PiPa: Custom Integration of Protein Interactions and Pathways

Sebastian Arzt, Johannes Starlinger, Oliver Arnold, Stefan Kröger, Samira Jaeger, Ulf Leser

Knowledge Management in Bioinformatics
Humboldt-Universität zu Berlin
Unter der Linden 6, 10099 Berlin, Germany
{arzt, starling, arnold, kroeger, sjaege, leser}@informatik.hu-berlin.de

Abstract: Information about proteins and their relationships to each other are a common source of input for many areas of Systems Biology, such as protein function prediction, relevance-ranking of disease genes and simulation of biological networks. While there are numerous databases that focus on collecting such data from, for instance, literature curation, expert knowledge, or experimental studies, their individual coverage is often low, making the building of an integrated protein-protein interaction database a pressing need. Accordingly, a number of such systems have emerged. But in most cases their content is only accessible over the web on a per-protein basis, which renders them useless for automatic analysis of sets of proteins. Even if the databases are available for download, often certain data sources are missing (e.g. because redistribution is forbidden by license), and update intervals are sporadic.

We present PiPa, a system for the integration of protein-protein interactions (PPI) and pathway data. PiPa is a stand-alone tool for loading and updating a large number of common PPI and pathway databases into a homogeneously structured relational database. PiPa features a graphical administration tool for monitoring its state, triggering updates, and for computing statistics on the content. Due to its modular architecture, addition of new data sources is easy. The software is freely available from the authors.

Keywords: Data Integration, Protein Database, Bioinformatics