Ontology-based Universal Knowledge Grid: Enabling Knowledge Discovery and Integration on the Grid

Yuhua Li, Zhengding Lu
College of Computer Science, Huazhong University of Sci. & Tech. , Wuhan, 430074,China
Yuhua_yy@163.com

Abstract

This paper proposes an ontology-based grid architecture model in terms of the Universal Knowledge Grid (UKG) for building large-scale distributed knowledge system on the grid. UKG emphasizes geographically distributed high-performance knowledge discovery applications and knowledge integration services. Five core components have been identified: an intelligent composer for user interface, an ontology server providing data integration ontology services, data mining ontology services, knowledge integration ontology services, a metadata directory server maintaining the metadata describing all the data, tools and knowledge, metadata database (db) stores metadata and a knowledge base as container for knowledge. An application example on Foreign Exchange Management (FEM) is also presented.

1. Introduction

In many scientific and business areas, massive data collections need to be analyzed. Moreover, in several cases data sets must be shared by large communities of users that pool their resources from different sites of a single organization or from a large number of institutions. Grid computing has been proposed as a novel computational model, distinguished from conventional distributed computing by its focus on large-scale resource sharing, innovative applications, and high-performance orientation. Today Grids can be used as effective infrastructures for distributed high-performance computing and data processing [1]. The Open Grid Services Architecture (OGSA) introduces service-orientated grid, and leverages the results of Web Services. OGSA is now emerging as a convenient model for building more integrable and interoperable Grid environments. The Globus Toolkit itself is currently being re-implemented under the OGSA model [2].

An ontology is an explicit specification of a conceptualization where definitions associate concepts, taxonomies, and relationships with human-readable text and formal, machine-readable axioms [3].

Ontology are used for [4]:

1. Communication between implemented computational systems, between humans or between humans and implemented computational systems;
2. Computational inference, e.g. for internally representing and manipulating plans and for analyzing the internal structures, algorithms, inputs and outputs of implemented systems in theoretical and conceptual terms;
3. reuse and organization of knowledge, e.g. for structuring and organizing libraries of plan and domain information.

As a motivating example, consider Foreign Exchange Management (FEM) domain. In FEM, the data involve the Foreign Exchange information from Finance Organization, Foreign Exchange Management Departments, Custom, Economic and Trade Departments and the money laundering and other finance crime case information from law-executed departments. Data distribution varies great: single data source, federated data sources with horizontal partitioning that storing parts of data of equal structure and semantics for a given problem domain in different sites, vertical partitioning that part of attributes are distributed over multiple sites, heterogeneous data sources schemes that is heterogeneous in structure and design. Some knowledge is law, statute and domain expert experiences, some money laundering knowledge should be acquired through data mining in large scale of data and integrated dynamically. So one kind of architecture is needed to facilitate database resource integration, data mining, and knowledge sharing and knowledge integration within FEM virtual communities.

This paper proposes an Ontology-based generic grid architecture model in terms of the Universal knowledge grid (UKG) that enables knowledge discovery and knowledge integration on the grid.

The rest of this paper is organized as follows. Section 2 describes the universal Knowledge Grid general architecture and the features of the main components.

Proceedings of the 2004 IEEE International Conference on Services Computing (SCC'04)
0-7695-2225-4/04 $ 20.00 IEEE

Section 3 gives an application example. Section 4 discusses related work and Section 5 concludes the paper.

2. The UKG architecture

The UKG architecture (figure. 1) is defined on top of Grid toolkits and services.

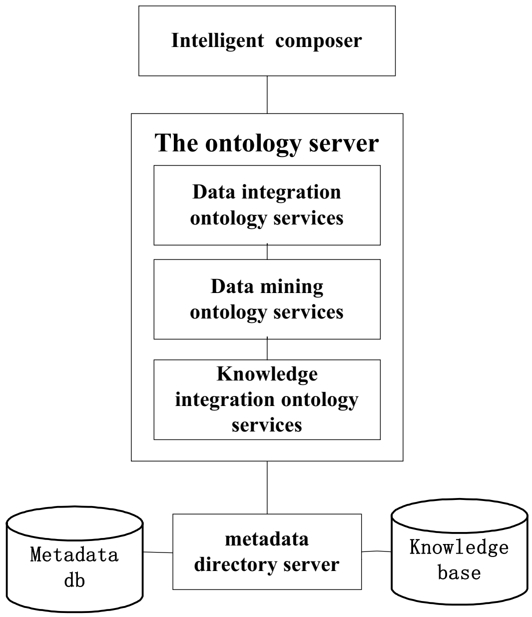

Figure 1. The architecture of the UKG

2.1 The intelligent composer

Current web browsers only know how to interpret the HTML tags and present it as a plain text document. We propose the intelligent composer as the user interface .The intelligent composer offers a set of graphical tool. The user can form a graphic representation of her/his knowledge requirements using common visual facilities or natural language. It displays the machine understandable semantics and explains the display. User can operate resources according to semantics.

1. Automatic suggestion service. The intelligent composer uses synonym and taxonomy to guide the user automatically to suggested search items. For an example , if one want to search "Grid" related documents, the taxonomy helps to suggest the "Data Grid" and "Knowledge Grid" related documents .
2. Advanced navigation. Using of ontologies provides the advanced navigation. The user gets easy access to relevant information by browsing through the modeled concepts and their relations.
3. Reasoning ability. Embedded inference engine helps users to retrieve and present implicit knowledge.

2.2 The ontology server

The ontology server is the center module. It is responsible for the management and query of explicit, declaratively represented ontologies. An ontology server offers some or all of the following services, data integration ontology services, data mining ontology services, knowledge integration ontology services.

2.2.1 Data integration ontology services

Data integration ontology services describes the semantic of web documents, bridges the gap between semi-structured and structured databases, facilitating data cleaning and data preparation, heterogeneous data sources integration.

2.2.2 Data mining ontology services

Data Mining Ontology Services offer a reference model for the different kind of data mining functionalities, algorithms and software available to solve a specified problem. The data mining ontology gives functionality class including clustering, classification, prediction, text mining, association mining, outlier analysis, link analysis, WEB mining etc. Figure 2 shows part of algorithm class hierarchy for the data mining ontology.

2.2.3 Knowledge integration ontology services

Knowledge integration ontology services include:

1. Maintaining a set of public available ontologies;
2. Responding to query for relationship among terms;
3. Facilitating the communication and knowledge interchange among different knowledge bases;
4. Translating and mapping between different ontologies;

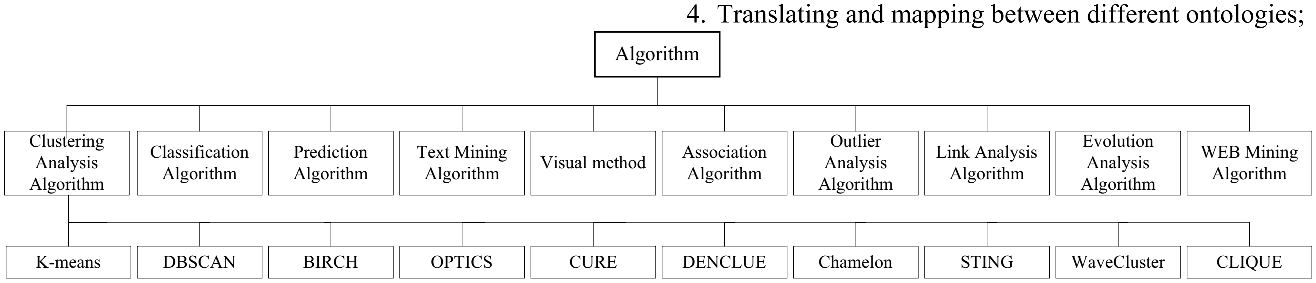

Figure 2. Algorithm class hierarchy for the data mining ontology

5. Scoring the pattern of data mining ontology and append the effective knowledge to knowledge bases;
6. Providing on-demand services to support problem and decision making;
7. Integrating knowledge resources of different levels (for example, concepts, axioms, rules and methods) to support problem analyzing and solving.

The ontology server keeps human users from directly interacting with the physical data resources, so they don't need to be concerned with the resources' form and location. It can also isolate any changes in the data resource. Ontology is built and maintained by the ontology server.

The process of building ontology is given below: (1) identify purpose and scope of target domain; (2) capture the key concepts and relationships; (3) code the ontologies; (4) integrate existing ontologies (eliminate redundancy and check consistency).

2.3 Metadata directory server

This service deals with maintaining the metadata describing all the data, tools and knowledge used in the Universal Knowledge Grid.

The metadata information are represented by XML (eXtensible Markup Language) documents and are stored in Metadata db and Knowledge base.

The metadata directory server is the meta-catalog of UKB. It can identify specific discovery and registration services, also provide the mechanism for integrating multiple knowledge bases into high-order knowledge base as an integral unit to serve a domain-specific functional purpose .In addressing these requirements, The metadata directory server will provide a rich set of behaviors for managing Knowledge base collections. The following are candidates:

1. Coordinated selection. A user can select knowledge of knowledge base from one or more directory servers simultaneously;
2. Knowledge propagation. A user or a virtual role can propagate knowledge multiple Metadata directory server easily through intelligent agent;
3. Knowledge management .A knowledge flow network [10] help to achieve effective knowledge sharing in dynamic virtual organization.
4. Integrated Knowledge Base. A subset of knowledge base within one or more directory servers can be associated to compose high-level integrated knowledge bases.

2.4 Metadata db

Metadata db stores metadata of heterogeneous data sources, tools and algorithms used to data integration and data mining.

2.5 knowledge base

Knowledge base stores knowledge obtained as result of the data mining process, i.e. learned models and discovered patterns and the knowledge gathered from application domain.

3 An application example

In the following, we discuss an example of finding a model of money laundering in FEM domain, useful to show how the problem solving can benefit from the UKG services.

In the scenario of the example, the UKB run through the following step (Figure 3):

1. User selection of the resources used in his problem and task creating using the intelligent composer;
2. The task is send to the ontology server;

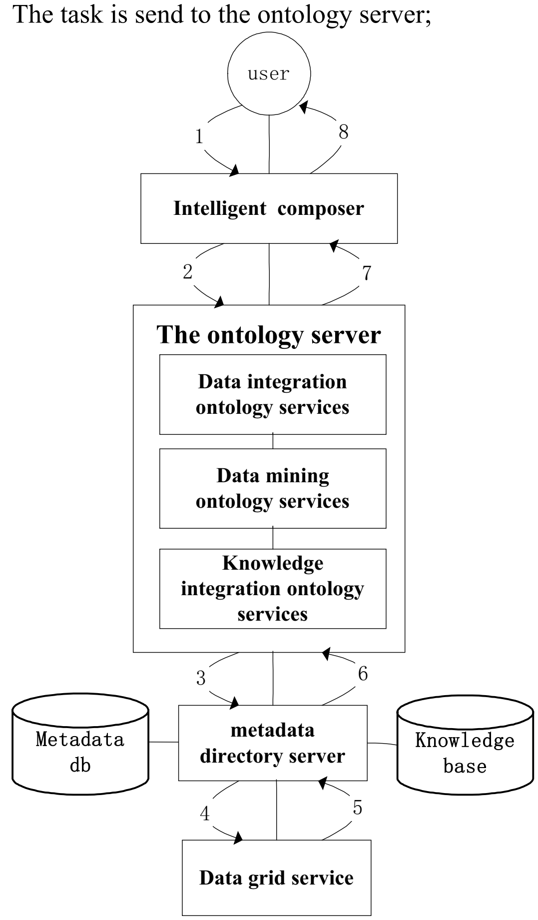

Figure 3. An application example of UKB

3. The ontology server explain the semantics of the task and transfer to the metadata directory server;

4. The metadata directory server locates the distributed crime case databases, BIRCH clustering algorithm of the task on the grid nodes and trans it to data grid service;
5. The data grid service executes the data mining task, the result is returned to the metadata directory server;
6. The effective mode (knowledge) is integrated in knowledge base through knowledge integration service;
7. The effective mode (knowledge) is browsed through the intelligent composer.
8. The user gets the problem solving.

4 Related work

There are a number of grid-inspired efforts recently that involve knowledge discovery and sharing.

– Semantic Grid

GGF has initialed a research group with the term of Semantic Grid [6]. However, their work is very similar to the semantic web service [5] effort. They focus on the semantic and ontology issues of grid service description and concern with the scheduling process among multiple grid services. They aren't known for Knowledge discovery services.

– Knowledge Grid

The knowledge grid offered by Knowledge Grid Lab in Italy used to perform data mining on very large data sets available over grids to make scientific discoveries, improve industrial processes and organization models, and uncover business valuable information [7]. It is ignorant of developing high-level knowledge integration services.

– VEGA Knowledge Grid

The VEGA Knowledge Grid effort in Chinese Academy of Sciences [8] has also proposed a worldwide resource, sharing and management platform. However, we argue that their platform has no awareness of developing high-level knowledge services and data mining services.

– TCM KB-Grid

TCM KB-Grid effort [9] is an in-progress work initialed by the Grid Computing Lab of Zhejiang University and TCM Information Research Institute of China Academy of Traditional Chinese Medicine. KB-Grid main enables Knowledge Sharing on the Semantic Web. It is not reported that TCM KB-Grid has data mining service.

Compared with them, UKG aims much more on data integration, knowledge management, discovery, and integration on the grid.

5 Summary and future work

The use of ontology to describe grid resources will simplify and structure the systematic building of grid applications through the composition and reuse of software components. We propose an ontology-based grid architecture model UKG for building large-scale distributed knowledge system on the grid. UKG emphasizes geographically distributed high-performance knowledge discovery applications and knowledge integration services. An application example on FEM is also presented. Future work is to finish the development of architecture model and rich data mining algorithm tools.

6 Acknowledgements.

This work is supported in part by National Foreign Exchange Management Information & Decision Support System subprogram of key technologies R&D program in finance information of the Tenth Five-year Plan of China Department of Science and Technology, under Contract 2001BA102A06-11.

Reference

[1] I. Foster, C. Kesselman, S. Tuecke. "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", *Intl. J. SupercomputerApplications*, 15(3), 2001.

[2] The Globus Project. Globus Toolkit 3.0 Alpha 3 Release, http://www.globus.org/ogsa/releases/alpha/.

[4] Ontology (Special Issue on), *CACM, Vol. 45,N.2*, 2002.

[5] Sheila A. Mcllraith, Tran Cao Son, and Honglei Zeng, " Semantic Web Service", *IEEE intelligent systems. 2001, Vol. 16, No. 2*, pp. 46-53.

[6]Global Grid Forum Semantic Grid Research Group: http://www.semanticgrid.org/GGF/

[7] Antonio Congiusta, Andrea Pugliese, Domenico Talia and Paolo Trunfio, "Designing Grid services for distributed knowledge discovery", *Web Intelligence and Agent Systems: An international journal 1*, 2003, pp. 91–104.

[8] H.Zhuge, " A Knowledge Grid Model and Platform for Global Knowledge Sharing", *Expert Systems with Applications, vol.22, no.4,* 2002, pp.313-320.

[9] Huajun Chen , Zhaohui Wu , Jiefeng Xu, **"**KB-Grid: Enabling Knowledge Sharing on the Semantic Web"**,** *Proceedings of the International Workshop on Challenges of Large Applications in Distributed Environments (CLADE'03) on* , 21 June 2003 pp. 70 – 7.

[10] Hai Zhuge, " A Knowledge Flow Model for Peer to Peer Team Knowledge Sharing and Management", *Expert Systems with Applications, vol.23, no.1*, 2002, PP.23-30.