BioMeXT Biomedical Information Extraction

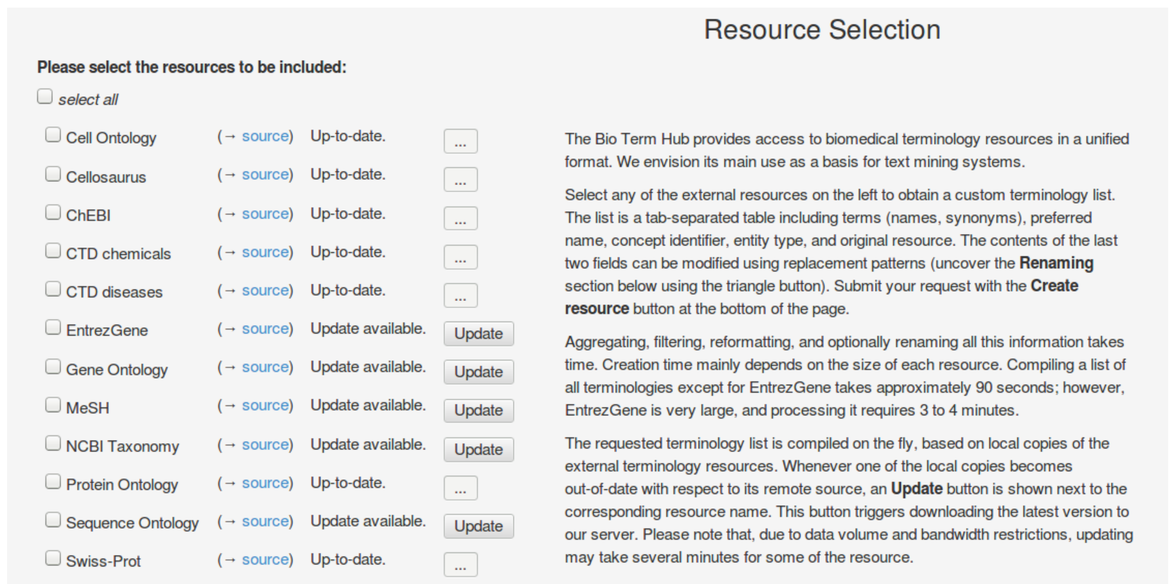
Fabio Rinaldi fabio@ontogene.org

Swiss Institute of Bioinformatics Institute of Computational Linguistics, University of Zurich, Switzerland
Fondazione Bruno Kessler, Trento, Italy Centro de Ciencias Genómicas, UNAM, Mexico

CTI – Start-up and Entrepreneurship, R&D Funding, KTT Support

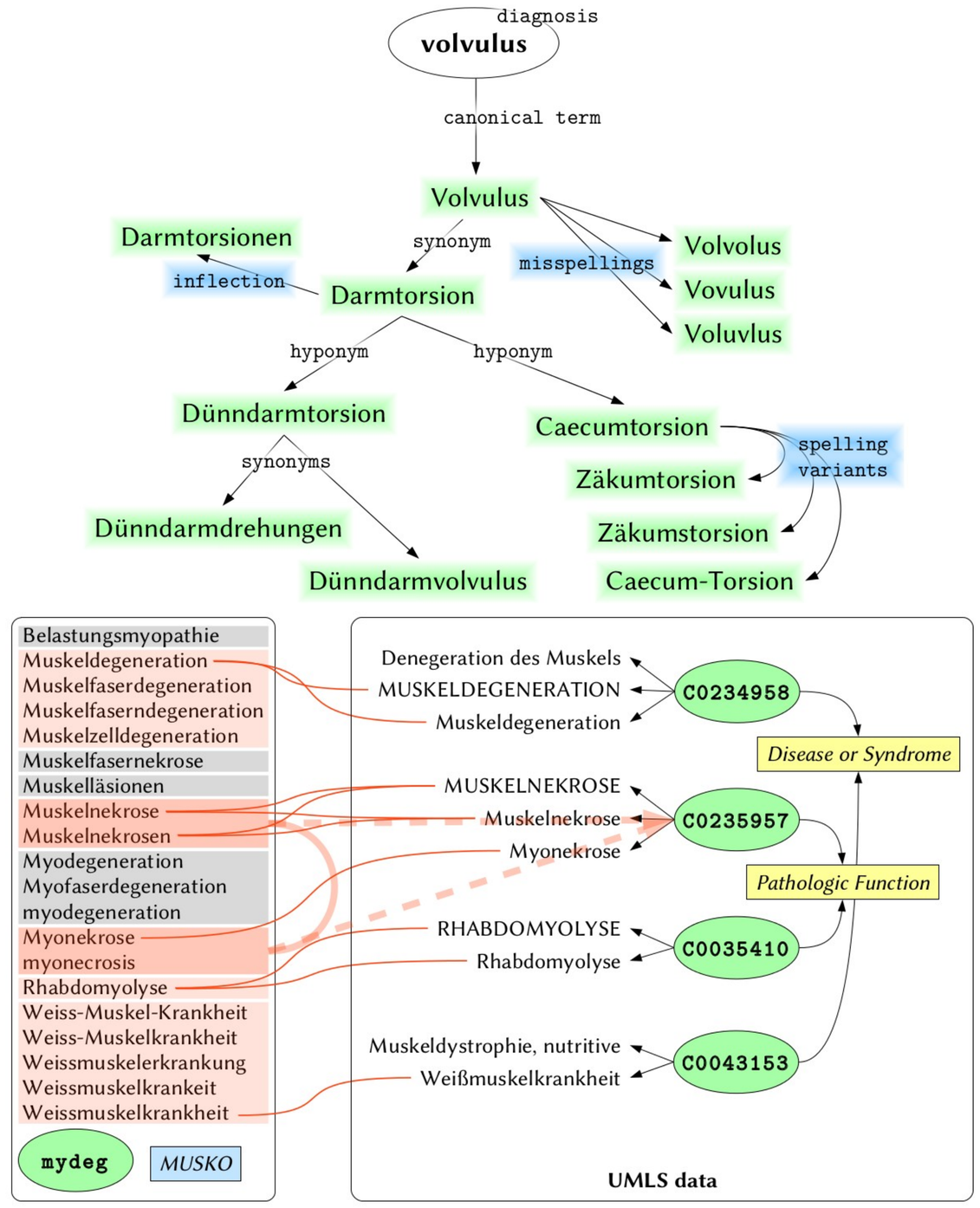
The **BioMeXT** group specializes in Information Extraction (IE), from the biomedical literature, as well as from other textual sources. Information extraction consists in the task of automatically extracting structured information from textual documents, and it is an important component of Text Mining systems. We focus in particular on the extraction of domain-specific entities (such as genes, proteins, drugs, diseases), and their semantic relations (e.g. protein-protein interactions, gene-disease associations). Our tools are often evaluated through participation in community-run evaluation challenges (e.g. **BioCreAtIvE**: Critical Assessment of Information Extraction systems in Biology). Additionally, we provide an environment for Assisted Curation (ODIN), which is currently being used in the curation pipeline of the RegulonDB database in a project funded by the US National Institutes of Health.

Bio Term Hub

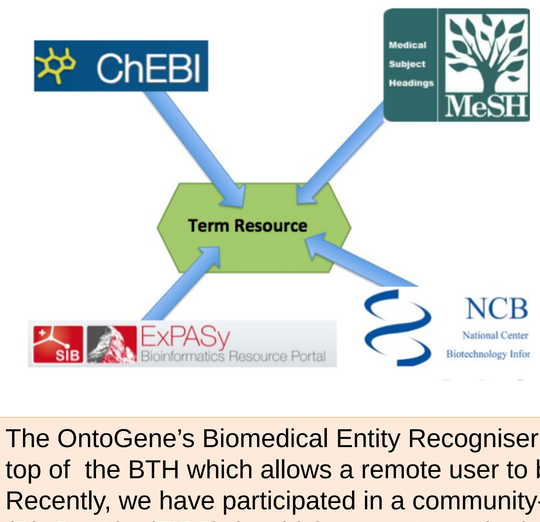


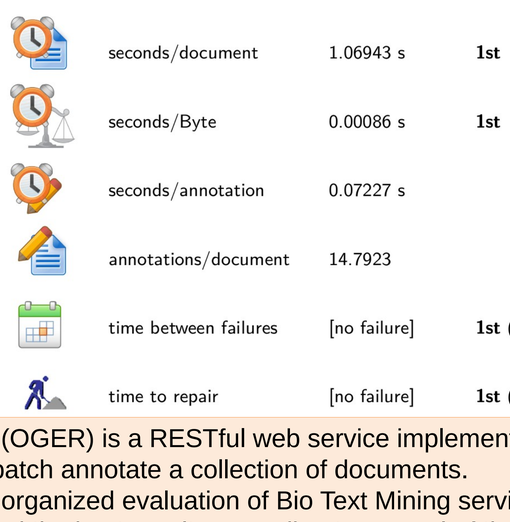
The Bio Term Hub (BTH) is a combined terminological resource created by dynamically sourcing entity names and their identifiers from reference databases. A web interface allows a user to access selected resources and download them in an uniform format.
http://www.ontogene.org/resources/termdb

Terminologies and Ontologies



OntoGene Entity Recognition (OGER)



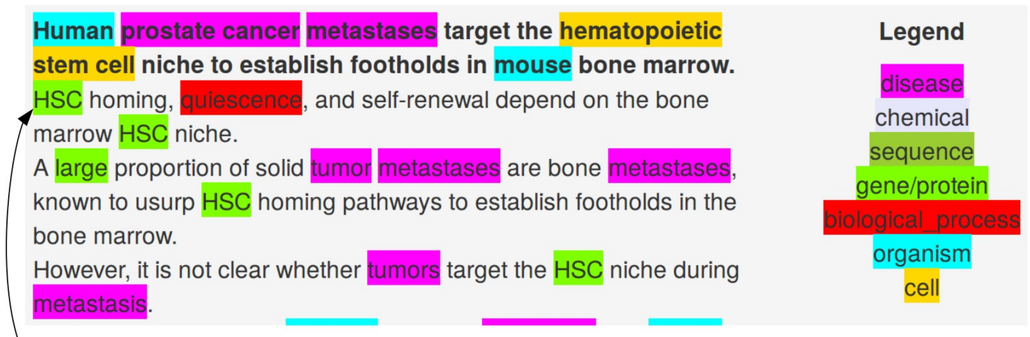


The OntoGene's Biomedical Entity Recogniser (OGER) is a RESTful web service implemented on top of the BTH which allows a remote user to batch annotate a collection of documents.
Recently, we have participated in a community-organized evaluation of Bio Text Mining services (BioCreative/TIPS), in which our system obtained the best results according to several of the evaluation metrics.
http://www.ontogene.org/resources/oger

Methotrexate enhances the anti-inflammatory effect of CF101 via up-regulation of the A3 adenosine receptor expression .

Abstract Methotrexate (MTX) exerts an anti-inflammatory effect via its metabolite adenosine , which activates adenosine receptors . The A3 adenosine receptor (A3AR) was found to be highly expressed in inflammatory tissues and peripheral blood mononuclear cells (PBMCs) of rats with adjuvant-induced arthritis (AIA) . CF101 (IB - MECA) , an A3AR agonist , was previously found to inhibit the clinical and pathological manifestations of AIA . The aim of the present study was to examine the effect of MTX on A3AR expression level and the efficacy of combined treatment with CF101 and MTX in AIA rats . AIA rats were treated with MTX , CF101 , or both agents combined . A3AR mRNA , protein expression and exhibition were tested in paw and PBMC extracts from AIA rats utilizing immunohistochemistry staining , RT - PCR and Western blot analysis . A3AR level was tested in PBMC extracts from patients chronically treated with MTX and healthy individuals . The effect of CF101 , MTX and combined treatment on A3AR expression level was also tested in PHA - stimulated PBMCs from healthy individuals and from MTX - treated patients with rheumatoid arthritis (RA) . Combined treatment with CF101 and MTX resulted in an additive anti-inflammatory effect in AIA rats . MTX induced A2AAR and A3AR over-expression in paw cells from treated animals . Moreover , increased A3AR expression level was detected in PBMCs from MTX - treated RA patients compared with cells from healthy individuals . MTX also increased the protein expression level of PHA - stimulated PBMCs from healthy individuals . The increase in A3AR level was counteracted in vitro by

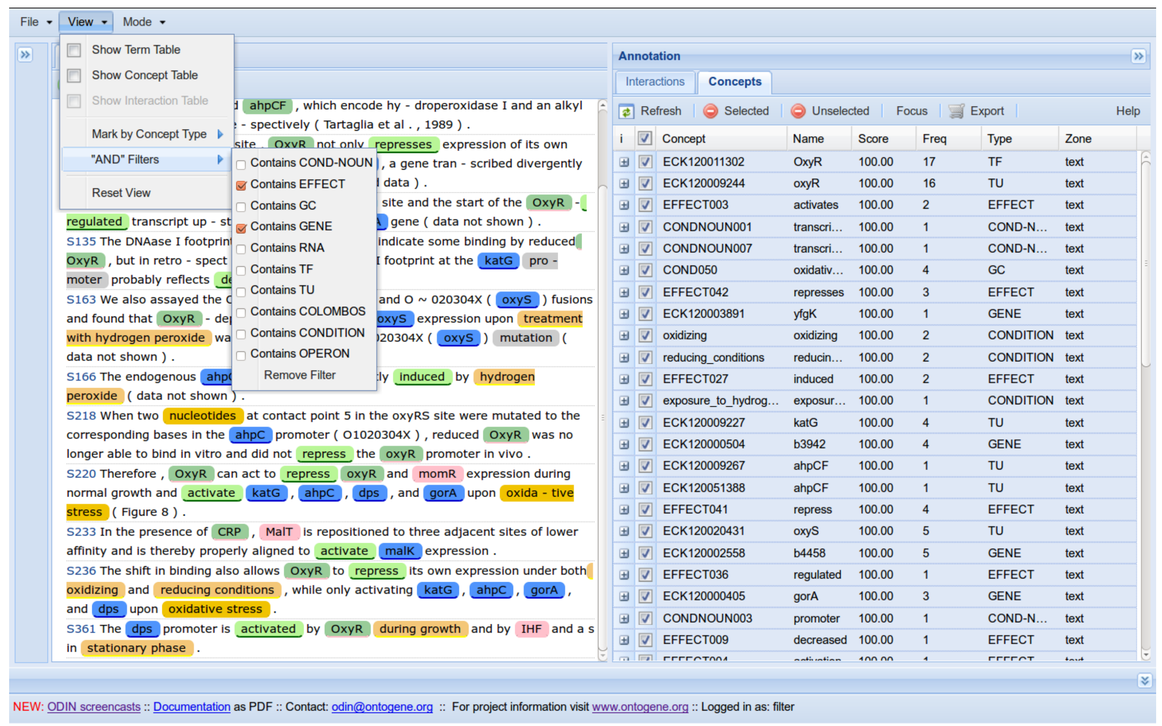
Disambiguation challenges



gene/protein	galactoside 2-alpha-L-fucosyltransferase 1	PR:000007702	Protein Ontology
cellular component	Hedgehog signaling complex	GO:0035301	Gene Ontology
gene/protein	chaperone protein HscA	PR:000022925	Protein Ontology
chemical	N-(3-carboxypropanoyl)-N-hydroxycadaverine	CHEBI:50443	ChEBI

The biggest problem in dealing with biomedical text entities is the large degree of ambiguity. The same name can refer to several different entities, and besides some common language names can be used occasionally as domain terms. Therefore disambiguation is a major challenge. In some types of text (e.g. clinical reports), other problems might emerge, such as misspellings and the usage of variants. The usage of reference Ontologies can help to deal with this problem.

Assisted Curation



In an NIH-funded project in collaboration with the RegulonDB group (CCG, UNAM), we are developing an interface for assisted curation (ODIN) which offers efficient manipulation of text and annotations. This system can use the predefined entities provided by the Bio Term Hub, or allow for customized vocabularies. The introduction of text mining technologies in an assisted curation pipeline is expected to increase the throughput of the curation process without compromising the quality of the results.

Funding

The OntoGene/BioMeXT group is partially supported by the Swiss National Science Foundation (SNF): grants 100014-118396/1, 105315-130558/1, CR30I1_162758 (F. Rinaldi). We are also involved in an SNF-NRP74 project, and in a large project funded by the US-NIH. Additionally we received funding from the COGITO foundation and the Federal Food Safety and Veterinary Office (FSVO). Recently a new project proposal has been approved by the Commission for Technology and Innovation (CTI).