Metadata, Ontologies, and Information Models for Grid PSE Toolkits Based on Web Services

Carmela Comito, ICAR-CNR, Italy
Carlo Mastoroianni, University of Calabria, Italy
Domenico Talia, University of Calabria, Italy

ABSTRACT

A PSE toolkit is a group of technologies within a software architecture through which multiple PSEs can be built for different application domains. The effective use of a PSE toolkit requires the management of the heterogeneity of the involved resources that can include computers, data, network facilities, sensors, and software tools provided by different organizations. A distributed implementation of a PSE toolkit can be envisioned through the exploitation of features and functionalities offered by a service-oriented Grid framework, so obtaining a Grid PSE toolkit based on Web services. This paper presents a metadata model for Grid PSE toolkits based on Web services and the architecture of an information system that exploits the proposed metadata model. These two components contribute to define a general model of metadata management for supporting the design and implementation of problem solving environments on Grids.

Keywords: grid; information system; metadata management; metadata model; ontology; PSE; Web services; WSRF

INTRODUCTION

A problem solving environment (PSE) is a computer system that provides the computational features necessary to solve a target class of problems, according to the well-known definition reported in (Gallopolous, Houstis, & Rice, 1994). PSEs for industry, commercial, and business applications are gaining popularity in the recent years. An advancement of the PSE concept is the PSE toolkit concept. A PSE toolkit is a group of technologies through which multiple PSEs can be built for different application domains.

PSEs can benefit from advancements in hardware/software solutions achieved in parallel and distributed systems. In particular, the Web service paradigm and the Grid emerged as very interesting computing models in the
area of parallel and distributed computing. The Web service paradigm enables flexible, platform-independent, and largely automated interactions between Web-resident services and applications, promoting the interoperation among them. The Grid is a novel infrastructure for network computing on local or geographical scales that can dynamically embody heterogeneous computing resources. Grid computing is today broadly used in many scientific and engineering application fields and is attracting a growing interest from business and industry.

The recently proposed OGSA architecture (Open Grid Services Architecture [Foster, Kesselman, Nick, & Tuecke, 2002]) aligns Grid technologies with Web services technologies to take advantage of important Web services properties, such as service description and discovery, automatic generation of client and service code from service description, compatibility with emerging higher-level open standards and tools, and broad commercial support. To achieve this goal, OGSA defines uniform exposed service semantics, the so-called Grid service, based on principles inherited from both the Grid computing and the Web services technologies.

The research and industry communities, under the guidance of the Global Grid Forum (GGF) (GGF, 2005), contributed to evolve OGSA toward the Web Services Resource Framework (WSRF) (WSRF, 2005), that completes the integration between Grid services and Web services. WSRF specifications define a generic and open framework for modeling and accessing stateful resources using enriched Web services referred to as WSRF Web services. This framework comprises mechanisms to describe views on the state and to support management of the state through properties associated with the Web services.

In order to fulfill the requirements of a PSE toolkit in a distributed environment, and according to the evolution trend discussed so far, this paper aims to exploit Grid and Web services features to enhance the functionalities of a PSE toolkit in a multi-domain environment. A “Grid PSE toolkit based on WSRF Web services” can indeed benefit from the advanced services and components offered by these novel technologies, such as security components, dynamic resource management services, resource discovery services, services for the parallel, and distributed execution of complex applications. In particular, this paper focuses on the development of an information system for a multi-domain PSE toolkit and on the definition of a flexible and semantically enriched metadata model.

An efficient information system is a key component because a PSE toolkit needs to manage a large variety of resources that can include computers, data, network facilities, sensors, and software tools provided by different organizations (Cannataro, Comito, Congiusta, Folino, Mastroianni, Pugliese, et al., 2004). The management of such heterogeneous resources requires the use of metadata that, through an accurate categorization of resources, provides useful information about the features of resources and their usage modalities.

As opposed to a single domain PSE, in a multi-domain PSE toolkit the structure of metadata information is not uniform: it depends on the type of the resource (i.e., software, hardware, data, etc.), and on the application domain in which the resource is used.

Accordingly, we propose a metadata model that can be flexibly exploited in a number of application domains, and at the same time is suited to be specialized in a particular application domain. In particular, we propose to associate a metadata document to each resource offered by the PSE toolkit and distinguish three sections within that document: an ontological metadata section that identifies the resource category, a semantic metadata section that characterizes resources in different application domains and assists discovery services, and a resource metadata section that gives details about how to use and access a resource. The rationale of such distinction comes from the consideration that, in a PSE toolkit, resources must be annotated with metadata information at different levels and at different times.

Moreover, this paper introduces a novel architecture for a Grid-based information