Biowep: a workflow enactment portal for bioinformatics applications

Romano P (1), Bartocci E (2), Bertolini G (3), De Paoli F (3), Marra D (1),
Mauri G (3), Merelli E (2), Milanesi L (4,5)

(1) National Cancer Research Institute, Genoa
(2) University of Camerino, Camerino (MC)
(3) University of Milan Bicocca, Milan
(4) National Research Council, Milan
(5) CILEA, Segrate (MI)

Motivation

The huge amount of biological information, its distribution over the Internet and the heterogeneity of software tools that are used in bioinformatics makes the adoption of new data integration and analysis network tools a necessity. Information and Communication Technologies (ICT) standards and tools, like Web Services (WS) and Workflow Management Systems (WMS), can support the creation and deployment of such systems. WS are network services usually communicating by using the Simple Object Architecture Protocol (SOAP), a framework for the distribution of XML structured information, over HTTP. They offer a standardized programming interface so that software tools can effectively make access to the information and services they are delivering. Hence, they allow software applications to identify and interpret the information and, when ontological metadata is added, the associated semantics. WS have been implemented at bioinformatics centers; examples are Entrez Utilities at NCBI and the SoapLab implementation at the EBI, through which all EMBOSS programs are available. Workflows, defined as "computerized facilitations or automations of a business process, in whole or part" (Workflow Management Coalition, WfMC), aim to implement data analysis processes in standardized environments. Their main advantages relate to effectiveness, reproducibility, reusability of intermediate results and traceability. Some WMS have been proposed in bioinformatics. The Bioinformatic Workflow Builder Interface - BioWBI, Pipeline Pilot and Taverna Workbench [1] are the most known. These WMS assume that end users know which bioinformatics resources can be reached through a programatic interface and that they are skilled in programming and in building workflows, but they are not viable to the vast majority of researchers that are customised to web interfaces. A portal enabling the vast majority of unskilled researchers to take profit from these new technologies is still missing. We present here a web system that can support selection and execution of a set of predefined workflows. It presents a user-friendly web interface that is able to simplify access to such workflows and it therefore is viable to all end users.

Methods

The conceptual architecture of our system includes a Workflow Manager (WM), a User Interface (UI) and a Workflow Executor (WE). The WM is external to the prototype. Its task is the creation of predefined annotated workflows. These can be created by using different WMS. Presently, we allow for two of them: the Taverna Workbench [1] and the BioWMS [2]. Workflows enactment is carried out by the FreeFluo tool for Taverna workflows and by BioAgent/Hermes [3], a mobile agent-based middleware for the design and execution of activity-based applications in distributed environments, for BioWMS ones. The main processing steps of each workflow are annotated on the basis of their input and output data, elaboration type and application domain. Annotations are defined by using a classification of bioinformatics data and tasks. The UI supports end users authentication and profiling, including the classification of users on the basis of their job/role and scientific interests. Workflows can be selected on the basis of users' profiles. Available workflows can be searched through their annotation. Results of the execution of workflows can be saved and later analysed and possibly reused.

Results

We designed a web based client application, as defined in the WfMC Reference Model, that allows for the selection and execution of a set of predefined, annotated workflows. A prototype system is

available on-line. It includes workflows that are devoted to the retrieval of data from IARC TP53 Mutation Database and from CABRI biological resources catalogues. Some of them have been made available both in Taverna and in BioWMS formats. Performances of the two approaches are under evaluation. The development and implementation of WS allowing the access to an exhaustive set of biomedical databases and analysis software and the creation of effective workflows through widely distributed WMS can significantly improve automation of in-silico analysis. biowep is available for interested researchers as a reference portal. They are invited to submit their workflows for insertion in the workflow repository. biowep is further being developed in the sphere of the Laboratory of Interdisciplinary Technologies in Bioinformatics – LITBIO.

Availability: http://www.o2i.it:8080/biowep/index.jsp

Contact email: paolo.romano@istge.it

References
1. T. Oinn, M. Addis, J. Ferris, D. Marvin, M. Senger, M. Greenwood, T. Carver, K. Glover, M. R. Pocock, A. Wipat and P. Li, Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics, 20(17):3045-3054, 2004
2. Bartocci E., Corradini F., Merelli E., Scortichini L., BioWMS: a web based Workflow Management System for Bioinformatics. Submitted to BITS 2006.
3. F. Corradini and E. Merelli. Hermes: agent-base middleware for mobile computing. In Mobile Computing, volume 3465, pages 234-270. LNCS, 2005.

Supplementary informations
This work was partially supported by the Italian Ministry of Education, University and Research (MIUR), projects "Oncology over Internet (O2I)" and "Laboratory of Interdisciplinary Technologies in Bioinformatics (LITBIO)". Our system is partially based on open source. biowep is itself available under the GNU Lesser General Public Licence (LGPL). The portal is still under test.
Links Taverna Workbench: http://taverna.sourceforge.net/
BioWMS: http://litbio.unicam.it:8080/biowms/ (demo under testing)
BioAgent: http://www.bioagent.net/ Hermes: http://hermes.cs.unicam.it/