

SWS: accessing SRS sites contents through Web Services

ID - 218

Romano Paolo¹, Marra Domenico¹

¹Bioinformatics and Structural Proteomics, National Cancer Research Institute, Genova

Motivation

Web Services (WS) and Workflow Management Systems (WMS) are new ICT tools that can support the creation and deployment of flexible network systems. WS have been implemented at bioinformatics centres; examples are Entrez Utilities at NCBI and the SoapLab implementation at the EBI. Some WMS have been proposed in bioinformatics, the Taverna Workbench probably being the most known. While new network systems are often developed by taking into account these new technologies, many already existing databases does not offer a programmatic access to them.

SRS (Sequence Retrieval System) is a well know indexing and search engine for biomedical databanks [1]. Through its many public sites, it offers access to more than 1,000 databanks and about 180 analysis tools. Unfortunately, these data are not accessible through WS. In this paper, we present SRS by WS (SWS), a tool that can improve the accessibility of information available in public SRS sites through a programmatic access. It does not only enable interoperability between SRS and WMS, but it also manages access to alternative equivalent sites in order to cope with network problems, by also selecting the most up-to-date among available systems.

Methods

Our system is based on the information on SRS public sites that is provided by BioWisdom. This reference site is checked daily and information regarding available sites, databanks and tools are stored into a local relational MySQL database.

Databases contents are kept up-to-date by running a perl script that check the status of the sites each hour.

Web Services were deployed by using Soaplab, a SOAP-based Analysis Web Service providing a programmatic access to local, command-line applications and to the contents of ordinary web pages [2].

The only requirements of Soaplab are the Apache Tomcat servlet engine with the Axis SOAP toolkit, a Java Virtual Machine and, optionally, perl and mySQL. New Web Services are deployed (added to the system) by defining simple descriptions of related execution commands. Definitions are written in the AJAX Command Definition (ACD) language and are then converted into XML before they can be used.

Web Services were implemented by perl and PHP scripts that query the local database for retrieving the needed information on available sites and databases and then perform the actually requested queries by searching the SRS and retrieving results by means of the GNU Wget non-interactive network downloader. SWS allows to query selected systems and to retrieve essential information on the sites, such as actual availability and lists of included databases and tools, and on available databases, such as sites where they are implemented and related sizes/versions. WS SWS can just be invoked by specifying the name of the databank to be queried and the query terms. It then automatically choose the best site, performs the query and returns the complete results. Users can also specify the following information: the SRS site to be queried, the fields where the information must be searched, the desired output fields. SWS is able to query sites having both the most recent versions of SRS, i.e. versions 7 and 8.

Results

The development and implementation of WS, that allows to make access to an exhaustive set of biomedical databases and analysis software and to create effective workflows through WMS, can significantly improve automation of in-silico analysis.

Many databanks are currently available, but only a few of them can be queried through programmatic interfaces. A list of public SRS sites list more than 1,000 databanks available in about 40 sites. These can be made accessible by developing a suite of WS that is able to query these sites and return results as a WS. We have presented in this paper SWS, a suite of WS allowing to query SRS sites and return results in a simple text-only format. SWS is being developed as an open source and it is still under test and further development.

Availability: <http://bioinformatics.istge.it/sws/>

Email: paolo.romano@istge.it