INTERNET TOOLS FOR COMPUTATIONAL BIOLOGY

R.DOELZ
BioComputing Basel, Biozentrum der Universität, Klingelbergstrasse 70
4056 Basel,Switzerland

Electronic access to computational tools in Biology is more important than ever. Electronic mail servers have been in use for a very long time already in order to access large archives of molecular biology data, such as the sequence data providers. The expansion of campus information systems (e.g., GOPHER) allowed to implement immediate, interactive access to resources in both query and retrieval. More recently, the burst of World Wide Web has accomplished to make data accessible on networks which have never been visible in this kind of presentation before. Currently, 'the Web' is the most sophisticated tool for information handling which can be accessed by novices. Even more sophisticated tools allow to gather information and computations more specifically, but require to employ specific client software rather than common-purpose browsers. The most important development today is to enhance the access protocol for client-server applications and to expand server-server communication for data update and data presentation issues.

1 Overview

Information in Biology is growing rapidly. There are various areas of information processing which benefit from networked information. Table 1 lists the methods and tools in use today. It is important to realize that the end user is mostly affected by the **client-server** communication which implies that the browser or special access program will need to run on desktop equipment. Sophisticated computational work is performed on the remote sever with very little use of networking, and the work load is shifted to the remote end entirely. The **server-server** communication, however, becomes increasingly important for database maintenance and requires that both ends are computationally active.

2 End-user aspects

The World Wide Web Virtual Library Biosciences page[a] soon became the reference point of many servers. The page was split into subsections, and various curators now maintain the biology area in a coordinated, collaborative effort. Site presentation, teaching, discussion forums and database access are the main domain of WWW access tools.

[a] http://golgi.harvard.edu/biopages.html

Table 1: Current Methods and Tools for Networked Biology

Feature	Community	Purpose	Protocol
Teaching	beginner	learning	various
Data mining	beginner, casual user, researcher, provider	information, retrieval, resource discovery	http (WWW) and various specialized
Discussion forums	end user, provider	discussion	nntp (NEWS)
Computational analysis	researcher	remote resource utilization	http (WWW) and various specialized
Data synchronization	provider	maintenance	specialized

Numerous computational tools are available. Some are implemented as standalone application using a specific client/server protocol. If WWW is used, the state of a multi-query dialogue has been transmitted using various approaches, such as retransmission of all data, the transmission of a 'identifier' and a stateful engine at the server site which inherits communication via WWW.

A very interesting approach is to utilize tools like the Moo, a virtual reality world which can be operated by keyboard and normal internet access. Those meeting places can be supplemented with WWW pages as discussion basis. More recently, tools became available to expand WWW into three dimensions. The Virtual Reality Modeling Language (VRML) extends the World Wide Web (WWW) interface with the additional ability to visualize three dimensional (3D) object oriented scenarios. Since most of the models in molecular science need 3D representations the new technique can be applied very effectively in biochemistry information networks.

3 Provider aspects

The time-dependency of databases in biology (exponential growth has been observed in the past) raises the difficult problem of synchronization. Quality control and multi-site updates are tackled with dedicated, special purpose protocols. Besides proper characterization of the data set itself as a function of time, the main challenge of synchronization will be an important issue in future biocomputing research on the internet.