Workflow-enabled Portal for Bioinformatics in a Grid Infrastructure

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 TüBiGrid 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.

The characteristics of Bioinformatics data are characterized by
- large data set size
- complexity
- diversity of data formats
- accuracy and development of standardized XML formats helps addressing these issues.

Typical users are researchers from biology, biomedicine, biochemistry. To avoid complex software installations, an intuitive and simple access to (graphical) user interface is the platform of choice.

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

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

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 existing WSDL files into portlet enabled Java classes.

References