Semantic Technologies in the SIMDAT Grid Project

Summary of activities

Jürgen Angele*, Moritz Weiten*

* ontoprise GmbH
Karlsruhe, Germany
{angele, weiten} AT ontoprise.de
http://www.ontoprise.de

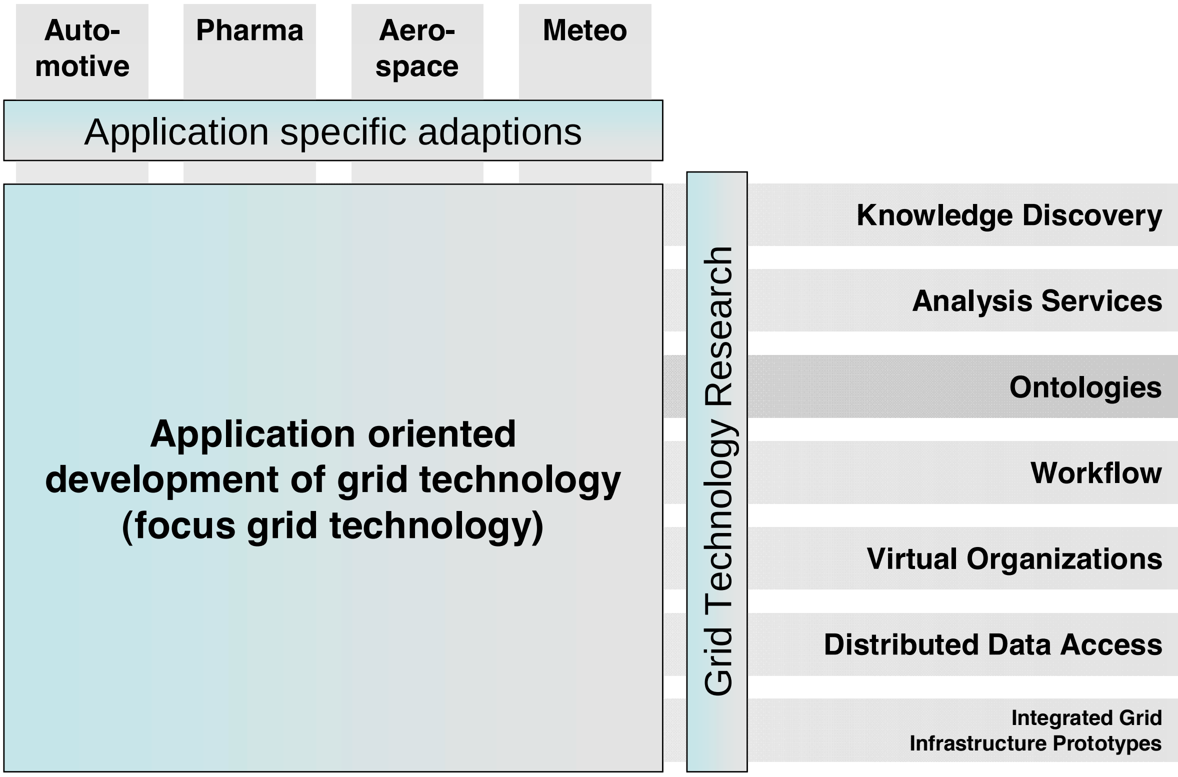

Figure 1: Structure of SIMDAT

The SIMDAT project aims at developing generic grid technology for the solution of complex application problems and using this new technology in several industrial application sectors. Figure 1 illustrates the structure of the SIMDAT project. The structure indicates the technology research activities in various areas as well as the application areas that the development is focused on. The grid technology research activity involves the collaboration of different technology providers and research institution across the technology topics and disciplines. Semantic technologies are expected to offer a significant added value to other technologies with respect to the management of resources on the process level and on the data level.

The strategic objectives of SIMDAT are

- to test and enhance data grid technology for product development and production process design,

Dagstuhl Seminar Proceedings 05271
Semantic Grid: The Convergence of Technologies
http://drops.dagstuhl.de/opus/volltexte/2005/382

- to develop federated versions of problem-solving environments (PSEs) by leveraging enhanced grid services,
- to exploit data grids as a basis for distributed knowledge discovery,
- to promote de facto standards for these enhanced grid technologies across a range of disciplines and sectors as well as
- to raise awareness for the advantages of data grids in important industrial sectors.

Semantic Technologies

As SIMDAT as a project is largely driven by industrial use cases, the application and further development of Semantic Technologies is based on a "bottom-up" approach. In a first step, the potential of different approaches (e.g. service discovery) is evaluated. Then the domain specific requirements are gathered and analysed. This leads to a scenario for a "demonstrator" showing the capabilities of semantic technologies in a distributed environment.
At the current stage of SIMDAT demonstrators for different application areas and different technological approaches are available (see below). The next steps target a more generalized architecture allowing to transfer solutions between application areas with similar requirements and thus exploiting synergy effects. This includes for example the integration of security features.

Regarding semantic technologies, SIMDAT currently has its focus on two basic aspects:

- the integration of distributed heterogeneous data sources;
- the semantic service discovery based on abstract service profiles.

Semantic Integration

The current approaches for the integration of distributed heterogeneous sources are based on the requirements defined within the automotive area. The goal is to provide a base for the semantic interoperability of different databases, dealing with data from real crash tests and simulated crash tests. The databases are hosted in different departments and heterogeneous in structure, even though there are semantic equivalences since both involve the management of crash test data.

In order to support the analysis and comparison of data sets from both worlds, an ontology-based integration concept has been developed and implemented in form of a demonstrator. It is based on a mediation scenario, where the internal data model of a CAE tool for the management of simulated crash test data ("CAE data") and a database for data of real crash tests ("CAT data") are combined. The goal is to provide the CAT data from the "view-point" of CAE engineers, i.e. in the model of the CAE tool used. This includes a technical integration (loose coupling via web-services), a syntactical integration (translation of queries between proprietary CAE language, semantic query language and SQL) and a semantic integration. The latter is realized via an articulating ontology.

The demonstrator has been implemented based on web-service technology, some special interfaces for the client PSE and ontoprise's core products - namely the OntoBroker and OntoStudio. The latter has been used to create a great part of the mappings between the models involved (with the help of a plugin developed in another research project). The OntoBroker serves as main component of the "semantic layer" and processes the declared mappings.

The approach chosen allows to resolve typical schema conflicts in declarative and flexible way. As part of the next steps include the development of a more generalized integration framework using existing features such as the connector architecture of OntoBroker. This activity will involve other application areas, mainly the Aerospace sector.

Semantic Service Discovery

The problem of service discovery was first of all targeted in the Pharma section. The goal was to provide flexible, dynamic descriptions of bio-informatic services in order to support service discovery. In order to achieve this goal, a concept for service discovery based on ontologies has been developed. It uses a compact but extensible design, based on rules for service profiles classifying different properties (e.g. data sources that can be accessed, type of analysis performed by service). While the current version of the concept uses parts of OWL-S, a systematic evaluation of frameworks/standards (including WSMO) will be performed based on the results of the first demonstration activities. Based on the concept mentioned an architecture for a semantic broker has been developed.

The semantic service broker will be capable to process queries specifying an abstract service profile in order to retrieve suited services. The architecture is based on the semantic description of service properties, containing information about the service characteristics, such as the type of analysis offered by a particular service, the databases that are accessed, etc. It will support domain experts in finding services that support a particular task (e.g. a sequence analysis on a set of data sources). This will be possible on different levels of abstraction, from very specific to very generic constraints.