BCB-Seminar

Wednesday, May 25th, 4 - 7 pm
Humboldt University of Berlin
Institute of Computer Science

Udo Hahn
Lehrstuhl für Computerlinguistik, Universität Jena

"Anforderungen an die Infrastruktur für das Text Mining in den Life Sciences"

Im Vortrag werden die infrastrukturellen Anforderungen skizziert, die sich für hochwertige Text-Mining-Dienste (etwa das entity und relation mining) in den Life Sciences ergeben. Hierzu zählen neben diversen computerlinguistischen Modulen (für die Namenserkennung, das Chunking oder Parsing) vor allem annotierte Korpora und breit abdeckende terminologische Ressourcen (Ontologien). Am Bespiel der (Nicht-)Verfügbarkeit life-science-spzifischer Annotationen werden die Konsequenzen für die Qualität der Analyseergebnisse gezeigt.

C. Lawerenz, **J. Eils**, R. Kabbe, R. Eils
Intelligent Bioinformatic Systems, German Cancer Research Center, Heidelberg

„iCHIP, an integrated database platform for clinical network-wide research"

iCHIP as a database platform for microarrays and comparative genomic hybridizations is established and integrated in different clinical networks. Its development has been funded by the German National Genome Research Network (NGFN). In the context of strengthening quality assurance within NGFN, a national quality control as a German initiative is being set off. Clinical parameter sets as well as standard operating procedures (SOPs) are part of the new efforts of this national quality commitee. We have created specified clinical parameter sets for different networks (e.g., cancer, inflammation/infection, neuro) in collaboration with clinical partners. The definition of other sets specific for cardiovascular diseases and environmental disorders are currently in planning. Clinicians of respective research groups have developed SOPs, which will be integrated into iCHIP and made available throughout the NGFN community.

To assist the researchers and clinicians in accessing and utilizing the specified clinical parameter sets and SOPs, we have integrated download facilities into iCHIP. The revised SOPs and parameter sets will be presented online. The MIAME requirements and the MGED Ontology standards have been used as guidelines for the development of these standards. The functionality of iCHIP will be extended to include 2D gel electrophoresis data from proteomics, matrix-CGH, tissue microarrays and cellular assays in combination with RNAi technology. The incorporation of enhanced quality standards is a necessary and important factor for the realization of clinical data exchange among each iCHIP node and data transfer to central repositories. Furthermore, the usage of comprehensive data within iCHIP facilitate to overcome

the situation of insufficient biomaterial data by combining several studies into one using methods of meta-analysis.

Bertram Weiss
Schering AG, Berlin

PhenomicDB: Comparative Phenomics in Pharmaceutical Research

Our genetic material determines in great part our *Gestalt*, or to use the Greek word for appearance: our phenotype. Any genotypic modification may ultimately become manifest in no, only slight, or rather severe changes in phenotype. Despite the importance of understanding how specific genotype alterations may contribute to the development of certain diseases, surprisingly little efforts have been made towards extracting and exploiting the current knowledge of genotype-phenotype relationships. Large numbers of genes were functionally annotated in the past with the help so-called "forward genetics" studies, i.e. going from a known phenotype to the identification of the respective mutated gene, in model organisms such as *D. melanogaster* or yeast. Analogous studies in higher organisms were hampered by the lack of sophisticated genetic methods even though examinations of transgenic or knockout animals as well as comprehensive SNP analyses have contributed significantly to today's knowledge of gene function.

This situation has now changed dramatically with the advent of RNA interference (RNAi) which allows to specifically silence one gene at a time and to observe the phenotypic outcome. This systematic large-scale characterization of genes in higher mammalian model systems in vitro will increase dramatically our knowledge on gene function during the upcoming years. It is now time to develop strategies on how to sytematically utilize this knowledge for the development of novel therapeutic approaches.

Although the path to massive genotype-phenotype examinations is now paved experimentally, we are still missing key tools for storing phenotypic information others than free text. There is not even a central repository for our available knowledge on genotype-phenotype relationships. The comparision of phenotypes over a range of species (i.e. comparative phenomics), although recognised as a great opportunity, has lacked practical applicability as key tools are still missing. Prerequisites like phenotype vocabulary and ontologies or standardized protocols for phenotypical assays have only started to be addressed since very recently and more effort is needed.

This talk will explain how we have approached some of the resulting bioinformatics opportunities in comparative phenomics. PhenomicDB is a first approach towards this goal.

Guests are welcome!

Rudower Chaussee 25
12489 Berlin
Haus 3, Raum 101