

## ONTOLOGY-BASED ASSISTED CURATION OF BIOMEDICAL DATA

**Plake C.<sup>1</sup>, Doms A.<sup>2</sup>, Dietze H.<sup>1</sup>, Wächter T.<sup>1</sup>, Alvers M. R.<sup>2</sup> and Schroeder M.<sup>1</sup>**

<sup>1</sup>Technische Universität Dresden Biotechnologisches Zentrum, Tatzberg 47-51, 01062 Dresden, Germany; <sup>2</sup>Transinsight GmbH, Tatzberg 47-51, 01307 Dresden, Germany

Manual curation of biomedical data is highly accurate but time consuming, and does not scale with the ever increasing growth of biomedical literature. Text mining as a high-throughput computational technique scales well but requires human expertise to produce highly accurate results. Ontologies can help organizing large quantities of unstructured information. Here we present three systems, namely GoGene, GoPubMed and GoWeb, employing biomedical ontologies and show how they can assist manual curation of biomedical data.

GoGene associates all genes from different model organisms to concepts of the Gene Ontology (GO) and the Medical Subject Headings (MeSH). The hierarchical structures of both terminologies support clustering and summarizing long lists of genes. Through the integration of known gene annotations from UniProt and EntrezGene with text-mined annotations from all abstracts in PubMed, GoGene currently contains up to 4,000,000 associations between genes and concepts from GO and MeSH for ten model organisms. The quality of all associations can be verified by following the links to their origin, that is, literature or database entries.

GoPubMed aims at reducing the limitations of classical keyword search. It handles inconsistent vocabulary such as synonyms and specialized terminology. It shows the most relevant concepts in GO and MeSH for a search and thus reveals information which otherwise remains buried in the masses of text. This feature as well as the entire bibliography of all authors in PubMed facilitate comprehensive literature search.

GoWeb translates these ideas to the World Wide Web and is thus not only limited to PubMed abstracts. GoWeb uses a standard web-search service and organizes search results based on GO, MeSH, and other concepts such as companies and institutions.