Databases and ontologies

REMORA: a pilot in the ocean of BioMoby web-services

Sébastien Carrere and Jérôme Gouzy*

INRA-CNRS Laboratoire des Interactions Plantes Micro-organismes (LIPM) BP 52627, 31326 Castanet Tolosan Cedex, France

Received on August 19, 2005; revised on November 28, 2005; accepted on January 12, 2006

Advance Access publication January 19, 2006

Associate Editor: Charlie Hodgman

ABSTRACT

Summary: Emerging web-services technology allows interoperability between multiple distributed architectures. Here, we present REMORA, a web server implemented according to the BioMoby web-service specifications, providing life science researchers with an easy-to-use workflow generator and launcher, a repository of predefined workflows and a survey system.

Contact: Jerome.Gouzy@toulouse.inra.fr

Availability: The REMORA web server is freely available at http://bioinfo.genopole-toulouse.prd.fr/remora, sources are available upon request from the authors.

INTRODUCTION

In the recent past, owing to the broad heterogeneity of bioinformatics resources, the interoperability between distributed systems was an obstacle for integration, both for bioinformaticians and life science researchers (Stein, 2002). Henceforth, several bioinformatics reference centres (Pillai et al., 2005; Wilkinson et al., 2005) consider the emerging web-services technology as the communication protocol that will provide a breakthrough in sharing distributed resources, as well as in the construction of high-level analysis workflows. Currently, two communication protocols, BioMoby (Wilkinson and Links, 2002) and mygrid (Stevens et al., 2003) are emerging in the bioinformatics research field. On the client-side, the field of workflow design is under investigation and software such as Taverna (Oinn et al., 2004) and framework prototypes such as GPIPE (Garcia Castro et al., 2005) are already available.

Owing to the early stage in the application of this technology in bioinformatics, the user interfaces provided by this type of software are generally designed for bioinformaticians and then are focused on abstraction and flexibility. As a consequence, the access cost increases for laboratory biologist end-users who generally prefer to avoid abstraction steps to focus directly on functionalities that can be used to analyse their own data. For these reasons we have developed the REMORA web server that extensively uses the BioMoby data types to guide users in the discovery of existing services. As gbrowse_moby (http://mobycentral.icapture.ubc.ca/cgi-bin/gbrowse_moby), but in contrast to Taverna, REMORA allows the discovery of services without a prior knowledge of tools. Unlike gbrowse_moby, but like Taverna, REMORA allows graphical construction and parameter settings of analysis workflows. As a complement to Taverna, which is primarily designed for use by bioinformaticians (Oinn et al., 2004), REMORA is an end-user oriented tool which aims to provide life science researchers with direct access to web-services.

MAIN FEATURES

REMORA has been designed to create and launch BioMoby workflows. This means that, in accordance with BioMoby specifications, REMORA manages parameters via secondary articles as well as namespaces and services with incoming or outgoing collections. From an ergonomic point of view we have simplified the interface by using the standard scheme of road crossing colour codes. Thus, the colour red represents blocking steps and running problems, while the colour green represents ready to go services and achieved computations and the orange code represents a punctual warning such as removal confirmation or a settings edition. In addition, colour blind people can run REMORA using a grey-scale based layout.

Workflow design and submission

The user starts the workflow design by adding a type or a namespace to the data he/she wants to analyse, then, depending on this information, REMORA displays a list of applicable web-services from which the user can select one or several services to be run. After submission, rectangular icons representing services will appear in the workflow graph as well as octagons representing output data.

Thus, the user can progress step-by-step in the workflow building process by ‘clicking’ on a data icon to get the list of applicable services and then adding more web-services to the workflow. In this way, REMORA acts not only as a discovery tool but also as a step-by-step generator of consistent workflows (in the limit of web-service semantic consistency). Once the graph has been designed, the user enters into the configuration mode, to first enter the query data (e.g. a string, a fasta file, etc.) and next to set the web-service parameters. These two actions can simply be performed by clicking on the service icon in order to get the appropriate web form. The design step ends with a submission form composed of two fields, one for a short workflow description and the other one to enter the user e-mail address to send the URL of the result web page after workflow completion.

Visualization

The same graphical representation is used for data visualization. In the same way that any service can be configured by clicking on its icon, one can click on a data node to access the web-service results. The REMORA graph can be saved to and reloaded from the user’s
computer either to see the results or to run the workflow again after editing if needed.

Survey mode
For many database searches, it is relevant to run the same workflow regularly. For this, the user receives in the result e-mail an additional internet address from which he/she can access a form to programme the workflow to run weekly or monthly.

Frequently Asked Workflow (FAW) section
By analogy with the 'Frequently Asked Questions', a repository of the most common workflows is available on the server. Compared with saving a workflow on the hard drive, registration in the FAW has the advantage that workflow development effort can be shared, as well as the possibility to see the workflow itself as a potential analysis resource after the initial step of user data typing. Furthermore, the FAW repository enables any visitor to quickly make use of a set of validated workflows.

Hook mode
This feature does not target end-users. Its purpose is to enable the remote control of REMORA from any other web server. Typically, this function provides an access to web-services or REMORA workflows from any web server on condition that data to be analysed are reachable via the web.

The full description of the 'hook' command is available in the 'Help' section. This section also provides several links to movie files that dynamically present all REMORA features.

REMORA is a PERL-CGI web server that uses the BioMoby perl API 0.8.2a. In creating REMORA, the ‘CommonSubs’ library has been largely rewritten using the xsltproc processor in order to speed up the processing of large sized XML files. The GraphViz (http://www.graphviz.org/) perl module is used both for graphic workflow representations and navigation. Some gbrowse_moby renderers have been adapted to REMORA. In the future, we will integrate further BioMoby specifications with the aim of establishing a compatibility link with Taverna workflows as soon as it is possible to do this without any functional loss. In addition, we will improve REMORA by integrating relevant evolutions that are expected to occur in the semantic MOBY project.

ACKNOWLEDGEMENTS
The authors thank Clare Gough and Thomas Faraut for feedback and helpful comments on the manuscript and the Bioinformatics platform of the Genopole Toulouse Midi-Pyrénées for technical support.

Conflict of Interest: none declared.

REFERENCES