Powering Semantic Analysis with Bio-ontologies

Olivier Bodenreider
olivier.bodenreider@nih.gov
National Library of Medicine, National Institutes of Health
Bethesda, MD, USA

ABSTRACT

In the past three decades, bio-ontologies have moved from esoteric artifacts to key resources for the semantic analysis of biomedical text. In this presentation, we will follow the evolution of the creation and integration of bio-ontologies, as well as their role in biomedical applications, including literature analysis.

Bio-ontologies include a variety of resources that provide a source of names for biomedical entities and specify relations among these entities. These resources are generally developed independently by individuals, collectives, institutions, and standard development organizations. Examples of bio-ontologies include the Medical Subject Headings (MeSH), developed by the National Library of Medicine to support the indexing and retrieval of the biomedical literature, the Gene Ontology (GO), developed by the GO Consortium to support consistent annotation and analysis of gene products across organisms, and SNOMED CT, developed by SNOMED International to support clinical documentation and analytics worldwide. While these three examples illustrate resources with a large scope, many other bio-ontologies focus on a specialized subdomain of medicine. Bio-ontologies use different formalisms and various degrees of formality for their representation.

Bio-ontologies play an important role in the semantic analysis of biomedical datasets, including the biomedical literature. Bio-ontologies provide a source of vocabulary for biomedical entities, used for named entity recognition (i.e., finding mentions of biomedical entities in text) and entity resolution (i.e., mapping mentions a specific reference). Bio-ontologies also provide semantic categorization for biomedical entities, which is leveraged for word-sense disambiguation and co-reference resolution (especially when a specific entity is referred to with a broader category). Finally, bio-ontologies provide a source of relations among entities, which can form the basis for relation extraction, hypothesis generation and, more generally, literature-based discovery.

Since most bio-ontologies are developed independently, but often need to be used together, ontology alignment techniques have been developed to identify correspondences among entities across ontologies. Repositories of bio-ontologies, such as the Unified Medical Language System (UMLS) Metathesaurus, the National Center for Biomedical Ontology (NCBO) BioPortal and the Open Biological and Biomedical Ontology (OBO) Foundry are useful sources of ontologies and contribute to the development of tools to support semantic analysis of biomedical text.

ACKNOWLEDGMENTS