First International Workshop on Semantic Infrastructure for Grid Computing Applications (SIGAW)

Workshop Preface

Chair: Line Pouchard[1], co-chairs: Luc Moreau[2], Valentina Tamma[3]
[1]*Oak Ridge National Laboratory, USA*
[2]*Southampton University, UK,* [3]*Liverpool University, UK*
[1]*pouchardlc@ornl.gov*, [2]*L.Moreau@ecs.soton.ac.uk*,[3] *valli@csc.liv.ac.uk*

1. Program Committee

Hafiz Farooq Ahmad, Communication Technologies, Sendai, Japan
Naveen Ashish, NASA Ames, USA
Mario Cannataro, University "Magna Græcia" of Catanzaro, Italy
Dan Cook, University of Washington, USA
Ewa Deelman, ISI, University of California, USA
David De Roure, Southampton University, UK
Ian Foster, Argonne National Laboratory, USA
Yolanda Gil, ISI, University of California, USA
Mike Huhns, University of South Carolina, USA
Rich Keller, NASA Ames, USA
Carl Kesselman, ISI, University of California, USA
Manolis Koubarakis, Technical University of Crete
Bertram Ludaescher, SDSC, University of California, San Diego, USA
Reagan Moore, University of California, San Diego, USA
Jim Myers, Pacific Northwest National Laboratory
Benno Overeinder, Vrije Universiteit, NL
Marlon Pierce, University of Indiana, USA
Daniel Rubin, Stanford University, USA
Andrew Woolf, Rutherford Appleton Laboratory, UK

2. Message from the chair

Pressing needs have emerged in several domain sciences and grid computing applications for an adequate description of the large volumes of data produced by data-intensive simulations and experiments on scientific instruments. The data produced by scientific applications including climate modeling, high throughput biology, proteomics, high energy physics, astronomy, and the knowledge derived from these applications may loose its value in the future if the mechanisms for inventory, cataloging, searching, viewing, retrieving, and presenting generated data are not quickly improved. For example, at the end of 2004, the volume of climate modeling data cataloged by the Earth System Grid was about 100 Terabytes (1.2 million files) distributed across several storage facilities. Other sciences such as biomedical science and bioinformatics produce smaller but thousands of diverse and widely distributed files stored on individual desktops and databases. Faced with an impending data crisis, scientists and data managers are forming partnerships with computer scientists for developing adequate solutions: semantic-based data descriptions, models, and services may play a crucial role.

The workshop investigates promising research and emerging technologies for semantic systems in the context of Grid computing. Technologies borrowed from the Semantic Web and the Digital Library community are prominent. Ontologies and ontology-driven systems are used to compose workflows, mediate between application semantics, and provide resource description. As successful prototypes move towards implementation and deployment the Semantic Grid is gaining recognition.

3. Summary of accepted papers

"Ontology-based Service for Grid Resources Description" presents the example of an ontology-driven application based on a description of static and dynamic states of resources. In "Semi-Automated Preservation and Archival of Scientific Data Using Semantic Grid Services," a prototype data preservation system is based on the development of an OWL-S ontology for reasoning over a description of 'preservation services.' The approach proposed in “Deductive Synthesis of Workflows for e-Science” uses theorem proving techniques to automate the construction of workflows. “Bootstrapping the Semantic Grid” presents the Scientific Annotation Middleware, an operational system that extracts existing metadata and the lessons learned in its implementation for a multi-scale chemical science collaboratory. “Semantic Integration of File-based Data for Grid Services” presents a use case for the Earth Sciences and work in progress for virtualization of file-based data. Finally, bioinformatics data integration is the topic for both “Using Semantic Web Technology to Automate Data Integration in Grid and Web Service Architectures” and “A Semantic Grid-based Data Access and Integration Service for Bioinformatics.” The former describes the development of a mapping language to convert representations of sequence data using OWL between bioinformatics applications; in the latter a mediator architecture is used to integrate bioinformatics knowledge.

4. Conclusion

The successful approaches presented in these papers illustrate various ontology-based systems that add some semantic capabilities to Grid computing. Domain sciences that have so far benefited the most are bioinformatics, the earth sciences, and the Collaboratory for Multi=Scale Chemical Sciences. Much remains to be done. For instance, a lightweight semantic architecture that offers flexible solutions for grid applications is needed. More tools for automatic capture of metadata and semantic-based searches should be developed to answer the specific needs of some domain sciences. Ontology repositories and ontology federation could be investigated for the creation of virtual data stores. Discussions at the workshop will hopefully bring some light on these questions.