BCB-Seminar

Time: Wednesday, April 21, 16.00 – 18.30
Place: Humboldt Universität zu Berlin, Institut für Informatik
(Humboldt-Kabinett)
Rudower Chaussee 25, 12489 Berlin

Prof. Dr. Ralf Zimmer, Ludwig-Maximilians-Universität München,
Methods for the analysis of gene expression data and biochemical networks

Coffee break

Prof. Dr. Ulf Leser, Humboldt Universität zu Berlin,
Mining Biomedical Literature

Dr. Stephan Heymann, Peter Rieger, Humboldt Universität zu Berlin,
***Ad hoc* Interfaces for Querying Genomic Data -Exemplified by Alternative Splice Form Evaluation**

Guests are welcome!

For further information, please visit http://www.bcbio.de

BCB-Seminar

(Abstracts of talks I)

Methods for the analysis of gene expression data and biochemical networks

Ralf Zimmer, Chair for Practical Computer Science and Bioinformatics, Department of Computer Science, Ludwig-Maximilians-Universität München, Amalienstrasse 17, 80333 München, Ralf.Zimmer@ifi.lmu.de

Understanding the molecular interactions and the cellular processes of genes and proteins is a major challenge for bioinformatics in the 'post genome era'. A major application problem is the identification and validation of molecular targets for drug research in pharmaceutical industry. We present methods for the analysis of large scale expression data (i.e. microarray gene expression data) in the context of biochemical (i.e. metabolic and regulatory networks).

We represent molecular network data as formal models called Petri nets, which are well-suited for the task and have been studied for decades such that an extensive mathematical theory is available for their analysis. The Petri net representation of relevant molecular networks are obtained via information extraction from relevant metabolic and regulatory databases, via input from biological experts from the respective fields, and via text mining of scientific literature. Text mining methods have been fine-tuned for the generation of molecular networks based on genes and proteins in particular for certain human diseases. Based on such (possibly very large networks) appropriate algorithms have been developed to compare and statistically analyse expression data, in particular in the context of networks and pathways. An overall goal of these methods is to identify pathways involved in a certain disease state as measured by the expression values of genes and proteins in that state as compared to the normal state.

We have developed ProMiner, a text mining program, which allows to process very large texts (i.e. all Pubmed abstracts) with acceptable sensitivity and specificity (top rank at the recent text mining competition BioCreative) and ToPNet a pure Java software system, which can be used for interactive analysis and visualization of expression and network data in an integrated fashion. ToPNet also contains a set of algorithms for the combined analysis of data in the context of networks (i.e. pathway scoring, pathway search, significant area search, co-clustering of networks and data, and pathway queries). The ToPNet tool, including documentation and tutorials, is freely available for academic use via http://www.biosolveit.de.

The methods and the tools have been developed in collaboration with Aventis pharma Frankfurt and have been applied to several Aventis disease-related research projects. In particular, we have analysed data from osteoarthritis research with the goal to identify and model relevant disease pathways both for target identification and target validation.

BCB-Seminar

(Abstracts of talks II)

Mining Biomedical Literature

Ulf Leser, Knowledge Management in Bioinformatics, Humboldt-Universität zu Berlin

Despite the many databases in the Life Sciences, most knowledge is still published in scientific papers, i.e., in the form of essentially unstructured, natural language text. Researcher face approximately 400.000 new publications in over 5.000 journals each year. At the same time, many results in modern biotechnology are based on high-throughput methods that generate little information about very many objects. For instance, microarray-based gene expression profiling can measure the expression intensity of more than 20.000 genes on a single chip; but for robust results, these measurements need to be combined with additional information about the single genes. Text mining is concerned with finding and extracting such information in scientific articles.

The talk will give an introduction into text mining in life sciences. The main application areas are presented. Special focus is put on two recent projects. First, we present a system for document classification which supports researches in Systems Biology. Second, we describe the specific problems of named entity-recognition in molecular biology and present our approach to this problem. The first project is joint work with the group of Edda Klipp, Max-Planck-Institute for Molecular Genetics. The second project is joint work with the group of Tobias Scheffer, Humboldt-Universität Berlin.

Ad hoc Interfaces for Querying Genomic Data - Exemplified by Alternative Splice Form Evaluation

Dr. Stephan Heymann, Peter Rieger, DBIS, Institut für Informatik, Humboldt-Universität zu Berlin

Over the last years there have been published multiple suggestions and improvements for how to access, retrieve, display, and exploit web-resident gene and protein sequence data as well as related information. However, in many occasions life scientists often experience the permanent need to ask questions different from those browsers and portal interfaces can immediately answer. Information bits are scattered over a number of autonomous, coarsely cross-referenced sources. In practice, it often remains a manual process to “re-grep” satisfying information from various open windows, and set-oriented calculi are even more tedious.

This presentation introduces the “Gene-EYe Genome Data Warehouse” (http://www.dbis.informatik.hu-berlin.de/research/bioinformatics/dbbio), which overcomes at least some of the above perils. Our platform mirrors publicly available data in relational form and eases flexible access by means of the common ACCESS QBE interface - As an example for accessing and manipulating complex genome data, our work addresses the search for alternative splice forms of human genes (http://www.bioinfo.de/isb/2004/04/0017/) which we also describe in our talk.