

Come together, right now!

Semantic matching of Grid Resource Descriptions

Summary made by Pascal BIHLER, 2005-02-07

Introduction

A main concept in the use of grids for distributed computing or shared data management is the concept of virtualization, that means user virtualization and of course resource virtualization. A single user should work on the shared resources of a grid without worrying about having a user accounts on all the different machines evolved in his calculation or data search process. This virtualization of the user is normally handled by a digital certificate which represents the user all long his grid usage process from the machine's view and is passed from one computer to the next if the user's process is referred to another machine.

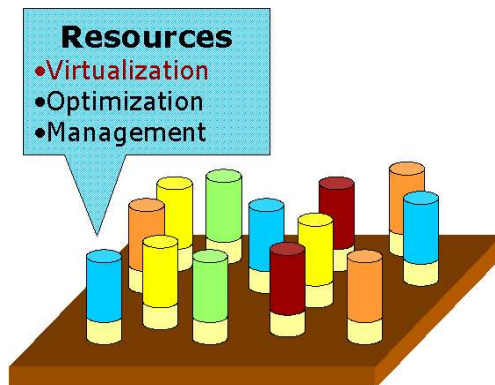


Image 1: Resources in OGSA context ([Atk04])

The virtualization of the grid resources is task of dedicated OGSA (Open Grid Service Architecture) Services, which take as well care of the grid-wide resource management and the resource use optimization. The concept of handling all the questions concerning the grid's meta-layer by specialized OGSA services is however a relatively new approach. Today's resource brokers rely on existing pre-OGSA middleware-structures just like the Globus Toolkit (GT2/GT3) or the UNICORE systems, both of them widely used throughout Europe and the United States. When merging different existing grids together to a virtual organization (VO), if not before, unavoidable the users are confronted with the situation of choosing one architecture

and restructure the existing grids which used other architectures before. This does not seem to be really an efficient way, but extending each broker with the capability to handle simultaneously interwoven requests on different grid architectures is a non trivial challenge as well. To provide a uniform access for resource brokers and other high level grid agents, a translation service based on the semantic correlations between the different grid structures has to be developed. In the article this summary is presenting ([Bro04]) the authors Brooke et al from the University of Manchester try to define a possible way to create in a semi-automatically way such a semantically translation service between the Globus Toolkit's GLUE-Schema and the UNICORE's resource description philosophy based on the UNICORE Incarnation Database (IDB).

Standards in describing grid resource semantics

The Globus Toolkit's way

The Globus Toolkit itself propose different protocols for finding and describing a resource, based on a network of Grid Resource Description Servers (GRIS) and Grid Index Information Servers (GIIS), which are managed using a Meta-Directory Service MDS-2 using a standard LDAP-Database. In the research of the Data Trans Atlantic Grid (DataTAG; encouraged by the CERN research laboratory in Geneva among others) one aim was to "insure interoperability mechanisms between US and Europe Grid domains and study the scalability issues which arise, including interoperation of grid security mechanisms and

interoperation of information services to permit inter-grid resource discovery" [Dat01]. Therefore the GLUE (Grid Laboratory Uniform Environment) schema was developed describing the resources of a grid managed with the Globus Toolkit as a meta model of the machine.

UNICORE's resource world

Another widely used grid technology is based on the architecture developed by the UNICORE Forum e.V. The UNICORE Forum, encouraged by the German research association, "promotes the use of high performance computers in science and research. It supports the Scientific Counsel's relevant recommendations by ensuring that authorized scientists at German research institutions and universities have easy and secure access from their workstation to national high performance computers." [Uni00] The software they have created is as well named UNICORE which stands for "uniform access to computer resources".

While GLUE is describing the resources independently from their implementing architecture or language, UNICORE is modeling a so called "universe of workflows", that means it reflects the way the resources are used in the description concept. Hence, it describes the resources as well as all the resource requests and information about available resources in a hierarchy of Java classes by creating Abstract Job Objects (AJO) to discover and manage the resources in the grid as well as the grid's jobs themselves. For communication between the grid's servers and the clients, the Java AJOs are serialized and sent throughout the network using the Basic UNICORE protocol layer (UPL) (on top of SSL) [Uni05]. The resources' virtualization is called Vsite and to each Vsite there is an Incarnation Database (IDB) associated responsible for the translation process between the AJOs and the concrete resource descriptions.

Getting two architectures together

Even though GLUE and UNICORE AJOs are designed to manage the same job, namely describing the grid's available resource and the request of them, they are approaching this aim in a completely different way and therefore, the description domains of UNICORE and GLUE are in almost no way congruent: While UNICORE is describing the user's point of view of the available software resources, GLUE addresses to the machine's view in describing dynamically the hardware resource which can be used. Nevertheless, a small intersection of the two domains can be examined and focusing on this area, a collaboration of the two network architecture seems to be possible using a translation service.

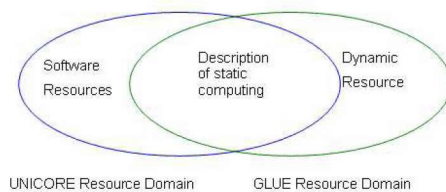


Image 2: Description resource domains ([Bro04])

The authors of the paper decided to use the AJO approach as "standard query language" for the resource broker and hereby AJO queries can be directly redirected to the corresponding EuroGrid brokering technologies where they have to be translated for requesting the GLUE world, which means the AJO requests have to be transformed into the corresponding LDAP searches on MDS-2. Finding these correspondences is not a negligible work so this part of the translation service was hand coded in the first prototypes of their uniform access middleware. Forming the heart of their work, this couldn't be the last word on the subject but rather they have found a way to create this ontology mapping in a more or less automatic way. They decided to base this automatically mapping on the documentation of the resource request models. For the UNICORE approach, it was possible to extract the JavaDoc data from the corresponding Java classes and extract the UNICORE ontology - the "descriptive language" from these documentation while GLUE already seemed to be sufficiently documented manually to take the GLUE ontology out of it's manual.

Extracting knowledge from prose

Unfortunately, documentation is not intended to be treated by a machine but basically for human beings saying that a document can be easily read by the operator where it is completely nonsense for the machine. Extracting knowledge from written text however is a challenge which is not only posed during a grid semantic matching process but as well in many other situations of an engineer's life (there is a specific research area called "knowledge engineering" handling with all these questions of how to acquire and process knowledge in an efficient way). The authors therefore decide to rely on a commercial solution, the PCPack tools provided by Epistemics Ltd, an "integrated suite of seven knowledge tools designed to support the acquisition and use of knowledge" [Epi04]. These tools support a number of key activities in knowledge treatment, such as "analyzing knowledge from text documents, structuring knowledge using various knowledge models, acquiring and validating knowledge from experts, publishing and implementing the captured knowledge and re-using knowledge across different domains" [Epi04]. The first job, analyzing knowledge from text documents, was needed in the context of deriving ontologies from the grid resource request architecture's documentation is treated by the so called "Ladder tool" which provides to build hierarchies of knowledge, as needed to describe the way of requesting a resource in the AJO style.

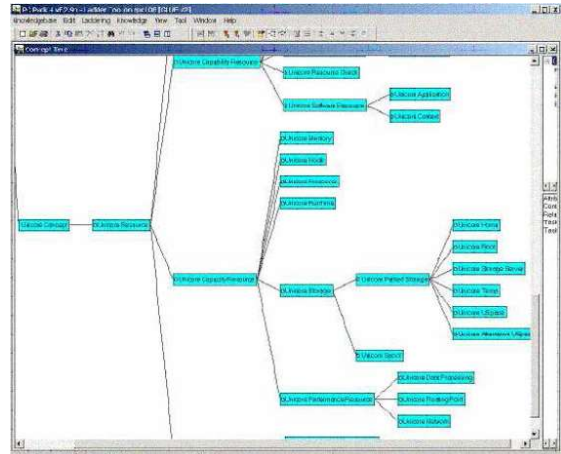


Image 3: Derived ontology hierarchy ([Bro04])

Hitches in automatic mapping

As mentioned above, the ontologies of UNICORE and GLUE cannot be mapped one-to-one due to their different point of view in describing the available and useable resources: While UNICORE is describing more or less the perception of the requesting user, GLUE is modeling the perspective of the resource providers. Certainly, this reflects on their ontology domains, that means on their way of describing the resources and the words they are using for it. UNICORE's "Priority value" can still be easily assigned to GLUE's "Priority", but already the "Maximum Memory Capacity Request" is described by four different parameters in Globus, namely "Host Virtual Main Memory Available", "Subcluster Virtual Main Memory Available", "Host RAM Main Memory Available" and "Subcluster RAM Main Memory Available". Because of these differences in the semantic of the ontologies, the mapping could just be realized "as far as possible".

To provide the translation from the UNICORE request to the Globus resource query, the derived ontologies are firstly translated from the PCPack format to an easier processible XML dataset, followed by creating an ontology structure in the main memory based on this XML data. With the help of this knowledge, the UNICORE resource checks are compiled to XPath queries for accessing the GT3 resource index servers. For accessing a UNICORE resource, the UNICORE IDB is directly queried based on an XPath query if a corresponding mapping exists for this resource, otherwise (because the matching is in no way exhaustive) the virtual resource is searched with the help of the Index Service.

This algorithm is implemented in a prototype, an "Ontological translator". The ontological data derived with PCPack from the documentation files is hereby provided in an XML form like the following snippet:

```
<translator>
  <domain name="Unicore">
    <element name="el1" id="1" native="uni1">
      <attribute ... />
    </element>
    ...
  </domain>
  <domain name="GLUE">
    ...
  </domain>
  <mapping name="Map1" from="Unicore" to="GLUE">
    <map from="el1" to="el2" />
    ...
  </mapping>
</translator>
```

Regrettably, the authors do not focus on the integration of this prototype into their broker's implementation explicitly nor do they provide any experimental results (Percentage of translatable requests for instance).

Critics and conclusion

This work seems to be the beginning in introducing a semantically ontology mapping into the existing grid broker's architectures, but the authors are right saying: "This is only a start, only a small subset ... can be mapped... However, it is a necessary start" [Bro04]. Maybe, this work will really be included into an OGSA service as aimed by the researchers and as long as the OGSA has not fully unified the world's grid architectures, there is a need for a translator service between the different grid systems for sure. But regarding this importance, the paper seems to be unfortunately small and sometimes uncompleted as there are spots where important information to follow their approach are missing or vague. In addition, the matching itself seems to be more or less a straight forward solution, originality can only be found in their creative way of deriving the architectures' resource request ontologies from their documentation. The absence of providing experimental results of their translation prototype is a pity, more grave is that they didn't mention with any word if their approach is as well suitable for data or computing grids of huge extensions saying that no kind of information or reflection about performance and scalability is given. But nevertheless, even if this concrete paper seems to be written a little bit in a rush, the research behind it seems to be profound and seminal.

Bibliography

- [And04] Andreozzi, S. *Glue schema implementation for the LDAP model [online]*, Technical Report. INFN/TC-04/16. 30 September 2004. <http://www.cnaf.infn.it/~sergio/publications/Glue4LDAP.pdf> (last visit: 2005-02-05)
- [Atk04] Atkinson, M.. *An Introduction to Data Services & OGSA-DAI [online]*. egee Grid Summer School - Vico Equense, 27 July 2004. <http://www.dma.unina.it/~murli/GridSummerSchool2004/presentations/Atkinson27/GridSummerschool-Vico-Equense-MPA-20040727.pdf> (last visit: 2005-02-05)
- [Bro04] Brook, J., Fellows, D., Garwood, K., Goble, C. (2004). *Semantic matching of Grid resource descriptions*. 2nd EUROPEAN ACROSS GRIDS CONFERENCE, Nicosia, Cyprus, Jan. 28-30, 2004
- [Dat03] DataTAG Project Office. *DataTAG - Project Description [online]* <http://datatag.web.cern.ch/datatag/project.html> (last visit: 2005-02-07)
- [Epi04] Epistemics Ltd. *PCPack [online]* <http://www.epistemics.co.uk/Notes/55-0-0.htm> (last visit: 2005-02-07)
- [Erw01] Erwin, D., Snelling, D., *Unicore: A Grid Computing Environment* LNCS 2150, Euro-Par 2001, Springer, pp 825-834
- [Uni00] UNICORE Forum eV. *By-laws of UNICORE-Forum [online]* <http://www.unicore.org/forum/statute.htm> (last visit: 2005-02-07)
- [Uni05] UNICORE Forum eV. *UNICORE at SourceForge: Detailed AJO description [online]* <http://unicore.sourceforge.net/ajo.html> (last visit: 2005-02-07)