Semantic Web for Translational Biomedicine: Two Pilot Experiments

Olivier Bodenreider, MD, PhD [1], Satya S. Sahoo, MS[2]
[1]National Library of Medicine, NIH, Bethesda, MD; [2]Kno.e.sis Center, Department of Computer Science and Engineering, Wright State University, Dayton, OH

Abstract

Semantic Web technologies (i.e., OWL, RDF and reasoning services) play an increasingly important role in biomedical data integration and represent an enabling technology for translational medicine. We conducted two pilot experiments to illustrate the use of these technologies for integrating genotype and phenotype information in glycomics, and gene and pathway information in the domain of nicotine dependence.

Introduction

The promise of translational bioinformatics and, more generally, translational medicine, hinges upon bridging basic research and clinical practice. One key element to the integration of the research and clinical communities is the integration of the information sources and data used in these communities. In practice, bridges need to be created both across domains (e.g., between genotypic and phenotypic information sources) and across knowledge bases within a domain (e.g., between genomic and pathway resources).

Semantic Web technologies have been used increasingly for data integration if the life sciences [1] and provide a useful framework for translational medicine [2]. Commonly used Semantic Web technologies include languages, namely the Web Ontology Language (OWL) and data models, such as the Resource Description Framework (RDF). Additionally, the use of a representation formalism based on a formal language enables software applications to reason over information.

We present two pilot experiments of the use of Semantic Web technologies for translational biomedicine recently conducted at the National Library of Medicine in the domains of glycomics (genotype-phenotype) and nicotine dependence (gene-pathway).

From glycosyltransferase to congenital muscular dystrophy

Is there a link between the molecular function *glycosyltransferase* and the disease *congenital muscular dystrophy*? This research question from glycomics researchers cannot be answered simply by querying resources such as *Entrez Gene*. After converting *Entrez Gene* to RDF and integrating it with the *Gene Ontology*, we added rules to enable subsumption reasoning. Using the query language SPARQL, we queried the resulting graph and identified a link between *acetylglucosaminyltransferase*, a type of glycosyltransferase, and *congenital muscular dystrophy type 1D* through the gene *LARGE* [3].

Genes and pathways for nicotine dependence

A study funded by the National Institute for Drug Abuse (NIDA) recently identified list of 449 genes putatively involved with nicotine dependence. In order to identify relations among these genes, we created an OWL ontology to represent the information model of *Entrez Gene* and *HomoloGene*, complementary to the BioPAX ontology representing pathway information from *KEGG*, *BioCyc* and *Reactome*. Then, we integrated genomic and pathway information in a single RDF store under these ontologies. We showed that the integrated resource supports complex biological queries including the identification of hub genes, common pathways across species and genes expressed in brain tissue.

Conclusion

The two pilot experiments presented here illustrate the usefulness of Semantic Web technologies as a framework for translational biomedicine.

References

[1] Cheung K, Chen H, Wang Y, Luciano J, Stephens S, Kashyap V, editors. Proceedings of the WWW2007 Workshop on Health Care and Life Sciences Data Integration for the Semantic Web; 2007.

[2] Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, Chen H, et al. Advancing translational research with the Semantic Web. BMC Bioinformatics 2007;8 Suppl 3:S2.

[3] Sahoo S, Zeng K, Bodenreider O, Sheth AP. From "glycosyltransferase" to "congenital muscular dystrophy": Integrating knowledge from NCBI Entrez Gene and the Gene Ontology. Medinfo 2007:1260-1264.