Towards the Wisdom Grid: Goals and Architecture*

Ivan Janciak[1], Peter Brezany[1], A. Min Tjoa[2]

[1] Institute for Software Science
University of Vienna, Liechtensteinstrasse 22, A-1090 Vienna, Austria
{janciak|brezany}@par.univie.ac.at
[2] Institute for Software Technology and Multimedia Systems
Vienna University of Technology, Vienna, Austria
tjoa@ifs.tuwien.ac.at

Abstract. Multi agent systems, Grid technology, Semantic Web, and Web Intelligence paradigm are three modern approaches in information technologies, which we link in the research effort described in this paper to create a new-generation infrastructure called the *Wisdom Grid* with the mission to maintain, share, discover, and expand knowledge in geographically distributed environments. The paper introduces motivating ideas for this project, proposes the system architecture of one instance of the *Wisdom Grid*, and describes its functionality by means of a case study of one medical application.

1 Introduction

1.1 Motivation

The Web has significant impacts on both academic and ordinary daily life. It revolutionizes the way in which information is gathered, stored, processed, presented, shared, and used. Moreover, the Web provides the infrastructure for the Grid an emerging platform to support on-demand "virtual organizations" for coordinated resource sharing and problem solving on a global scale [15]. The Grid is sometimes heralded as the next generation of the Internet or the Web. There are strong connections between Grid, the Internet, and Web developments, as will be discussed later in this section.

Early Grid efforts (early to mid 1990s) started as projects to link supercomputing sites; at this time this approach was known as *metacomputing*. The objective was to provide Web computational resources to a range of high performance applications. Today the Grid infrastructure is capable of binding together more than just a few specialized supercomputing centers. It is more ubiquitous and can support diverse applications requiring large-scale computation and data.

Grid technologies are currently evolving toward an Open Grid Services Architecture (OGSA) [9], in which the Grid provides an extensible set of services that virtual

* The work described in this paper is being carried out as part of the research projects "Aurora" and "Advanced Data Analysis on Computational Grids" supported by the Austrian Research Foundation.

organizations can aggregate in various ways. OGSA development is a natural extension of the activities defining the standards for Web Services.

As already mentioned, Grid computing began with an emphasis on compute intensive tasks which, while benefiting from massive parallelism for their computation needs were not data intensive; the data that they manipulate was not proportional to the computation performed. Later, this focus shifted to more data-intensive applications [6], where significant processing was done on very large amounts of data and recently several research projects also addressed knowledge discovery in large databases attached to the Grid [4, 1, 5].

Meanwhile, "*a new generation of Web technology, called the Semantic Web, has been designed to improve communications between people using different terminologies, to extend to interoperability of databases, to provide tools for interacting with multimedia collections, and to provide new mechanisms for the support of "agent-based" computing in which people and machines work more interactively.*" [3]. These ideas led Grid scientists to the notion of *Semantic Grid*, where they plan to apply Semantic Web technologies in Grid computing developments [11].

Web Intelligence (WI) is a new direction for scientific research and development that explores the fundamental roles as well as practical impacts of Artificial Intelligence (AI) and advanced Information Technology (IT) on the next generation of Web-empowered products, systems, services, and activities [17]. Our vision is that a similar research direction, correlated with WI research, is also needed in the Grid research domain. Therefore, this paper introduces an analogous paradigm, *Grid Intelligence*, as a basis for developing a new-generation information technology infrastructure, the **Wisdom Grid** (WG), which will allow the creation of Grid applications that will help people achieve better ways of living, performing scientific work, treating patients, working, learning, etc.

1.2 Motivating Scenario

We believe that the best way of grounding the subsequent discussion on the Wisdom Grid architecture, its design goals, functionality, and the techniques used is in terms of a scenario. To this end, we will use the following scenario, which is derived from discussions with health care professionals involved in treatment of patients with traumatic brain injuries (TBIs) [4].

A physician has a patient with a specific diagnosis and wants to know what should be done next in the treatment of this patient, what the possible risks are, what kind of drugs should be used, what outcome for this and similar patients can be assumed, etc. The physician (user) has several options to determine answers to these questions. Traditionally, he could discuss these issues with his colleagues, with a specialist via on-line chat, he could look into a patient database to see what developments are available from other patients with the same or similar symptoms and diagnosis, he could search the Internet for appropriate medical articles and reports on this problem, etc. However, in this way it is not possible to obtain urgent information, which may often be a very critical issue in many medical applications, like TBI, or in crisis management environments in general.

To solve the above problem, an advanced information technology support which fulfills a set of requirements is needed. These requirements include: (1) ability to access and analyze a huge amount of information that is typically heterogeneous and geographically distributed; (2) intelligent behavior - ability to maintain, discover, extend, present, and communicate knowledge; (3) high performance (real-time or nearly real-time) query processing; and (4) high security guarantee. This combination of requirements results in complex and stringent demands that, until recently, could not be satisfied by any existing computational and data management infrastructure.

In tackling these problems, we have initiated a research effort to design and experimentally implement a novel infrastructure called *Wisdom Grid* (WG), which aims to fulfill the requirements outlined above. The WG concepts are mainly based on the following technologies: basic Grid [10], Data Grid [6], Knowledge Grid [2], Semantic Web and Grid [3, 16], and agents [14].

Now a new information and knowledge acquisition style based on the cooperation of the physician (user) with the WG components can be briefly described by the following scenario.

We assume that the user has his own *Personal Agent* (P-Agent), which could be a specialized medical agent having detailed information about the user, his professional field, department specialization etc., or it could be a very simple agent that is only able to ask questions and provide the answers. The user sends such an agent, together with information about the problem, to our *Domain Agent* (D-Agent). These agents can only communicate if they understand each other, so first they negotiate about the communication language and the domain in which they are going to communicate. D-Agent informs P-Agent about the domain of interest it is able to answer, and P-Agent passes this information to the user. The user then creates a particular question and fills it in with real input data. When D-Agent receives this question from P-Agent, it tries to find the answer in its *knowledge base*. There can be an exact answer or just a reference to the resource (location) where the desired information is stored and how to retrieve this information from the resource. If the information is somewhere on the Grid (e.g. in a special medical database) then a specific Grid service, e.g. data mining service, is queried for this information.

A special *Knowledge Explorer Agent* (E-Agent) is submitted to ask another agent, for example specialized medical agent, who has required information or is able to search for it. The answer returned to P-Agent could also be a location of resources on the web or in a medical library where the user can find information about the problem. In addition, it could be an email address or a phone number of a specialist who has appropriate knowledge about the user's request. Finally, P-Agent receives complex information about the user's problem and it is up the P-Agent to represent this information to the user.

1.3 Organization of the Paper

The structure of the rest of the paper is organized as follows: Section 2 presents the architecture and functionality of the Wisdom Grid. The knowledge base organization is introduced in Section 2.3 and we briefly conclude in Section 3.

2 The System Design

In this section we describe the system architecture, outline the functionality of the components involved, and describe interactions between individual components.

Architecture. The architecture of our Wisdom Grid system is sketched out in Fig 1. The agents provide distributed intelligence services, which involve communication and decision making activities. The Grid is the basic infrastructure, which provides secure access to distributed data and knowledge resources.

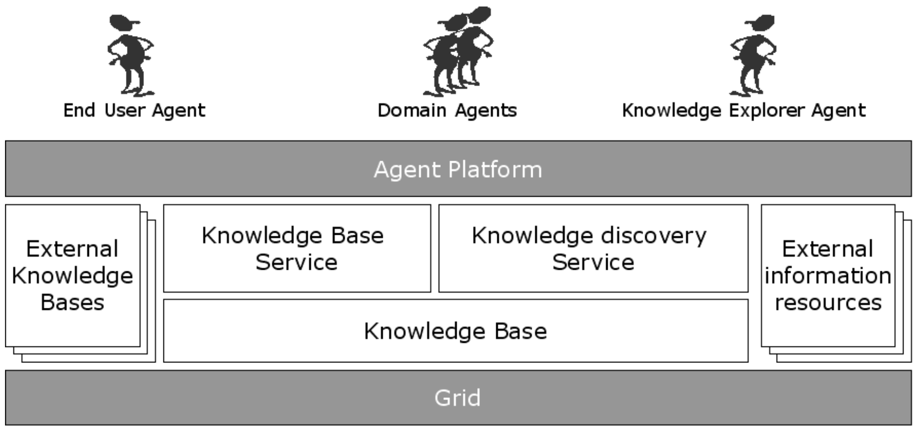

Fig. 1. System Architecture

WG services can be geographically distributed and can use all advantages of grid computing. Taken actions in knowledge discovery are invoked and controlled by internal agents which communicate with services and transport data between them. This agent manages and executes services in ways described in internal workflow scenarios.

2.1 Agent Platform

We decided to use agent technology [14] for our communication purposes. Agents build the communication aspect of our system and represent it to the outside environment. This part of the *Wisdom Grid* plays the main role in the interaction of agents and WG services. Agents' messages are transformed to exact actions and results of these actions are converted back to messages. This platform closely cooperates with the knowledge base service, which is always queried for information about all the services and resources that can be used in the search for knowledge. FIPA Standards [8] are used for the agent life cycle and two types of agents are managed by this platform:

a) **Domain Agent.** *(D-Agent)* These agents represent the system to the outside world and they are mediators between knowledge demands and knowledge provided by WG. From the outside world point of view, they are intelligent software agents specialized for a particular domain of knowledge stored in the knowledge base. This part of the knowledge base is used as agent's ontology with which it presets itself to other agents submitted to query WG for knowledge. Those personal agents (P-Agents) are able to address questions to the D-Agents and present answers to the user. The success of this activity strongly depends on the mutual understanding of the P-Agent and D-Agent and also on the query construction technique. P-Agent is not part of the system because it is created by the user and is not registered on our agent platform. It can ask the D-Agent for the list of domains, about which it has knowledge, or P-Agent can inform the D-Agent about the ontology, which describes its domain and merges it together with D-Agent's ontology to select one common domain.

b) **Knowledge Explorer Agent.** *(E-Agent)* This agent is used by the WG in the process of searching for other knowledge providers and extends the knowledge base with information about services or resources which can be later queried by the Knowledge discovery service. The agent is able to query other agents and also to search the Semantic Web [3], which can be considered as an external knowledge base.

2.2 Knowledge Discovery Service

This service acts as a client of external information resources or uses prepared clients or agents in communication with them. It provides external services with input data and returns their output. The knowledge base service supports the knowledge discovery service with information about resources. Such information usually consists of location, type of resource, communication protocols, and everything needed for successful cooperation between them. Semantic markup for web services (DAML-S) [7] is used for this purpose. Knowledge discovery service uses instances of the DAML-S ontology to identify their profile, model and grounding. Resource or output data from it can be used as an input parameter for another external service. Therefore workflows from the knowledge base are used to fulfill this requirement. For example, references to databases integrated into the Grid or other databases provided by cooperating institutions (hospitals, research institutes, etc.) can be used as an input for external services.

External information resources. As external information resources we consider resources, which provide WG with some kind of useful knowledge. It could be a Web/Grid service, web page, agent, database, online library, etc. We mainly focus on Web/Grid services able to perform knowledge discovery (advanced data analysis like data mining, Online Analytical Processing (OLAP), etc. [12]). Locations and descriptions of external services are stored in the knowledge base.

2.3 Knowledge Base Service

This service operates over the knowledge base. Its function is to parse, search, query and extend knowledge base. It supports the Knowledge Discovery service with data about information resources. It uses a reasoner to determine which resource meets the query restrictions. The DAML Query Language (formal language and protocol for querying repositories) is used for this purpose. The E-Agent extends this service with the ability to search in external, geographically distributed knowledge bases and also in the Semantic Web, often presented as global knowledge base.

Knowledge Base *is a database that stores particular data about real objects and relations between these objects and their properties in particular domains*. The knowledge base mainly stores data about information resources; what they do, how they work, how to query them and what knowledge they produce. The knowledge base is divided into individual domains of knowledge provided by WG. Every domain is associated with a D-Agent. This agent can operate over more then one domain. Ontologies are sets of associated slots of classes, their properties, relations, and restrictions that describe individual domains. It means that the ontology provides exact consequence to the objects stored in the knowledge base. The ontology is written in the DAML+OIL [13] language and for better search and query performance it is stored in the database. Ontologies in the knowledge base are used to describe semantics of the information resources. For example a real database of patients in a hospital is an instance of a database ontology, and on the other hand, it is also an instance of the ontology defining patient class and his properties. This fact that one object can be an instance of more ontologies is useful in the process of mapping between real names and their semantic. It can be used, for example, in the process of integrating different databases (different in structure, names, location etc., but having the same semantics).

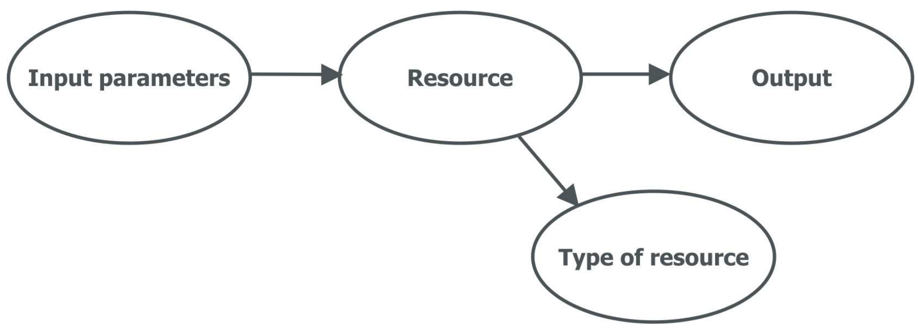

Fig. 2. Top ontology of the information resource

The knowledge base can be extended manually by a domain expert, a person who understands the context of a particular domain, or automatically by the knowledge base service which takes into account knowledge founded by the Knowledge discovery service or by the E-Agent. Such information is usually a location of a knowledge resource and its description. The knowledge base also stores semantic descriptions of data used as an input and output parameters for services and information about quality of service. From this point of view the knowledge base can simply be considered as a semantic registry of the services and other information resources.

The top level information resource ontology is depicted in Fig. 2. This high level abstraction simply says that the resource need input parameters and gives some output. It can be used for simple dynamic web pages, databases as well as for Grid services. For example, the service at *http://gridminer/decision-tree.gwsdl* with input parameters *jdbc://hospital/db*, *user=test*, *table=test*, and *class=outcome* gives a decision tree as an output located at *http://gridminer/decision-tree/tree.xml*, or another example: the resource *jdbc://hospital/db* returns a list of patients if input is query: *"SELECT patients FROM test"*. Detailed descriptions of resources and properties are specified in the DAML+OIL language and for every type of resource, specialized templates exist to express resource requirements and properties in some goal achievement process.

The knowledge base also contains internal information about WG services, their detailed description, and their functionality and ability in processing information flows. For this purpose workflows diagrams, also stored in the knowledge base and defined by specialized ontology, are used.

External Knowledge Bases. For external knowledge bases, we consider ontologies available on the web and freely accessible. We assume that the advent of the Semantic Web will bring the ontologies of many different domains as its components. Therefore, the E-Agent is proposed, which is able to search widely distributed knowledge bases.

3 Conclusions

This paper has outlined our vision of the Wisdom Grid as a future infrastructure for advanced applications. We have also presented an instance architecture of the Wisdom Grid, described its main components and the structure of the services as well as the concepts of the Wisdom Grid knowledge base in the context of a medical application example. The prototype implementation of the knowledge discovery services is currently in progress. The future work is to create WG portal to simplify cooperation between the real user and his agent through web browser.

References

1. Cannataro, M., Talia, D., Trunfio, P.: Design of distributed data mining applications on the KNOWLEDGE GRID. Proceedings NSF Workshop on Next Generation Data Mining Baltimore (November 2002) 191–195
2. Cannataro, M., Talia, D.: KNOWLEDGE GRID: An Architecture for Distributed Knowledge Discovery. Communications of the ACM (January 2003) 89–93
3. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (May 2001)
4. Brezany, P., et al.: Knowledge grid support for treatment of traumatic brain injury victims. Proceedings of the Conference ICCSA 2003 Montreal (May 2003)
5. Cannataro, M., Talia, D., Trunfio, P.: Distributed data mining on the grid. *Future Generation Computer Systems.* (2002) 1101–1112

6. Chervenak, A., Foster, I., Kesselman, C., Salisbury, C., Tuecke, S.: The data grid: Towards an architecture for the distributed management and analysis of large scientific datasets. (1999)
7. The DAML Services Coalition: DAML-S: Semantic markup for web services. http://www.daml.org/services (2002)
8. Foundation for Intelligent Physical Agents: Fipa - agent management specification. http://www.fipa.org (2000)
9. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The physiology of the grid: An open grid services architecture for distributed systems integration. (2002)
10. Foster, I., Kesselman, C., Tuecke, S.: The anatomy of the Grid: Enabling scalable virtual organizations. Intl. J. Supercomputer Applications 15(3) (2001)
11. Goble, C., De Roure, D.: The semantic grid: A future e-science infrastructure. www.semanticgrid.org
12. Han, J.: Data Mining. Concepts and Techniques. Morgan Kaufmann (2000)
13. Horrocks, I.: DAML+OIL: a reason-able Web ontology language. in Proceedings of EDBT (March 2002)
14. Jennings, N., R., Wooldridge, M., J.: Agent technology : Foundations, Applications, and Markets. Springer-Verlag (1998)
15. De Roure, D., Baker, M., A., Jennings, N., R., Shadbolt, N., R.: The evolution of the grid. In F. Berman, A. J. G. Hey, and G. Fox, editors, *Grid Computing: Making The Global Infrastructure a Reality*, John Wiley & Sons (2003) 65–100
16. De Roure, D., Jennings, N., R., Shadbolt, N., R.: The semantic grid: A future e-science infrastructure. In F. Berman, A. J. G. Hey, and G. Fox, editors, *Grid Computing: Making The Global Infrastructure a Reality*, John Wiley & Sons (2003) 437–470
17. Zhong, N., Liu, J., Yao, Y., (eds.): *Web Intelligence*. Springer-Verlag (2003)