

POSTER 142 GenMapper – A System for Flexible Integration of Molecular-biological**Do H, Rahm E****List of topics**

Immunology and Rheumatology
 Endocrinology
 Neurosciences
 Oncology
 Bioanalytics
 Bioorganics
 Biotechnology
 Receptors and signal transduction
 Biochemical pharmacology and toxicology
 Molecular biology and biochemistry
 Molecular cell biology and -medicine
 Evolution
 Biophysics
Bioinformatics
 Stem cell biology
 Microbiology and virology
 Molecular Vascular Biology and Metabolism
 Clinical Science

Molecular-biological annotation data is continuously being collected, curated and made accessible in numerous public data sources. Many applications require annotation data to be integrated from different sources. This integration task is a major challenge in bioinformatics due to the high degree of heterogeneity of the annotation data and the constantly evolving structure and contents of the sources. We present GenMapper, a system, which is able to perform integration of heterogeneous annotation data in a flexible way. In particular, we use a generic data model to uniformly represent different kinds of annotations originating from different data sources. The generic representation makes it easy to add new sources or to incorporate changes for existing ones. From the generic representation, tailored annotation views can be flexibly generated to serve specific analysis needs. An example query for such a view is "Given a set of UniGene clusters, retrieve all associated Enzymes and GeneOntology functions". By combining existing annotation relationships, GenMapper can automatically derive new annotations if they are not available in a public source. It also allows the user to specify his own way to derive the required annotations. For annotations based on hierarchical taxonomies, such as GeneOntology and Enzyme, it is possible to exploit the source structure to perform significance analysis and functional comparison of different sets of objects. Currently, GenMapper integrates annotation data for 65 different data sources, a vendor-based datasource of annotations for genes used in microarray experiments.

→ **Hong-Hai Do**
Interdisciplinary Centre for Bioinformatics (IZBI),
 Kreuzstr. 7b , 04103 Leipzig
 email: rahm@informatik.uni-leipzig.de

POSTER 143 Databases and Data Integration**Kirsten T, Rahm E, Do H****List of topics**

Immunology and Rheumatology
 Endocrinology
 Neurosciences
 Oncology
 Bioanalytics
 Bioorganics
 Biotechnology
 Receptors and signal transduction
 Biochemical pharmacology and toxicology
 Molecular biology and biochemistry
 Molecular cell biology and -medicine
 Evolution
 Biophysics
Bioinformatics
 Stem cell biology
 Microbiology and virology
 Molecular Vascular Biology and Metabolism
 Clinical Science

Flexible analysis of molecular-biological data is at the heart of many bioinformatics research projects. The general objective of WG1 is to support the data management and analysis needs of molecular-biological research and service projects in all working groups especially those requiring access to large amounts of data and the integration of diverse data and metadata (annotations). Moreover, we aim at advancing the state of the art in matching and integrating annotations and coupling data mining and large databases.

A current focus is database support for microarray-based gene expression analysis which is a major request from our experimental and medical partners. It will become of high priority to manage the large amounts of data on gene expression but also on other molecular data (e.g. sequence data, cytogenetic data, proteom data, cell-phenotype data, pathway data) obtained from heterogenous sources like cell biological experiments or clinical trials from different laboratories in a comprehensive database.

Annotations from several sources (public, expert knowledge) also need to be integrated to help the users in interpreting detected gene expression patterns. For gene expression and related molecular data, these tasks can be best served by a data warehouse approach. The desired functionality is neither available commercially nor in prototypes. We therefore decided to design a data warehouse and a conceptual data model. All data access and analysis routines are accessible through a uniform web-based user interface.

The data warehouse is completed in a first version and presently running on a test system.

→ **Toralf Kirsten**
IZBI

Kreuzstr. 7b, 04103 Leipzig
 phone: (0341) 14951 16
 email: kirsten@izbi.uni-leipzig.de