Developing next generation web as collaboration media

KOJI ZETTSU, YUTAKA KIDAWARA

National Institute of Information and Communications Technology, Japan
{zettsu, kidawara}@nict.go.jp

YASUSHI KIYOKI

Faculty of Environmental Information, Keio University, Japan
kiyoki@sfc.keio.ac.jp

Keywords: Global Knowledge Grid, service-oriented architecture, knowledge web

Beyond the semantic web era, digital contents are not the key assets of the web any more, but the knowledge acquired from the web contents plays an important role for users to understand situations, make decisions and/or take actions. Considering knowledge as a key asset, it is worthwhile to renovate the traditional content-centered framework of web lifecycle so as to be a knowledge-centered one. We introduce the Global Knowledge Grid, a distributed knowledge service computing environment based on the service-oriented architecture (SOA). It facilitates data-centric collaborations of different types of knowledge services: knowledge capture, knowledge association, and knowledge provision. An evolving network of knowledge is built by interconnecting heterogeneous knowledge sites over the global knowledge grid. The web browser is not restricted to hyper-document metaphor any more, but it facilitates adaptation of knowledge delivery by controlling contents and their screen layout for proactive navigation of user's information demands. Some typical applications are demonstrated to show how well our framework works.

1. Introduction

In today's networked society, knowledge-intensive work involves communication among communities of people and the social practices that occur in a particular context. Knowledge-intensive work engages in a large number of communication, coordination, and cooperation practices that cross the boundaries of organizations, countries, cultures, and/or disciplines. Beyond the semantic web era, digital contents are not, by themselves, the key assets of the web any more. Instead, the knowledge acquired from the digital contents plays an important role for users to know what is going on, make decisions, and take actions. Considering knowledge as a key asset, it is worthwhile to renovate the traditional content-centered framework of web lifecycle as a knowledge-centered one. Our intention is to change the nature of the web as an information-based resource to a knowledge-based resource.

The Global Knowledge Grid [1,2] is an integrated infrastructure for coordinating knowledge sharing and problem solving in distributed environments. The concept of knowledge grid was originally introduced as a parallel distributed knowledge discovery and data mining (PDKDD) environment, which uses the basic functions of a grid and defines a set of additional layers to implement the functions of distributed knowledge discovery [3]. We extend the original idea of knowledge grid with particular emphasis on the following concepts.

Harnessing collective intelligence

The web provides a platform for establishing networks made up of communities of people (or organizations or other social entities) connected by social relationships such as friendship, collaboration, or information exchange based on common interests. These web-supported social networks can be regarded as virtual communities.

From a sociological perspective, knowledge is considered to be socially constructed. Social processes influence the processes of generating and applying knowledge. As a consequence, knowledge can be described not as objective truth but as what a social system considers to be true. On the web, individuals are responsible for identifying knowledge sources according to their own demands on the basis of the collective intelligence [4] to be harnessed. The unit is shifting from documents to semantically coherent units of content representing, for instance, opinions, explanations, and interpretations.

A conceptualization provides a context in which knowledge elements can be uniformly organized on the basis of a specific common understanding. A conceptualization is never universally valid, but rather, it is valid for a limited number of users committing to that conceptualization. Web-supported social networks or virtual communities manage and analyze distributed knowledge. It becomes necessary to have a global platform for sharing and processing heterogeneous knowledge repositories across community boundaries. The "architecture of participation" becomes inclusive defaults for aggregating user data and building value as a side-effect of the ordinary use of the application. The mechanism that "users select for value" retains knowledge with a high-reuse rate but discards knowledge with a low reuse rate.

Virtual organizations on grid architecture

The web-based sharing that we are concerned with is not primarily web document exchange but rather direct access to data, software services, and other resources, as is required by a range of collaborative problem-solving and resource-brokering strategies emerging in industry, science, and engineering. Because of our focus on dynamic, cross-organizational sharing, Grid architecture complements, rather than competes with, existing web-based distributed environment. The real and specific problem that underlies the grid concept is coordinated resource sharing and problem solving in dynamic, multi-institutional *virtual organizations* (VO) [5].

This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules form a VO.

Service-oriented knowledge

Service orientation is widely acknowledged for its potential to revolutionize the world of computing by abstracting from all resources as services in a service-oriented architecture (SOA) [6]. The service-oriented knowledge utilities model [7] stems from the necessity of providing knowledge and processing capabilities to everybody, thus promoting the advent of a competitive knowledge-based economy. Here, the knowledge services are instantiated and assembled dynamically, so that the structure and location of the knowledge services are changing at run-time on the basis of the user needs and issues. In addition, it leads to the idea that a fundamental change will be seen in the world of knowledge base so that knowledge bases move away from the traditional relational and object models, toward the *associative model* of knowledge elements, such as graph representations and triple stores.

2. System overview

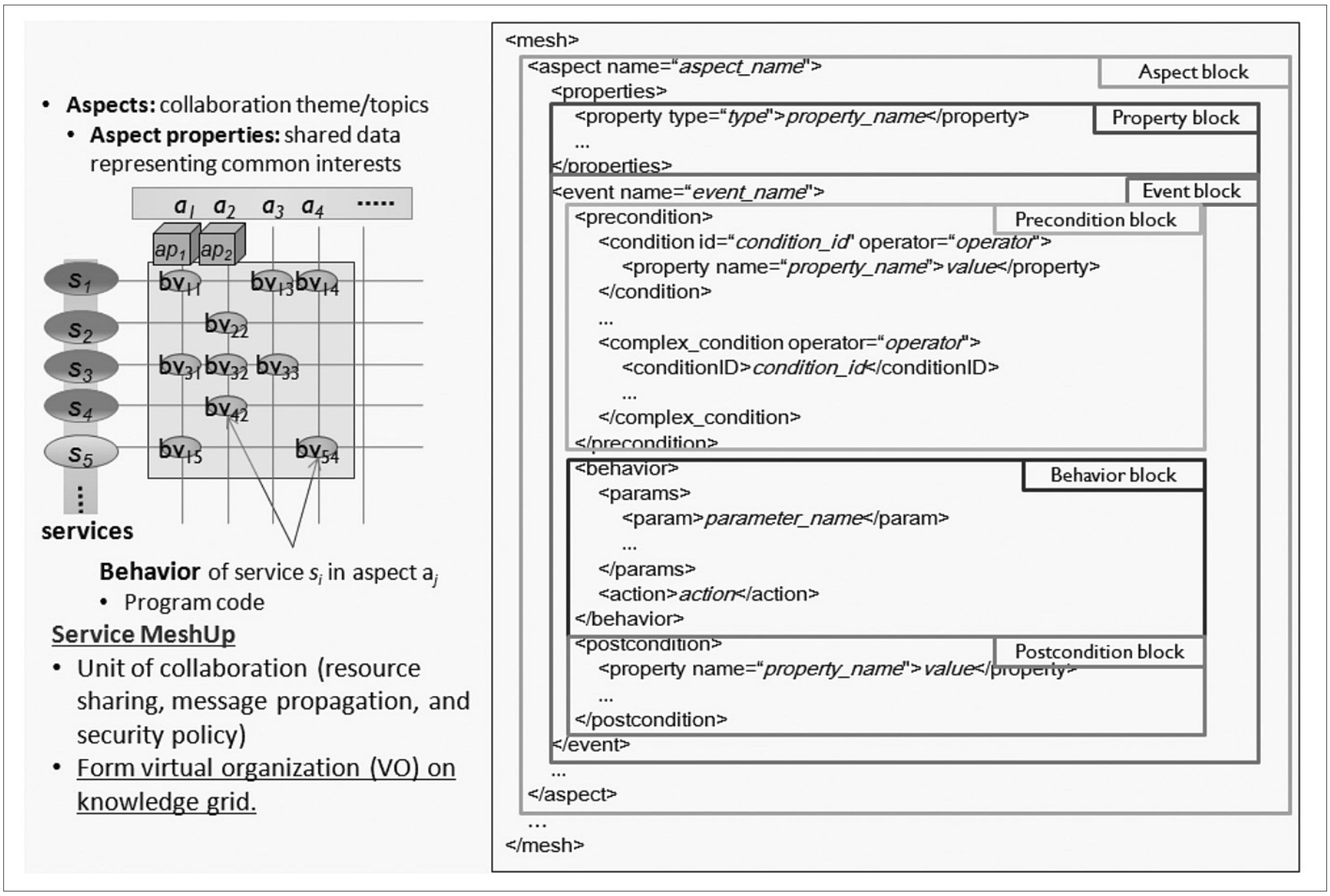
Knowledge services on the Global Knowledge Grid are classified into the following categories:

• **Knowledge Discovery Service:** It provides concept descriptions of an information source on the basis of the knowledge discovery approach by data mining, such as segmentation, classification, summarization, and ontological annotations.

• **Knowledge Association Service:** It produces various kinds of associations between multiple information sources on the basis of their concept descriptions provided by the corresponding knowledge discovery services.

• **Knowledge Delivery Service:** It is responsible for (visual) presentation and navigation of the results from knowledge discovery services and knowledge association services. It also handles user interfaces and interactions.

Figure 1. An example of Service MeshUp description



An application on the Global Knowledge Grid is realized by a specific collaboration of the abovementioned services. The Global Knowledge Grid provides the following service collaboration models:

• **Workflow model:** It designs a control flow of service execution. It is described using web service business process execution language (WS-BPEL) [8], a de facto standard for describing workflow-based service collaborations.

• **Data-centric model:** It designs the interests and behaviors of services with respect to shared data. It is described using *Service MeshUp description language*, our original description language for data-driven service collaborations.

In the rest of this section, we explain only the data-centric model. To realize the data-driven service collaborations, instead of the conventional service mashups we have proposed a novel approach called Service MeshUp.

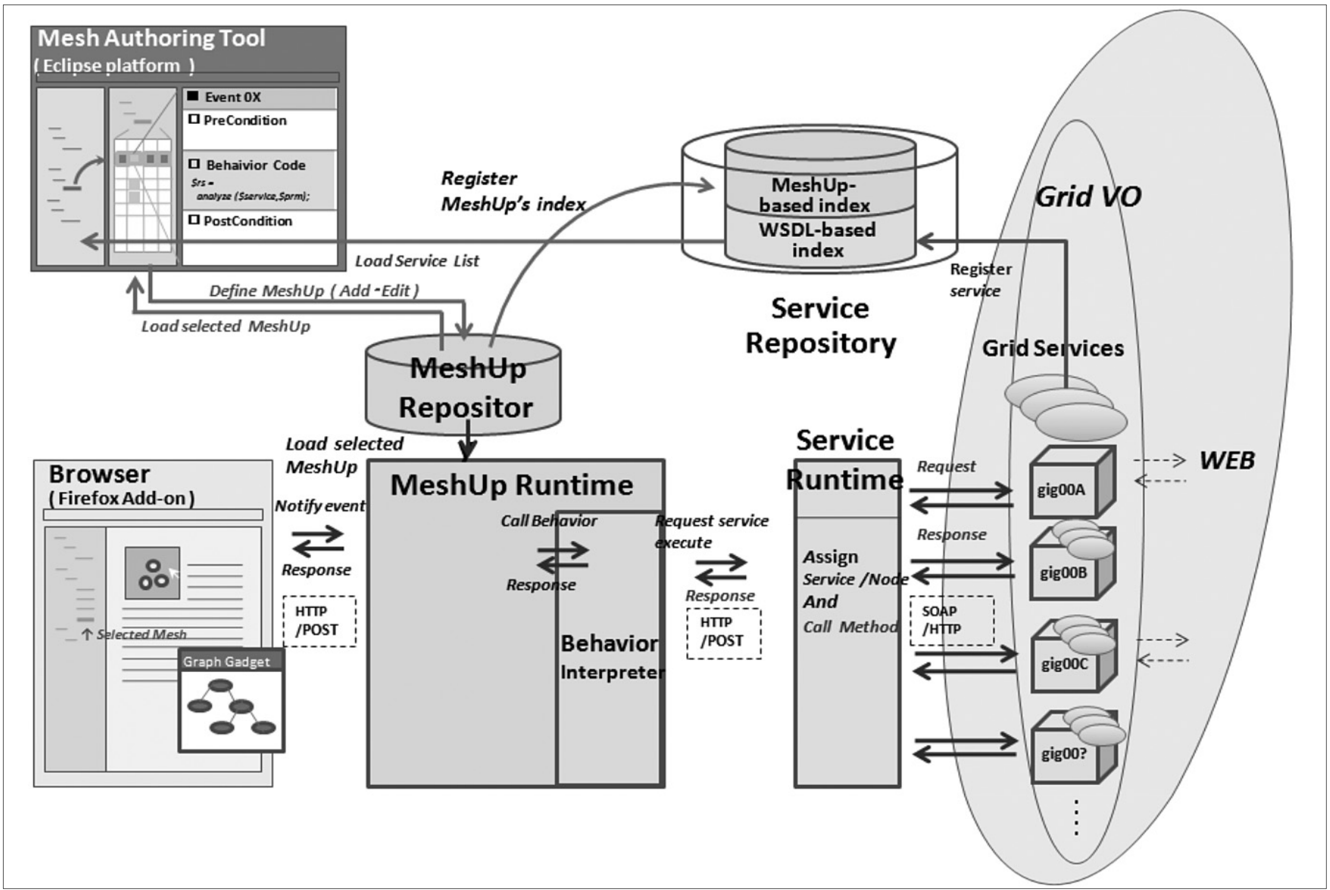
Fig. 1 shows an example of the *Service MeshUp* description. The Service MeshUp description consists of a set of *aspects*, each of which defines a single task or function in an application. Each aspect has its own properties *(aspect properties)* to be shared by the services. For each aspect, an application developer determines a set of services to be involved, and then, for each service, he/she defines *precondition, behavior,* and *postcondition* with respect to the aspect properties.

Here, the precondition describes the conditions for activating the service with respect to the aspect properties. For instance, the precondition becomes true when a specific aspect property is modified or has a specific value. Of course, Boolean composition of the conditions is also specifiable. The behavior defines what the service will do when it is activated. Basically, it invokes the functions of the corresponding service. The postcondition defines what will be done after the behavior. In most cases, the data obtained as the result of the behavior will be set to the aspect properties.

In contrast to conventional service mashup methods based on workflow model, our Service MeshUp aims at realizing service collaborations sharing "common interests," each of which is represented by the aspect properties. While both models are convertible with each other, our Service MeshUp model is more suitable for self-organization processes and ever-evolving processes such as building collective knowledge or monitoring situations than it is for transactional processes like business workflows.

The Global Knowledge Grid consists of three layers: grid layer, knowledge service layer, and application layer. In the grid layer, a computational grid network is constructed using de facto standard grid middleware Globus Toolkit [9]. In contrast with the conventional grid for building high-performance computer clusters, the Global Knowledge Grid nodes are distributed all over the world like traditional web servers. On individual grid

Figure 2. Component architecture of knowledge service layer



nodes, the knowledge services are developed independently and in parallel, like traditional web pages. The grid nodes are connected by an extensive secure network. The knowledge service layer is responsible for implementation, deployment, search, and collaboration of the knowledge services.

Fig. 2 illustrates in detail the knowledge service layer. The knowledge service layer consists of the following components:

• **Mesh runtime:** executes Service MeshUp descriptions to realize collaborations of the knowledge services.

• **Mesh repository:** stores and retrieves the Service MeshUp descriptions.

• **Service runtime:** invokes the knowledge services on local/remote grid nodes.

• **Service repository:** creates the catalog of the knowledge services, which have the mappings between service names (used in Service MeshUp description) and their end-point-references (used by Service Runtime for service invocations).

The Mesh Runtime interprets a given Service MeshUp description, and then resolves the end-point-references (EPRs) of the service names by looking up the Service Repository. According to the EPRs, the Mesh Runtime creates the VO consisting of the grid nodes where the services are deployed. The dynamically created VO provides a secure playground for sharing aspect properties and activating (i.e., invoking) the services. Once the VO is created, the Mesh Runtime initializes the aspect properties and starts the MeshUp process. While the MeshUp process is running, the knowledge services are invoked through Service Runtime.

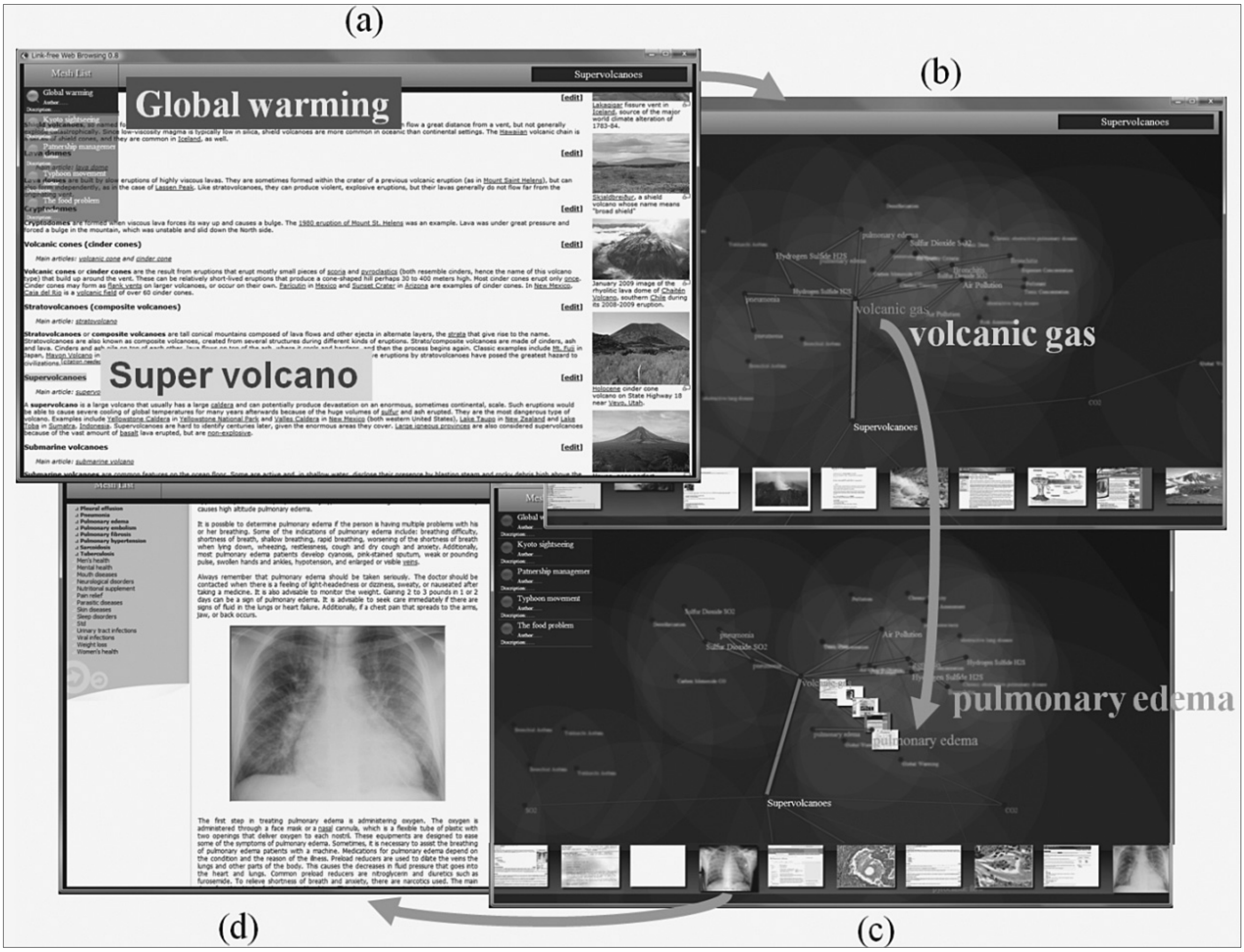
The Service Repository is managed in a decentralized manner. Once a new knowledge service is deployed to a grid node, the service information is stored in the local repository of the grid node first and then propagated to the Service Repositories on other grid nodes.

3. Application Example

Fig. 3 shows the application implemented on the Global Knowledge Grid. It allows users to browse thematically correlated web contents in a specific subject, rather than simply traversing hyperlinked web pages [10].

Here, let us suppose that when a user opens a web page to learn about volcanoes, a question ("What effect has global climate change had on us?") suddenly comes up in his/her mind. The user specifies the question to the

Figure 3. An application on Global Knowledge Grid for browsing thematically correlated web contents



browser, and then the browser shows a list of knowledge association services, each of which provides the context for leading him/her to a particular answer from the current page. Supposing that the user selects the knowledge association service "global warming," the browser first collects the relevant information from the content of current page (i.e., volcanoes) and then tries to describe the information by using, for instance, natural disaster concepts. It is done by the corresponding knowledge discovery service.

As the result, the user will discover concept terms like "volcanic gas" and "volcanic eruption". Supposing that the knowledge asset "global warming" is designed to find the possible effects to healthcare and ecosystem caused by the natural disasters related to global warming, the browser tries to discover the healthcare concepts and ecosystem concepts dependent on the natural disaster concepts, and then searches the web for information representing those concepts. The browser shows the causal relationships between the concepts using the semantic network layout, because it is considered the most direct way of visual navigation for causal relationships and is also specified to the knowledge asset "global warming" as the knowledge delivery method. It is done by the corresponding knowledge delivery service.

The user explores the causal relationships by following the edges of the semantic network. In addition, the information used for analyzing the causal relationships can be viewed by the corresponding edges. If the user clicks on the thumbnail of web information related to the concept node, the browser jumps to the web page containing the information, and then, the conventional web page browsing will be resumed.

4. Summary

Towards next generation web as collaboration media, the key concepts are:

(1) collaboration on demand with collective intelligence,
(2) grid architecture with emphasis on virtual organization, and
(3) everything as a service.

Our Global Knowledge Grid is an integrated infrastructure for coordinating knowledge sharing and problem solving in distributed environments, where three categories of knowledge services – knowledge capture, association and provision services –, collaborate on global-scale grid network. We explained the mechanisms for data-centric service collaboration, called Service Mesh Up, and dynamic VO construction to provide multi-institutional and secure playground for the service collaboration.

Currently, there are more than 400 services running on the Global Knowledge Grid. Our future work includes developing a service search engine for better performance and scalability.

References

[1] Zettsu, K., Nakanishi, T. Iwazume, M., Kidawara, Y., Kiyoki, Y., "Knowledge Cluster Systems for Knowledge Sharing, Analysis and Delivery among Remote Sites," Information Modeling and Knowledge Bases, Vol. XIX, IOS Press, pp.282–289, 2008.

[2] Zhang, R., Zettsu, K., Kidawara, Y., Kiyoki, Y., " A Decentralized Architecture for Resource Management of Group-based Distributed Systems," Journal of Frontiers of Computer Science in China (FCSC), pp.224–233, 2008.

[3] Cannataro, M., Talia, D., "The Knowledge Grid: Designing, Building, and Implementing an Architecture for Distributed Knowledge Discovery," Communications of the ACM, Vol. 46, No. 1, pp.89–93, 2003.

[4] Zettsu, K., Kiyoki, Y., "Towards Knowledge Management based on Harnessing Collective Intelligence on the Web," In Proc. of the 15th International Conference of Knowledge Engineering and Knowledge Management – Managing Knowledge in a World of Networks – (EKAW2006), Lecture Notes in Computer Science, Vol. 4248, pp.350–357, 2006.

[5] Foster, I., Kesselman, C., Tuecke, S., "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," International Journal of High Performance Computing Applications, Vol. 15, No. 3, pp.200–222, 2001.

[6] Papazoglou, M.P., Georgakopoulos, D., "Service-Oriented Computing," Communications of the ACM, Vol. 46, No. 10, pp.24–28, 2003.

[7] Future of European Grids, Grids and Service-Oriented Knowledge Utilities, Next Generation Grids (NGG) Expert Group Report 3, 2006.

[8] Fu, X., Bultan, T., Su, J., "Analysis of Interacting BPEL Web Services," In Proc. of the 13th International Conference on WWW, pp.621–630, 2004.

[9] The Globus Alliance, *http://www.globus.org/*

[10] Nakanishi, T., Zettsu, K., Kidawara, Y., Kiyoki, "Approaching to Interconnection of Heterogeneous Knowledge Bases on a Knowledge Grid," In Proc. of The International Conference on Semantics, Knowledge and Grid (SKG 2008), pp.71–78, 2008.