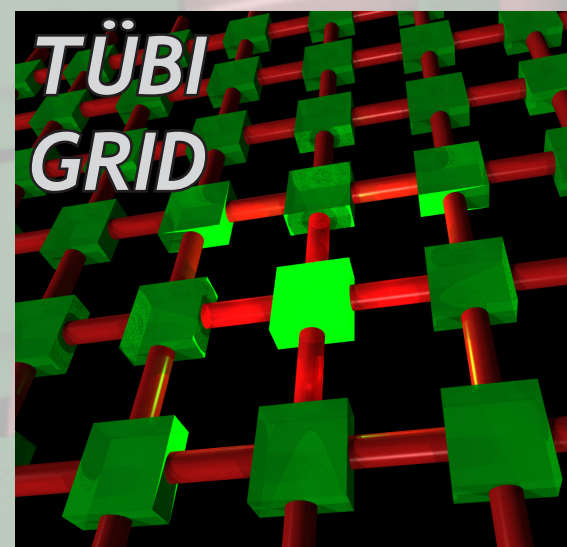


Muriel Quenzer, Sandra Gesing, and Oliver Kohlbacher

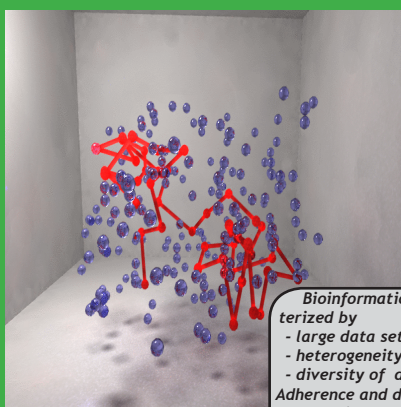
Simulation of Biological Systems, Center for Bioinformatics, Eberhard-Karls-Universität Tübingen, 72076 Tübingen, Germany

The area of bioinformatics covers a broad range of rather different topics. In all these fields an enormous number of complex and sophisticated algorithms and tools has been developed to aid the analysis of large datasets as a basis for further research activities. To enable efficient interaction with these tools and to allow various combinations, we are designing a workflow-enabled portal. This portal provides a single gateway of easy access to different tools as web services to a portlet based interface.

This approach is motivated on one hand by our role in the ***TüBiGrid project*** and on the other hand by our experience with the installation of complex bioinformatics software in the heterogeneous environments of our experimental cooperation partners. The goal of the TueBiGrid project is to build a grid infrastructure for parallel and distributed applications in the area of bioinformatics, particularly in the field of proteomics [2] and systems biology. This poster reports on the planned project and its current status.



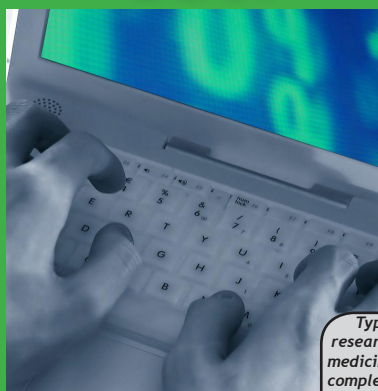
Data



- Bioinformatics data is characterized by
 - large data set size
 - heterogeneity
 - diversity of data formats

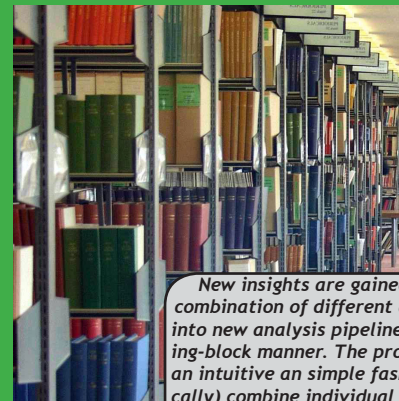
Adherence and development of standardized XML formats helps addressing these issues.

User



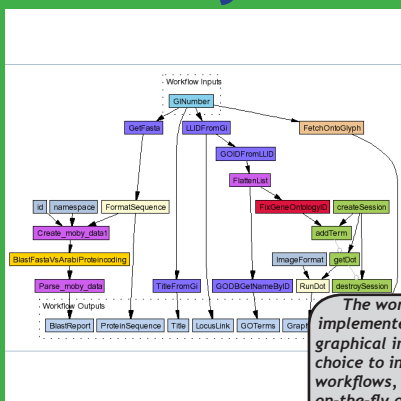
Typical users are researchers from biology, biomedicine, biochemistry. To avoid complex software installation and administration, a web-based user interface is the platform of choice.

Results



New insights are gained from the combination of different analysis tools into new analysis pipelines in a building-block manner. The projects aims at an intuitive an simple fashion to (graphically) combine individual combination of prefabricated tools into new analysis steps and also a user-friendly presentation of the results for further human analysis.

Workflow



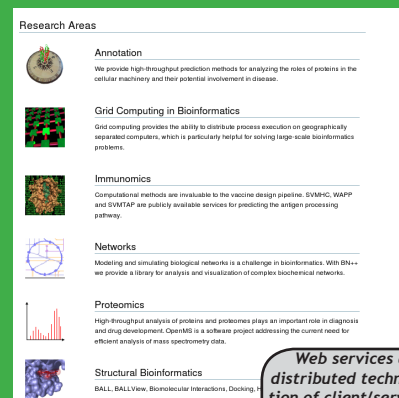
The workflow engine will be implemented as a portlet with a graphical interface. The user has the choice to interact with predefined workflows, to modify workflows on-the-fly or to create workflows. The workflow engine converts data automatically between data in case this is needed.

Portal



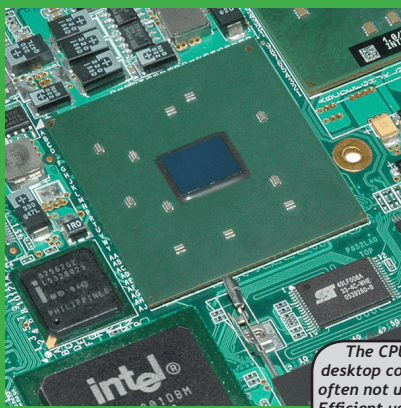
(based on GridSphere[1])

Web Services



Web services are yet another distributed technology for the creation of client/server application. As web services use standard XML languages they are platform-independent and language-independent. Another important advantage of web services using HTTP is that most of the internet's proxies and firewalls do not mess with HTTP traffic.

CPU Power



The CPU power of existing desktop computers and clusters is often not used to the extent possible. Efficient use is largely hindered by making these resources available to non-expert users of different organizations (virtual and real). Abstract WS-based interface can facilitate this.

Storage



The characteristics of the processed data dominates the storage management and leads to the question of efficient, low-priced and secure management of distributed data. Paradigm: Data coming to the computation or computation coming to the data.

Software



Various software tools require specific environments which might be available within a virtual organisation but not at every geographical site of this organisation. Due to the globalisation organisations are built around the world sharing a mutual funding but being spread at various locations.

Status

Currently we are in an early stage of the project and working on miscellaneous aspects. These include the evaluation of existing middleware solutions and tools like Globus Toolkit 4, gLite, Sun Grid Engine and Condor G. Another aspect is the development of first web services and portlets for our TOPP modules[3]. Additionally we are working

on the specification for the lightweight workflow engine based on the standard language BPEL (Business Process Execution Language). To enable fast and reliable integration of these services we project a code generator which facilitates the creation of graphical portlets by automatically parsing existings WSDL files into portlet enabled Java classes.

References

- [1] GridSphere portal framework, <http://www.gridsphere.org>
- [2] OpenMS - A Software Platform for Shotgun Proteomics, <http://open-ms.sourceforge.net>
- [3] TOPP - The OpenMS Proteomics Pipeline, <http://open-ms.sourceforge.net/TOPP/index.php>