Chapter XVI
Bioinformatics Web Portals

Mario Cannataro
Università "Magna Græcia" di Catanzaro, Italy

Pierangelo Veltri
Università "Magna Græcia" di Catanzaro, Italy

INTRODUCTION

Bioinformatics involves the design and development of advanced algorithms and computational platforms to solve problems in biomedicine (Jones & Pevzner, 2004). It also deals with methods for acquiring, storing, retrieving and analysing biological data obtained by querying biological databases or provided by experiments. Bioinformatics applications involve different datasets as well as different software tools and algorithms. Such applications need semantic models for basic software components and need advanced scientific portal services able to aggregate such different components and to hide their details and complexity from the final user. For instance, proteomics applications involve datasets, either produced by experiments or available as public databases, as well as a huge number of different software tools and algorithms. To use such applications it is required to know both biological issues related to data generation and results interpretation and informatics requirements related to data analysis.

Bioinformatics applications require platforms that are computationally out of standard. Applications are indeed (1) naturally distributed, due to the high number of involved datasets; (2) require high computing power, due to the large size of datasets and the complexity of basic computations;

(3) access heterogeneous data both in format and structure; and finally (5) require reliability and security. For instance, applications such as identification of proteins from spectra data (de Hoffmann & Stroobant, 2002), querying of protein databases (Swiss-Prot), predictions of proteins structures (Guerra & Istrail, 2003), and string-based pattern extraction from large biological sequences, are some examples of computationally expensive applications. Moreover, expertise is required in choosing the most appropriate tools. For instance, protein structure prediction depends on proteins family, so choosing the right tool may strongly influence the experimental results.

Recently, there has been much interest from database community and computer science community for bioinformatics. Nevertheless, what is still missing is a high-level environment able to classify tools and provide Web-based easy to use application programming interfaces. In such a way, users can concentrate on the logic of application (i.e., biological aspects) leaving to such platform the work to compose applications, format input data, provide options and parameters, and collect results.

Another important requirement is the accessibility of such platform through a Web portal, that is, by using the user interfaces and protocols of the World Wide Web. A bioinformatics Web portal is thus a Web portal that allows access to bioinformatics tools and databases through a Web browser. Moreover, due to the complexity, diversity and a huge number of bioinformatics tools and databases, a bioinformatics Web portal should also support problem formulation, application composition and execution, results visualisation and annotation. A possible approach to solve these issues —high-level modeling and Web-based user interfaces—can be obtained by adding semantics links between biological problems and bioinformatics resources through ontologies (Baker, 1998), and by decoupling Web-based user interfaces from high-performance back-end platforms.

In this article we review main requirements of distributed bioinformatics applications and related bioinformatics Web portals, and report the proposal of a grid-based bioinformatics portal allowing choosing and composing of bioinformatics tools with the help of a domain ontology describing data and software resources.

BACKGROUND

Bioinformatics researchers, among the other directions, are investigating through: (1) data modeling to manage heterogeneous datasets (e.g., see HUPO, n.d., the HUPO, Human Proteome Organization—Proteomics Standard Initiative); (2) specialised services for protein sequences searching, and data mining techniques to extract meaningful information from datasets; (3) ontologies and metadata for a high-level description of the goals and requirements of applications; and (5) high performance computational platforms to execute distributed bioinformatics applications.

Many applications have been defined to support biological researchers for solving problems on different topics where large computing power is required. Grid community (Foster & Kesselman, 2003) has recognised that bioinformatics and postgenomic applications are both a challenge but especially an opportunity for distributed high performance computing and collaboration. The Life Science Grid Research Group of the Global Grid Forum (see LSG, n.d.) aims to investigate how bioinformatics requirements can be fitted and satisfied by grid services and standards, and vice versa, what new services should grids provide to bioinformatics applications. Some bioinformatics grids projects are also appearing, for example, the EuroGrid project (EuroGrid, n.d.), the Bio-GRID work package (Bio-Grid, n.d.) used to access portal for biomolecular modeling resources, the *myGrid* (Stevens, Robinson, & Goble, 2003) system, and the Asia Pacific Grid (AsiaGrid, n.d.).

Related Content

An Efficient Concurrency Control Algorithm for High-Dimensional Index Structures
Seok Il Songand Jae Soo Yoo (2004). *Journal of Database Management (pp. 60-72).*
www.irma-international.org/article/efficient-concurrency-control-algorithm-high/3316

Retrieval Optimization for Server-Based Repositories in Location-Based Mobile Commerce
James E. Wyse (2009). *Advanced Principles for Improving Database Design, Systems Modeling, and Software Development (pp. 255-281).*
www.irma-international.org/chapter/retrieval-optimization-server-based-repositories/4302

Profiling the Users of High Influence on Social Media in the Context of Public Events
Lu An, Junyang Hu, Manting Xu, Gang Liand Chuanming Yu (2021). *Journal of Database Management (pp. 36-49).*
www.irma-international.org/article/profiling-the-users-of-high-influence-on-social-media-in-the-context-of-public-events/276498

Toward a Visual Query System for Spatio-Temporal Databases
Valéria M.B. Cavalcanti, Ulrich Schieland Claudio de Souza Baptista (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 987-1002).*
www.irma-international.org/chapter/toward-visual-query-system-spatio/7954

Semantic Integration and Knowledge Discovery for Environmental Research
Zhiyuan Chen (2009). *Advanced Principles for Improving Database Design, Systems Modeling, and Software Development (pp. 39-61).*
www.irma-international.org/chapter/semantic-integration-knowledge-discovery-environmental/4291