

Poster B-24

WebLab: a workflow-based platform integrating bioinformatics resources



Authors:

Jianmin Wu (*Center of Bioinformatics, Peking University, Beijing 100871, China*)

Ge Gao (*Center of Bioinformatics, Peking University, Beijing 100871, China*)

Xiaoqiao Liu (*Center of Bioinformatics, Peking University, Beijing 100871, China*)

Lei Kong (*Center of Bioinformatics, Peking University, Beijing 100871, China*)

Zhe Li (*Center of Bioinformatics, Peking University, Beijing 100871, China*)

Di Liu (*Center of Bioinformatics, Peking University, Beijing 100871, China*)

Wen Tang (*Center of Bioinformatics, Peking University, Beijing 100871, China*)

Jingchu Luo (*Center of Bioinformatics, Peking University, Beijing 100871, China*)

Short Abstract: We developed a bioinformatics platform WebLab integrating various sequence analysis tools. WebLab was designed biological end-users with a single and simple Web interface which allows users to input sequence data and manage analysis results. WebLab can also be used by bioinformaticians as a workflow-based system for further development.

Long Abstract:

With the great amount of genome data generated every day, a paradigm shift has been emerged in biological science and biotechnology during the past decade. Wet lab biologists access molecular databases and analysis tools on a daily basis. It is no doubt that “half day on the Web, saves you half month in the lab”. However, novice users may have problems to use the bioinformatics packages efficiently due to the heterogeneity of the available tools and to manage the various input data and analysis results. We have developed a bioinformatics platform WebLab to solve this problem.

WebLab was designed to provide the end-users in the molecular biological community with a single and simple Web interface by integrating various sequence analysis packages including the European molecular biology open software suite (EMBOSS), the NCBI database search tool BLAST, the phylogenetic analysis package Phylip, the multiple sequence alignment program ClustalW, as well the KEGG orthology annotation system (KOBAS) developed locally. Users can input their sequence data through typical cut-and-paste or browse-and-upload approach, or use accession number or sequence ID to retrieve entries from the databases indexed on the server. Users can edit their data such as the descriptions of FASTA format sequences. The output results can be shown in either text or graphics format. Both the input data and the output results can be stored in a user space created on the server and managed by a database system running at the background. Users can manipulate their data in the tree-like folder and file system, and share their data with trusted users, for example, another user working in the same lab. With the aim of making the routine analysis easier, WebLab borrowed the protocol approach from wet lab biological experiments and implemented several sequence analysis pipelines. Users may follow the built-in protocols to do step-by-step analysis for single DNA or protein sequence, such ORF finding, translation, profile scaling, similarity search, motif scanning, as well as multiple sequence analysis such as sequence alignment, phylogenetic tree construction and motif

search. Users may also create their own protocols on-fly and store them on the server. WebLab can also be used by bioinformaticians as a workflow-based system for further development. Bioinformatics analysis for problem solving is often a complex task that involves numerous bioinformatics tools. We used directed acyclic graph (DAG) to model the workflow and divided workflows into two types based on whether user interactions are allowed. Tools appear as nodes and state information is represented as conditions attached to edges. Flexibility is carefully considered in design and adding new tool only needs writing an XML file. Extension mechanism is supported for renders to format input or output data. Bioinformatics developers can plug-in their own extension classes to the platform. WebLab was developed using the platform-independent Java language. Apache tomcat was used as container for Java Servlet and JSP. The architecture of the system has a layered topology using proxy model. The portal server has a unified interface for the analysis tools, and acts as a proxy to accept all user requests. It is made up of separate layers for job scheduling, renders and database action. Once a workflow is requested, the scheduling layer will traverse the workflow model represented as DAG and find initial analysis, then look up for appropriate computing servers and send requests. The dispatch daemon running on the computing server will execute the required tool after accepting a request. The result will be sent back to the portal server after the analysis is completed and saved into backend database. If it has subsequent analysis, the scheduling layer will automatically send request to appropriate computing server. Thus it allows tools in a workflow to distribute on different servers.

The first version of WebLab is available for testing at <http://weblab.cbi.pku.edu.cn/>.