

Abstract: Integration of molecular-biological Gene Annotation

Toralf Kirsten, Christine Körner, Hong-Hai Do, Erhard Rahm

We present a new approach [1, 2] to integrate annotation data from public sources for the expression analysis of genes and proteins. Expression data is materialized in a data warehouse supporting high performance for data-intensive analysis tasks. On the other hand, annotation data is integrated virtually according to analysis needs. Our virtual integration utilizes the commercial product SRS (Sequence Retrieval System) of LION bioscience. To couple the data warehouse and SRS, we implemented a query mediator exploiting correspondences between molecular-biological objects explicitly captured from public data sources. This hybrid integration approach has been implemented for a large gene expression warehouse and supports functional analysis using annotation data from GeneOntology, Locuslink and Ensembl. The presentation motivates the chosen approach, details the integration concept and implementation, and provides results of preliminary performance tests.

- [1] Körner, C., Kirsten, T., Do, H.-H., Rahm, E.: Hybride Integration von molekularbiologischen Annotationsdaten. Proc. BTW Conference, Karlsruhe 2005.
- [2] Kirsten, T., Do, H.-H., Körner, C., Rahm, E.: Hybrid Integration of molecular-biological annotation data. 2nd DILS Workshop, San Diego, 2005.