

Semantic Data Integration and Knowledge Representation Using Biomedical Ontologies

Ranga Chandra Gudivada

Departments of Biomedical Engineering¹ and Pediatrics², University of Cincinnati and Division of Biomedical Informatics³, Cincinnati Childrens Hospital Medical Center, Cincinnati OH-45229, USA

Traditional approaches to identify or analyze candidate disease genes are usually done in the laboratory in a laborious process of experimental elimination. However, with the advent of the Human Genome Project and its major contribution to the understanding of genetic level implications to the human health, several new avenues have opened up to apply in silico approaches. The fulcrum of these bioinformatics approaches is an effective integration and analysis of genetic information, gene expression data, gene regulatory networks, protein–protein interactions, gene structure variation, homologs and orthologs data with clinical information. This integration of the massive amounts of accumulating heterogeneous genetic information in the clinical environment is expected to support tailor-made medicine, where clinical diagnosis and treatments will be supported by information at molecular level. The inherent problem for such data integration is lack of widely-accepted standards for expressing the syntax and semantics of the data present in various heterogeneous biomedical databases. The current approach is based on representing biomedical knowledge as a formal, explicit ontological model at various levels of granularity spanning from high-level physiological processes to molecular-level functions. This granular representation will further enhance application for inference and reasoning to provide or capture new testable-hypotheses. The project specifically utilizes semantic web technologies which include Resource Description Framework (RDF) for describing objects and relations, Ontology Web Language (OWL) for specifying ontologies, BioPax for pathway knowledge representation, SPARQL for querying and also other XML-related technologies. As a test case, we have applied this knowledge-aggregation method to enunciate cardiovascular diseases.