WebLab: a data-centric, knowledge-sharing bioinformatic platform

Wang Jun 王珺
Center for Bioinformatics
Peking University
wangjun@mail.cbi.pku.edu.cn
Introduction

Why great demands are proposed for an integrative system?

1. Large biological data are produced
2. Thousands of literatures are pouring
3. Plentiful complex bioinformatic analysis tools
4. Collaboration of researchers from various fields
Data Management: “My Data”

Data can be uploaded locally or retrieved from remote database.

“Tag” View

“Comment” knowledge
Define your own databases

Blast against your own databases
Analysis Service: “Program”

- 264 tools in total
- Uniform interface
- Three types of tools:
  1. Command-line programs
  2. Web-services
  3. Grid-services
Analysis Service: “Workflow”

- **Workflow**
  Complex analysis tasks can be run automatically

- **Two types:**
  Macro Protocol

- **Customized**
Analysis Service: “Resource”

We provide uniform interface for different data sources.
Analysis Service: “Utility”

Utility includes some mini popular bioinformatics analysis tools written in javascript or applet. If you know some other applications which are pretty useful, please give us some advice at your free time. 🛠️ utility help

- Expand all - Collapse all

- Bioreverse
- Dotlet
  Dotlet(applet) - Dotlet is a program for consensus
- GlycanBuilder
- Vimida
- Sequence Manipulation Suite
  - Sequence Analysis
  - Sequence Figures
  - Format Conversion
  - Miscellaneous
  - Random Sequences
- jalview
  jalview(applet) - Jalview is a multiple align...
- WebMol

[Image of a software interface with a genetic sequence analysis window open]
Team Work: “Group Strategy”

User can create a Virtual Research Group (VRG)

User attended VGRs
Team Work: “Sharing Mechanism”

Share data/literature to a user or a group

Share workflow to a user or a group
WebLab's whole source code is released as “Free Software” under the GNU General Public License version 3 (GPLv3), and freely available to anyone for downloading.

WebLab 1.0 Source Code
Web interfaces

Most web interfaces to EMBOSS provide a form for setting options and running applications. Difficulties can arise because of 'dependencies' in the application input. These include where the value of one parameter depends on the value or property of another, for example sequence length, or where the available options depend on the input, for example whether a sequence is DNA or Protein. Some web interfaces handle dependencies better than others.

To use a web interface, all you need are a web browser and internet access. Some of these interfaces run 'anonymously' so any user can run the applications. In most cases however you will be expected to register with the site before you can use their services.

The interfaces we have seen so far are:

**WebLab**

WebLab collect provide easy web access to the most popular bioinformatics programs. It is task-oriented instead of tool-oriented. It provides pre-defined protocols to guide biologists to work step by step through the entire analysis procedure.

**Pise**

PISE (french pronunciation like the English word 'peas') was developed by Catherine Letondal at the Institute Pasteur. Using an XML definition for each application, PISE generates a Web interface and other common interfaces automatically.

For EMBOSS, the ACD files were converted with scripts to generate the XML definitions. Where ACD files caused problems, we simply fixed the ACD files.

**EMBOSS Explorer.**

An extremely popular web interface to EMBOSS from Luke McCarthy at Plant Biotechnology Institute, Canada.
BioMart is a query-oriented data management system developed jointly by the Ontario Institute for Cancer Research (OICR) and the European Bioinformatics Institute (EBI).

The system can be used with any type of data and is particularly suited for providing ‘data mining’ like searches of complex descriptive data. BioMart comes with an ‘out of the box’ website that can be installed, configured and customised according to user requirements. Further access is provided by graphical and text based applications or programmatically using web services or API written in Perl and Java. BioMart has built-in support for query optimisation and data federation and in addition can be configured to work as a DAS 1.5 Annotation server. The process of converting a data source into BioMart format is fully automated by the tools included in the package. Currently supported RDBMS platforms are MySQL, Oracle and Postgres.

BioMart is completely Open Source, licensed under the LGPL, and freely available to anyone without restrictions.

Powered by BioMart software:
- Dictrybase
- Wormbase
- Stamene
- Eurohomo
- Uniprot
- Rat Genome Database
- DROSPD
- ArrayExpress
- Tuxexpress
- GermOnline
- PRIDE
- PepSeeker
- VectorBase
- Pancreatic Expression Database
- Reactome
- EU Rat Mart
- Paramacium DB

Third party software with BioMart Plugin:
- Bioclipse
- BiomaRt
- BioConductor
- Cytoscape
- Galaxy
- Taverna
- WebLab

http://www.biomart.org