A multilayer architecture to support bioinformaticians of today and tomorrow

Bartocci E (1), Cannata N (1), Corradini F (1), Merelli E (1), Milanesi L (2), Romano P (3)

(1) Dipartimento di Matematica e Informatica, Universita' di Camerino, Camerino
(2) Istituto Nazionale di Ricerca sul Cancro, Genova (3) ITB-CNR, Milano

Motivation

In bionformatics fundamental importance are acquiring cyberinfrastructures [1] that will permit multidisciplinary, geographically dispersed, data and computation intensive science. Cyberinfrastructures include peer-to-peer technology, web services and grid technology. In particular grid technology can support virtual communities through sharing of computational and data resource. Simultaneously is growing the request for semantics and the WWW started to become Semantic Web [2]. Nevertheless, scientists difficultly can keep up with the fast development of a specific research area, due to the continuous appearing of new knowledge, data and computational resources. The quest for resources, therefore became a very demanding and time-consuming activity. Bioinformatics deeply changed molecular biology making in-silico experiments a routine task, beside in-vivo and in-vitro ones. In the age of e-Science [3], bioinformaticians can intuitively compose their experiments in the form of workflows. Tasks, designed at a higher conceptual level, are dynamically bound at runtime to physical resources -data and computational ones- taking also into account issues like workload, resource availability and optimization. The integration of all the bio-molecular and "omics" pieces of knowledge requires a significant effort. Built on this premise, systems biology [4] aims at the analysis, modeling and simulation of biological systems and processes, through the supply of mathematical and computational models. Therefore the availability of a virtual desk, on which would be easy to progressively engineer models of biological systems and to simulate and validate them, undoubtedly constitutes another important requirements in modern and future biology.

Methods

To fulfill bioinformaticians needs we propose a multilayer architecture. At the user layer, it is intended to support in-silico experiments, resource discovery and biological systems simulation. The pivot of the architecture is a component called Resourceome [5] which keeps an "alive" index of resources in the bioinformatics domain using a specific ontology of resource information. The Resourceome directly assists scientists in the hard navigation in the ocean of bioinformatics resources. A Workflow Management System, called BioWMS, provides a web-based interface to define in-silico experiments as workflows [6] of complex and primitives activities. In this case high level concepts concerning activities and data could be indexed in the Resourceome. The Resourceome itself would dynamically support workflow enactment, providing the related resources available at runtime. A set of tools for systems biology allows user to intuitively create and refine agent-based models [7] of biological systems and processes. Also in this case Resourcesome can be used to retrieve important related resources like e.g. organism-specific parameters of metabolic pathways. An Agent-based middleware provides the necessary flexibility to support data and computation intensive distributed applications. A middleware permits to develop complex software systems without taking into account at design time who is actually executing them and where they are physically executed. A GRID Infrastructure allows a transparent access to the high performance computing resources required, for example in the biological systems simulation. Beside the computation-intensive aspect, other important issues are taken into account today from grid architectures, like e.g. service grids and knowledge grids.

Results

We conceived the proposed architecture in the context of the MIUR-FIRB LITBIO project(http://www.litbio.org/). The main goals of LITBIO are: to serve the research community with Bioinformatics tools and database and to develop a virtual Laboratory for Interdisciplinary Technologies in Bioinformatics applied to Genomics, Transcriptomics, Proteomics, Systems

Biology and Metabolomics.

Availability: http://www.litbio.org/

Contact email: emanuela.merelli@unicam.it

References

1. T. Hey and A. E. Trefethen. Cyberinfrastructure for e-Science. Science, 308(5723):817821, 2005.
2. T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. Sci Am., 284:3443, 2001.
3. D. De Roure and J. A. Hendler. E-science: The grid and the semantic web. IEEE Intelligent Systems, 19(1):6571, 2004
4. H. Kitano. Foundations of Systems Biology. MIT Press, 2002
5. N. Cannata, E. Merelli, and R. B. Altman. Time to organize the bioinformatics resourceome. PLoS Comput Biol., 1(7):e76, 2005
6. E. Bartocci, F. Corradini, and E. Merelli. Enacting proactive workflow engine in e-Science, Proc. of ICCS 2006, to appear in LNCS
7. N. Cannata, F. Corradini, E. Merelli, A. Omicini, and A. Ricci. An agent-oriented conceptual framework for systems biology. In T. Comp. Sys. Biology, volume 3737 of LNCS, pages 105122, 2005.