LifeSKIM: Application for large scale biomedical semantic annotations

Vassil Momtchev[1], Georgi Georgiev[1], Deyan Peychev[1]

[1] Ontotext AD, Sofia, Bulgaria
{vassil.momtchev, georgi.georgiev, deyan.peychev,
}@ontotext.com
http://www.ontotext.com/lifeskim/

Abstract

Data integration and interpretation is a very challenging problem for the fundamental biomedical research and drug development industry. There is an emerging need for the development of new tools and applications which to transparently link the unstructured and semi-structured information to the biological database. LifeSKIM 0.1 is flexible and scalable application for semantic annotations generation, indexing and retrieval of biomedical entities and documents. Life Ontology is OWL-Horst knowledge base which contains information about genes, gene ontology terms, organisms and diseases. It consists of 100 million RDF statements generated from public biomedical databases Entrez-Gene, GeneOntology, SNOMED, NCBI Taxonomy and DrugBank. The application demonstrates a corpus of over 1.2 million documents annotated against Life Ontology resulting in total 10,884,032 semantic annotations to 1,204,063 different entities. LifeSKIM demonstrates high scalability and the powerful query capabilities for 1) querying and navigation of knowledge generated from structured (biological databases) and unstructured (biomedical document) sources; 2) semantic indexing and retrieval of documents with respect to an ontology; 3) ontology population and learning of new types of entities from text e.g., gene, cell line and type, DNA and RNA molecules; 4) Efficient reasoning against the extracted and structured information, e.g., "type I programmed cell death" is a specific process of "Apoptosis of neutrophils"; 5) co-occurrence tracking and ranking of entities. The application provides multiple user and programming interfaces for access and it is fully compliant with the W3C standards and recommendations.