clinical investigations.

Keywords: Translational bioinformatics, Semantic integration, Genetic disorders, Orthology information, Semantic Web Technology.

1 Methodology and Results

In this work, we address the semantic integration of genetic disorders information provided by the Online Mendelian Inheritance in Man (OMIM) [1] and the OGO ontological repository of orthology-related information and knowledge [2], which was recently developed by our research group.

First, the global domain ontology was obtained through the reuse of the OGO ontology and its subsequent extension with the concepts, relations, properties and restrictions of the domain of genetic diseases. The inclusion of domain knowledge of genetic diseases with domain knowledge of orthologous genes and proteins allows the relations between individuals to be analyzed for translational bioinformatics.

The genetic diseases information was integrated into the OGO repository taking the relationships and restrictions of the domain into account in order to define the mapping rules. Thus, the information is properly categorized in the ontology, allowing only the instantiation of the relationships defined in the ontology and controlling the integration process by means of checking the defined restrictions and properties of the domain.

A semantic query interface is provided in order to exploit the integrated repository. This interface provides two query methods. One method allows to query the repository by means of gene identification and to define conditions in the query like the organism and repository resource to look into. The retrieved results contain the orthologous genes and its associated genetic diseases links.

Following these genetic diseases links more information about genetic diseases can be retrieved.

The other method consists in searching by genetic diseases names; the result contains the properties information defined in the ontology and the references to related genetic diseases, the PubMed articles with information about them and the genes references that are involved in the genetic diseases and provides links to its orthologous genes information.

2 Conclusions

As a result of this work the addition of genetic diseases knowledge domain in the OGO ontology was obtained. Since the ontology contains the formal conceptualization of the orthology and genetic disease domain it is possible to share the knowledge model and reuse it in other systems. Due to properties and restrictions, the domain ontology allows us to perform more complicated reasoning than other formal knowledge representation. This integration associates semantically the genes that are involved in a human genetic disorder with those genes that are orthologous to from other organisms, what can facilitate the labour of researchers.

Using a global ontology that cover the domain of the information resources facilitates the integration process and reduces information heterogeneity. The semantic integration is performed by defining mapping rules that take into account the domain knowledge and facilitating the evaluation of the consistency of the integrated repository.

Acknowledgments. Jose Antonio Miñarro is supported by the Fundación Séneca through the fellowship 07836/BPS/07. Marisa Madrid is supported by the EMBO Long-Term Fellowship ALTF 212-2008. This work has been possible thanks to the Spanish Ministry for Science and Innovation through the project TSI2007-66575-C02-02.

References

1. OMIM: Online mendelian inheritance in man, mckusick-nathans institute of genetic medicine, johns hopkins university (baltimore, md) and national center for biotechnology information, national library of medicine (bethesda, md). http://www.ncbi.nlm.nih.gov/omim/
2. Miñarro-Gimenez, J.A., Madrid, M., Fernández-Breis, J.T.: Ogo: an ontological approach for integrating knowledge about orthology. BMC Bioinformatics **10:S10:S13** (2009)