# 6-4 The Language Grid: Multi-Language Infrastructure based on Service-Oriented Collective Intelligence

**MURAKAMI Yohei**

The Language Grid is a service-oriented multi-language infrastructure for supporting intercultural collaboration. Language resources like dictionary data and machine translation software that are traditionally distributed with CD copies and are downloaded should be transformed to language services that can be easily available through the Internet and allow users to combine them according to intercultural collaboration fields. In this paper, we introduce the infrastructure software that enables collecting, sharing, and combining language services, the intercultural collaboration environment that enables easy usage of language services registered in the infrastructure software, and several typical use cases.

***Keywords***
Service-oriented architecture, Multi-language infrastructure, Service grid, The Language Grid, Intercultural collaboration

## 1 Introduction

The Internet connects people all over the world. However, we still have a language barrier. People on the Internet use a variety of languages and have no standard language. According to the survey by Global Reach, English users on the Internet occupy 35.2%, Asian language users 26%, and European language users 28%. There are various types of information shared on the Web, but it is extremely difficult to understand the content of the information as it requires people to understand many languages. Also, the Internet has many language resources (in the forms of data and software), but it is difficult for people, except specialists, to use them for intercultural collaboration. Complicated contracts, intellectual properties, data structure, and diversity of interfaces make the use of the language resources difficult.

An effective solution to these problems is not the separate action of individual users but the intelligent action of user groups, called collective intelligence [1][2]. In recent years, World-Wide Web activities to form a collective intelligence such as the Wikipedia have been developing, forming a knowledge foundation for all human. The currently-formed collective intelligence is a content-oriented collective intelligence based on the accumulation of documents, photographs, and videos. In this study, on the other hand, we propose a service-oriented collective intelligence which is based on assembling of services such as machine translation or speech recognition. In particular, we explain "the Language Grid" [3], a collective intelligence platform of language services for the support of intercultural collaboration. With the Language Grid, end users can combine existing language services provided by researchers and specialists and add their own language services to create a new language service for their purposes. In this paper we explain the following three issues that we worked on in the development of the Language Grid.

-**Development of service-oriented multi-language infrastructure**: To collect and share language services, it is necessary to have platform software that combines them based on atomic services. In addition, users have to be able to easily develop an application system that supports intercultural collaborations using the language services.
-**System design of operation model**: The service-oriented collective intelligence has various stakeholders. Users have their own requirements and providers have their own policies. For the coordination of these different stakeholders, the operator has to design an operation model taking account of both parties' incentives.
-**Implementation of user-involved design**: The more language services are provided, the more benefits users can receive from the services. It is therefore necessary to have the users and their community proactively join the designing of a Language Grid in order to form a service-oriented collective intelligence*1.

Studies on the Language Grid, whose target is language services, face a number of general problems about services computing. For example, one of the problems we faced is how the services should be accumulated and shared in an open environment or how new services developed by users or their community could be supported [5].

In this paper we first explain the platform software of the Language Grid and then describe an operation model of the Language Grid that uses the software. Finally, we introduce the implementation of the Language Grid in a local community and global community as an example of the user-involved designing.

## 2 Language Grid architecture

### 2.1 Design policy

The Language Grid takes an approach of collective intelligence. Thus, it is designed as an environment where specialists and various local users can share and use their developed language resources (Fig. 1). The characteristic of the Language Grid is to share the language resources as services. In this environment, there are three types of stakeholders, i.e. Language Grid operators, service providers, and service users. Language Grid operators manage the Language Grid and control the implementation of language services. Service providers wrap the language services, such as machine translations, morphological analyzers and dictionaries, as a language service and register it to the Language Grid. Service users invoke the registered language services and use them for intercultural collaboration activities.

The Language Grid is thus a platform that combines the language services provided by different organizations. Previously, DFKI's Heart of Gold [6] and IBM's UIMA [7] attempted to combine language processing programs. These platforms, mostly for researchers and developers, provide a pipeline that links various language processing programs owned by them. On the other hand, the Language Grid is a platform where applications can use the language resources based on a service-oriented architecture with focus on the management of intellectual properties. Since the above two approaches have different purposes, a joint research was conducted to make a connection between DFKI's Heart of Gold and the Language Grid [8].

### 2.2 System architecture

As shown in Fig. 2, the Language Grid concept consists of the following four layers [9]. The lowest layer, P2P service grid, has an aim to connect two kinds of nodes, a core node and a service node. The core node manages the registered services and makes access control and service composition. On the other hand, the service node contains the service entities, i.e. language resources, and their wrappers.

Atomic service is a Web service that cor-

*1 Growth of the collective intelligence is made by spontaneous effort of users [4].

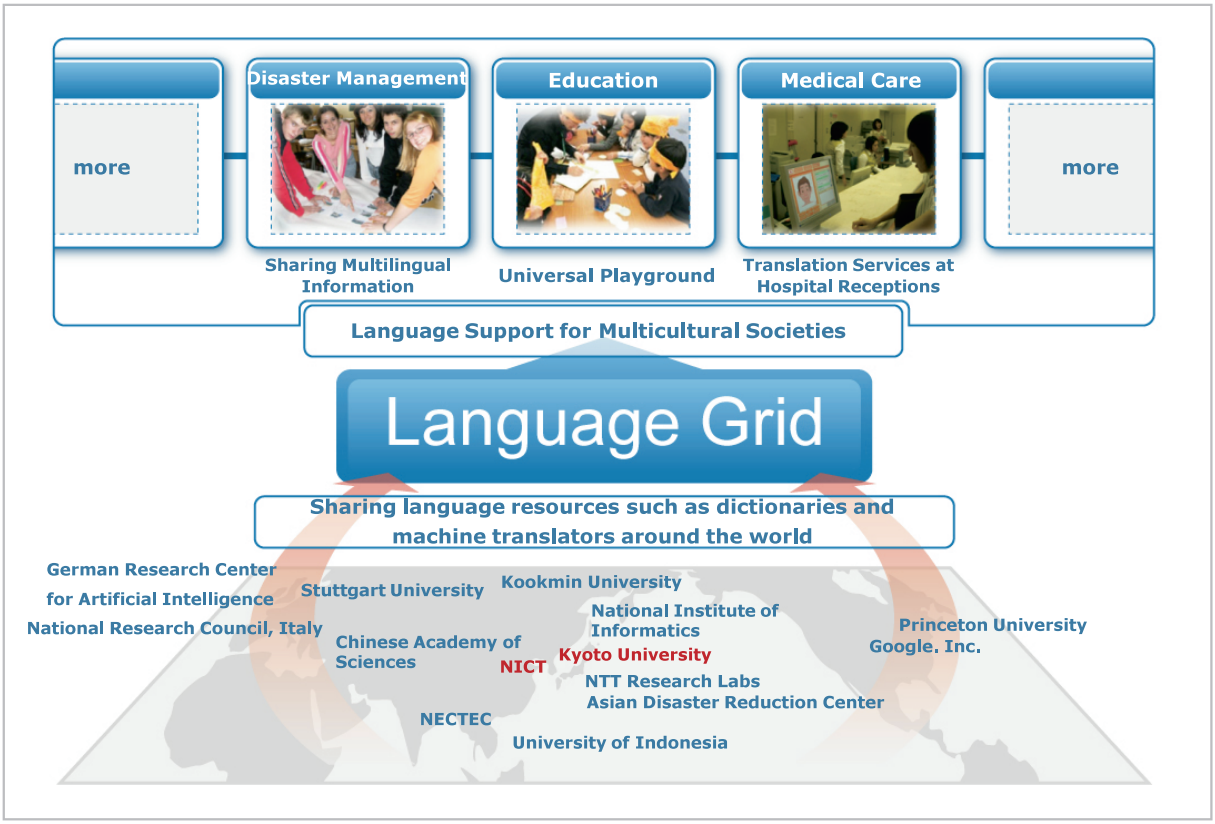


**Fig.1** *The Language Grid*

responds to each language resource and is provided on the service grid. It includes typical language resources such as machine translations, morphological analyzers, dictionaries, and parallel texts. These language resources are wrapped on the basis of a standardized service interface. An ontology was already proposed to standardize service interface layers of various language data and language processing programs [10].

Composite service is a combination of atomic services made by a workflow [11]. The workflow is described in WS-BPEL and interpreted and executed by BPEL execution engine [12]. In the language domain, a backtranslation service that translates a translated text to its original language, a specialized translation service that combines a technical term dictionary service and a translation service, and other various composite services are developed.

Application system is an application dedicated to each field in order to support intercultural collaborations. By using the composite

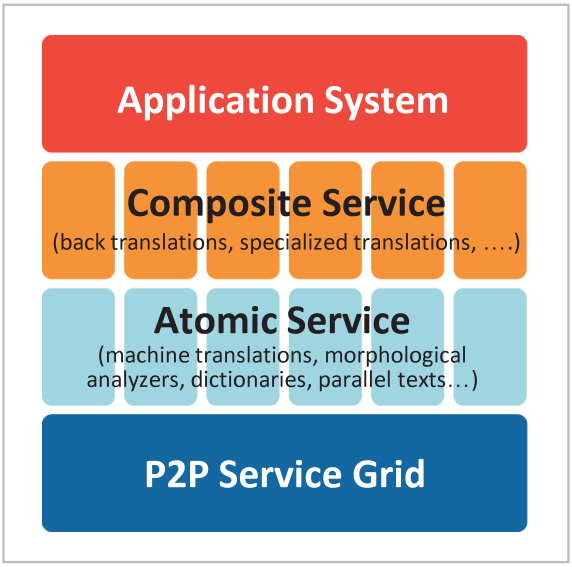


**Fig.2** *Layer structure of the Language Grid*

service and the atomic service, the application system can concentrate on the development of a user interface that handles interactions with users, which can reduce development cost.

Figure 3 shows a system configuration of P2P service grid that implements the four-lay-

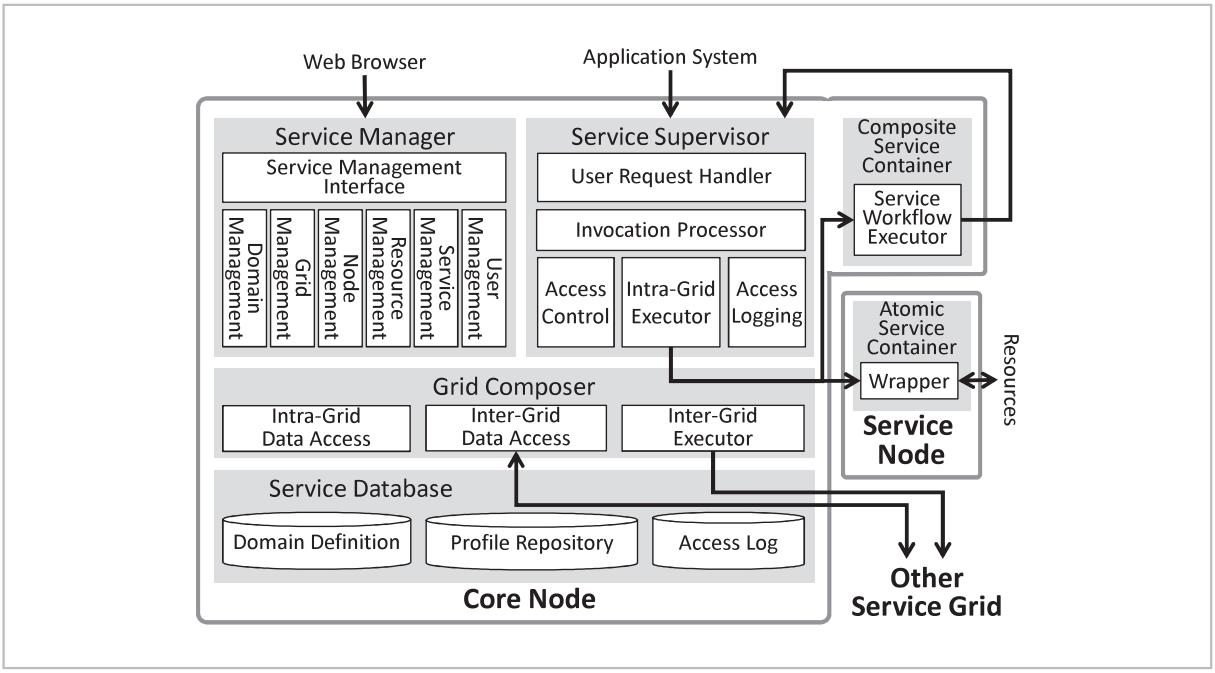


**Fig.3** *System configuration of P2P service grid*

er structure of the Language Grid [13]. After wrapping the language resources, service providers register a WSDL file, which describes the interface of the Web service, copyright information of the service, license information, and access restriction to Service Manager. The Service Manager receives the WSDL file, extracts the interface information and endpoint's URL, and creates a virtual endpoint of the same interface on Service Supervisor. The aim of the virtual endpoint is to prohibit direct access to the service, monitor the usage status of the service, and control access to the service according to the designated access constraint.

To use the service, a SOAP request is sent to the virtual endpoint and invokes the service. Service Supervisor receives the request and verifies if it satisfies the access constraint which was designated when the service was registered. If the access constraint is satisfied, the Service Supervisor acquires an actual endpoint of the service from its profile repository and accesses the service. If the service is a composite service, the request is sent to the service workflow executor and a corresponding workflow is executed. The service workflow executor sends a new request to the Service Supervisor according to the workflow. The responses of the service are accumulated in an access log to be used for the verification of satisfaction with the access constraint or for the monitoring of the use of the service.

The information accumulated in the access log and the service information registered by the service provider are shared by other core nodes through grid composers. With this system, service users can receive the same service from any of the multiple core nodes.

## 3 Management of Language Grid

### 3.1 Centralized operation model

To start management of the Language Grid, we collected the requirements of service providers and service users, who are both stakeholders, and developed a centralized operation model which focused only on nonprofit operations. In the centralized operation model, a single organization manages the Language Grid and user organizations all conclude a memorandum with this operation organization. The service providers can control access to the services that they provide. For example, they can monitor the usage statistics (access counts,

access log, data traffic), select users who are allowed to access the services, and set regulations for the use of the services. Also for the prevention of unauthorized access, service users must not make their tools open to public or allow accesses from general users. Namely, service users need to be able to identify each end user when they release a tool [14]. As a result, NPO, NGO, or local governments who are facing intercultural collaborations do not use the Language Grid, contrary to expectations, and the grid is mostly used by universities. Figure 4 shows the transition of the organizations that use the Language Grid. At present, 139 organizations in 16 countries signed the memorandum. For example, research institutes including Chinese Academy of Sciences, CNR, DFKI, and NII, universities such as University of Stuttgart, Princeton University, Tsinghua University, and many Japanese universities, NPO, NGO and public agencies. NTT, Toshiba, Oki Electric Industry, and Google also joined the project and provide services such as machine translations without charge.

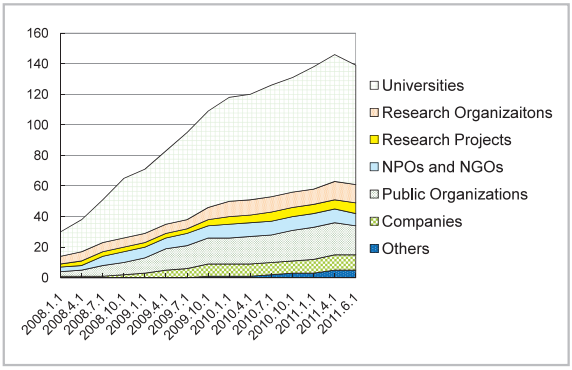


**Fig.4** *Number of participants of the Language Grid*

## 3.2 Federated operation model

The existing centralized non-profit operation model has two problems. One is that domestic user organizations occupy more than 70% of the user organizations of the Language Grid. The other is that large portion of the user organizations are universities. For the first problem, we designed a federated operation model which is managed by multiple operation organizations (Fig. 5). In the federated operation model, two roles, i.e. affiliated operation centers and affiliated users, are introduced so that not a single operation organization but multiple operation organizations operate the Language Grid collaboratively. Affiliated operation centers are user organizations that carry on their own Language Grid using the same memorandum. Affiliate users are user organizations that sign a memorandum with the affiliate operation centers and do not sign a memorandum directly with the operation organization. By allowing the affiliate users to use the services on the Language Grid, they can use the services which might not be usable if the Language Grid is managed by a single operation organization, and language services can be collected from the user organizations.

For the second problem, we expand the utilization purpose of the Language Grid and

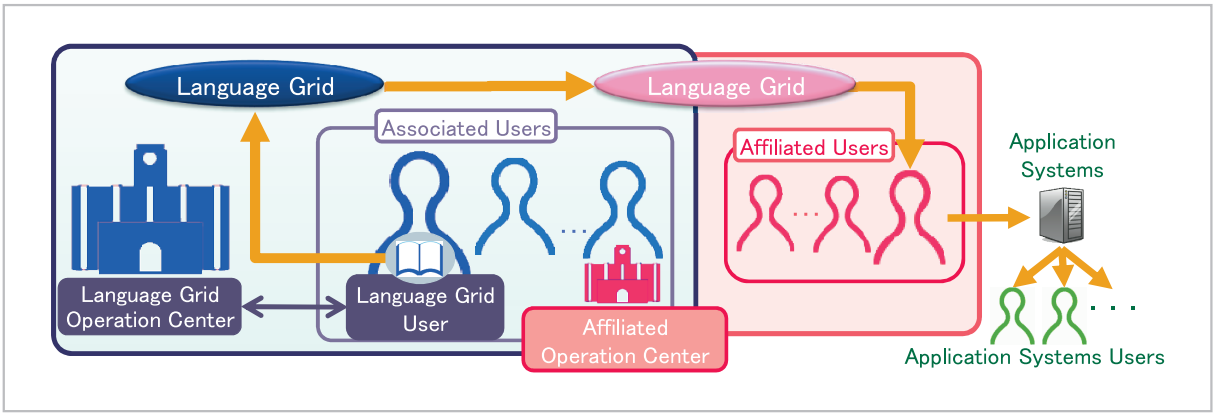


**Fig.5** *Federated operation model*

the range of the usage method and allow users to choose one. When registering a service, service providers designate its utilization purpose and utilization form that they allow. On the other hand, users present their utilization purpose and utilization form when they use the service. The providers and the users make mutual matching on the system for the access control. With this system, commercial enterprises can use the Language Grid for research purposes, or tools can be released from a server. It is hence expected that the diversification of the usage methods would increase the variety of user organizations.

In February 2011, NECTEC in Thailand founded a Language Grid operation center in Bangkok and launched a federated operation with the operation center of Kyoto University [15]. As a result, more than 120 language services are now registered in the Language Grid (in Kyoto and Bangkok), where various atomic services and composite services of 20 service types, such as translation, bilingual dictionary, parallel text, morphological analysis, and text-to-speech, are shared.

## 4 Use of the Language Grid

### 4.1 Environment for supporting intercultural collaboration

In the collective intelligence, the platform grows only through spontaneous activities of users. As the service providers provide more services, the service users can receive more benefits from the services. Therefore, we have employed a user-involved design since we launched the Language Grid, in collaboration with NPOs, potential service users. These NPOs, NGOs, schools and other non-profit organizations create their own dictionaries or parallel texts and provide them as services to improve translation quality, and use them in combination with ordinary translation services through the Language Grid. In this study, we developed intercultural collaboration support tools using a language service tuned for each field, with focus particularly on disaster prevention, education, and medical care. In this section we introduce two intercultural collaboration support tools that are versatile and customizable and that were developed to reduce the development cost of intercultural collaboration support tools of each field.

Language Grid Playground is an application system developed by a student group of Kyoto University. Playground allows users to access various language services on the Language Grid through a Web browser. Playground has a basic component that is used to use atomic services, advanced component that is used to use composite services, combinations of atomic services, and customized component that is dedicated to being used for intercultural collaboration activities. The customized component is formed by combining the basic component and the advanced component in accordance with the application purpose. Figure 6 shows the customized component of a multilingual chat tool developed for Fujimi Junior High School in Kawasaki City.

On the other hand, Language Grid Toolbox is a module group that supports intercultural collaboration in a community. It has a function of supporting communications such as multilingual BBS. It is provided based on CMS of an open source software called XOOPS, and can be extended by each community when necessary. At present, NPO PANGAEA is developing a multilingual community site using the Language Grid Toolbox (Fig. 7).

### 4.2 Use by local community

With the increase of the number of foreigners in Japan, communication with foreign patients who cannot speak Japanese well has become a problem in medical front. For medical treatment, medical staffs have to correctly inform foreign patients of the symptoms, drugs and insurance system. In Kyoto, supports by volunteer medical interpreters are available and the demand of this supports has been still increasing.

In this situation, Wakayama University and Center for Multicultural Society Kyoto developed Medical Interview Sheet Translation System $M^3$ that supports commu-

nications between medical staffs and foreign patients using parallel-text samples (Fig. 8) [16]. In medical front, particularly at reception desk of hospitals, since frequently-used examples are necessary, medical parallel-text collection system TackPad was developed to help volunteer medical interpreters to collect parallel-text samples.

At present, M$^3$ is introduced in Kyoto City Hospital, Kyoto University Hospital, Rakuwakai Otowa Hospital, and the University of Tokyo Hospital to be used for multilingual

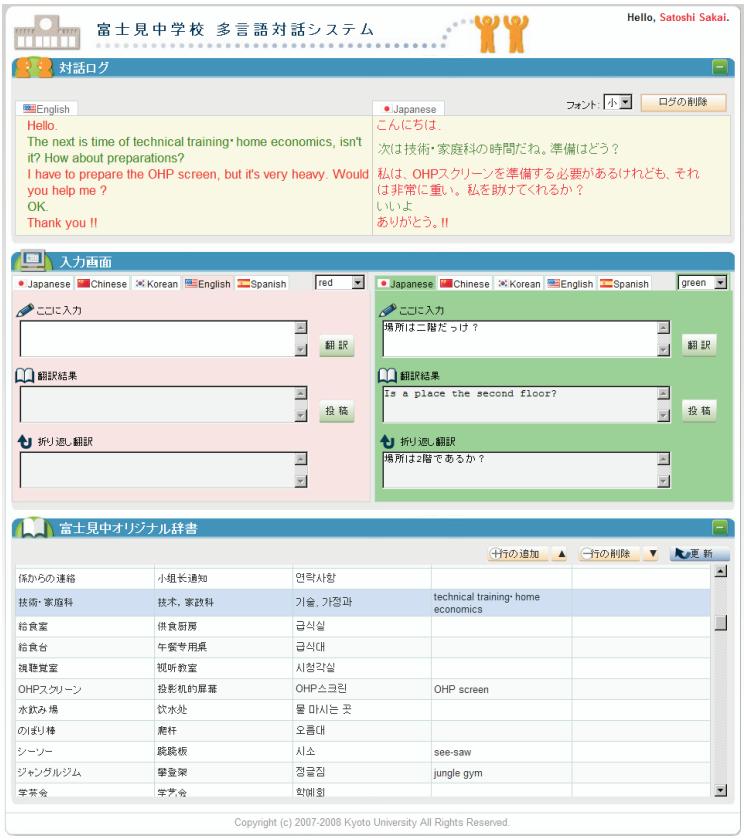


**Fig.6** *Language Grid Playground*


**Fig.7** *Use case of Language Grid Toolbox*

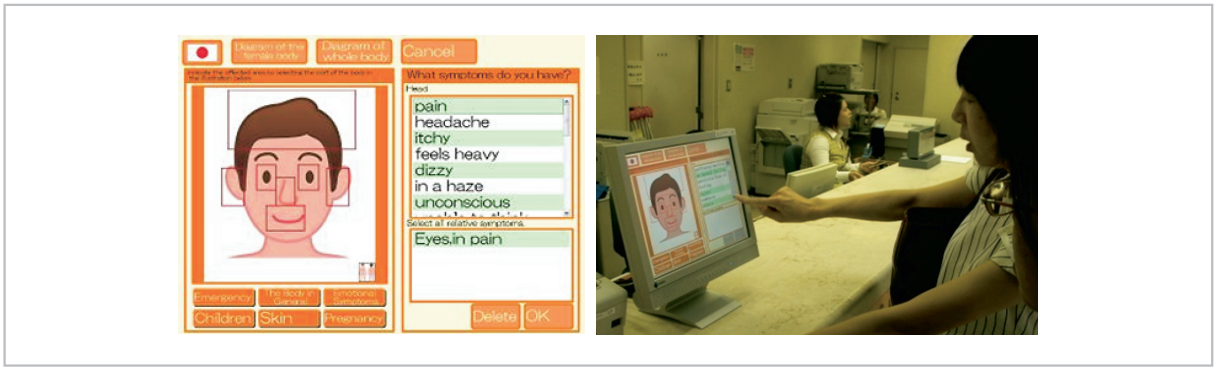


**Fig.8** *Multilingual Medical Questionnaire System M³*

medical questionnaire. Web-based M³ and mobile M³ are also released for medical support of patients before they visit a hospital.

### 4.3 Use by global community

Wikipedia is provided in about 270 languages for information sharing since it can be created and edited by anyone. The articles in Wikipedia, written with their own cultural backgrounds, are information resources for intercultural mutual understanding. However, there is a large bias in the number of articles written in different languages. For example the number of English articles is 3,540,000, that of Japanese articles is 730,000, and that of Thai articles is 60,000. For the acceleration of translations of knowledge, a multilingual bulletin board where translations can be discussed would be necessary.

We, therefore, developed a multilingual bulletin board utilizing the Language Grid on MediaWiki in collaboration with Wikimedia Foundation*2. Wikipedia volunteers over the world can use this multilingual bulletin board to make questions and answers for the translation of articles.

In the course of the development, we first created a Language Grid extension on MediaWiki to provide access means to the Language Grid (Fig. 9). Using this, we then developed a multilingual bulletin board, Multilingual Liquid Thread, an extended version of single language bulletin board, Liquid Thread. On the Multilingual Liquid Thread, one can create a multilingual glossary, which

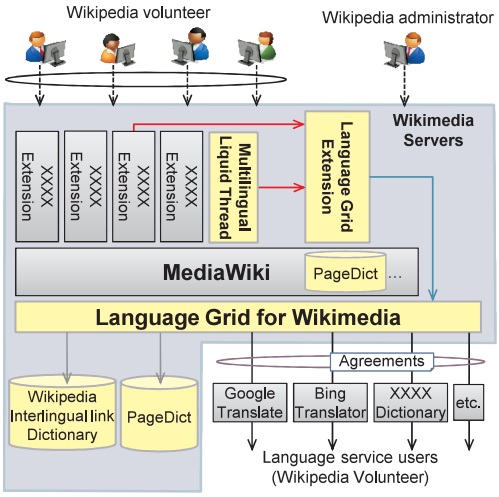


**Fig.9** *System configuration of Multilingual Liquid Thread*

can be used to customize machine translation to improve the translation accuracy of articles.

## 5 Conclusions

In this study, we developed a service-oriented multi-language infrastructure to support the collection and sharing of language services while protecting intellectual property rights of the language resources. The contributions made in this study are the following three:

-**Development of service-oriented multi-language infrastructure**: The Language Grid consists of four layers: P2P service grid,

*2 MediaWiki is a platform of Wikipedia and other services provided by Wikimedia Foundation.

atomic services, composite services, and application system. The P2P service grid is used to share information between core nodes so that service users can use the same service on any nodes and that service providers can make the same control of access to any nodes.

- **System design of operation model**: We proposed an operation model to make coordination of different types of stakeholders in the Language Grid. This operation model is designed to make the incentive of the service users fit that of the service providers. For the improvement of the accessibility of the Language Grid, we also provided a federated operation model that enabled collaboration of multiple operation organizations.
- **Implementation of user-involved design**:We showed that the service-oriented approach and the versatile intercultural collaboration support tool actually accelerated the user-involved designing. We showed a case example that the users, such as, schools and NPO developed an intercultural collaboration environment using the Language Grids, Playground and Toolbox.

The Language Grid is a service-oriented multi-language infrastructure for the development of multilingual environment in accordance with the purpose of users. It allows users to freely combine the language services provided by universities, research institutes, and companies. As a result, the Language Grid is now used for multilingual support activities for schools and local shopping area communities. For example, CoSMOS (Collaborative Safety Maps on Open System), a system for supporting disaster-prevention cooperative studies by sharing disaster-prevention safety maps drawn by children over the world, was developed [17].

The Language Grid is a multi-language infrastructure, but its platform software or operation model is not dedicated only to language domain. It can be applied to other domains by newly defining an interface of services. In future, we will develop a service platform that promotes utilization of big data through a service-oriented collective intelligence approach by systematizing large-scale science data services or archive data services and large-scale data analysis services.

## Acknowledgement

The author would like to thank Professor Toru Ishida of Kyoto University who gave me the research motivation and led the project, the Language Grid Kyoto Operation Center (Graduate School of Social Informatics, Kyoto University) and Bangkok Operation Center (NECTEC) which actually conduct daily operation of the Language Grid, and the Language Grid Association which supports the Language Grid user organizations.

## *References*

1 T. Gruber, "Collective Knowledge Systems: Where the Social Web meets the Semantic Web," Journal of Web Semantics, Vol. 6, No. 1, 2008.

2 P. Levy, "Collective Intelligence: Mankind's Emerging World in Cyberspace," Cambridge, MA: Perseus Books, 1999.

3 T. Ishida, "Language Grid: An Infrastructure for Intercultural Collaboration," IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp. 96–100, keynote address, 2006.

4 A. Weiss, "The Power of Collective Intelligence," networker, Vol. 9, No. 3, pp. 16–23, 2005.

5 M. P. Papazoglou, P. Traverso, S. Dustdar, and F. Leymann, "Service-Oriented Computing: a Research Roadmap," Int. J. Cooperative Inf. Syst. Vol. 17, No. 2, pp. 223–255, 2008.

6 U. Callmeier, A. Eisele, U. Schafer, and M. Siegel, "The Deep Thought Core Architecture Framework," LREC 2004, pp. 1205–1208, 2004.

7 D. Ferrucci and A. Lally, "UIMA: An Architectural Approach to Unstructured Information Processing in the

Corporate Research Environment," Natural Language Engineering, Vol. 10, pp. 327–348, 2004.
8 A. Bramantoro, T. Tanaka, Y. Murakami, U. Schäfer, and T. Ishida, "A Hybrid Integrated Architecture for Language Service Composition," IEEE International Conference on Web Services (ICWS-08), pp. 345–352, 2008.
9 Y. Murakami and T. Ishida, "A Layered Language Service Architecture for Intercultural Collaboration," International Conference on Creating, Connecting and Collaborating through Computing (C5-08), 2008.
10 Y. Hayashi, T. Declerck, P. Buitelaar, and M. Monachini, "Ontologies for a Global Language Infrastructure," Proc. of ICGL2008, pp. 105–112, 2008.
11 R. Khalaf, N. Mukhi, and S. Weerawarana, "Service-Oriented Composition in BPEL4WS," Proceedings of the 2003 World Wide Web Conference, 2003.
12 T. Andrews, F. Curbera, H. Dolakia, J. Goland, J. Klein, F. Leymann, K. Liu, D. Roller, D. Smith, S. Thatte, I. Trickovic, and S. Weeravarana, "Business Process Execution Language for Web Services," 2003.
13 Y. Murakami, M. Tanaka, D. Lin, and T. Ishida, "Service Grid Federation Architecture for Heterogeneous Domains," International Conference on Services Computing (SCC-12), 2012.
14 T. Ishida, A. Nadamoto, Y. Murakami, R. Inaba, T. Shigenobu, S. Matsubara, H. Hattori, Y. Kubota, T. Nakaguchi, and E. Tsunokawa, "A Non-Profit Operation Model for the Language Grid," International Conference on Global Interoperability for Language Resources, pp. 114–121, 2008.
15 T. Ishida and Y. Murakami, "Institutional Design for Service-Oriented Collective Intelligence," The Transactions of the Institute of Electronics, Information and Communication Engineers D, Vol. J93-D, No. 6, pp. 675–682, Invited Paper, 2010.
16 M. Miyabe, T. Yoshino, and A. Shigeno, "Development of a Multilingual Medical Reception Support System Based on Parallel Texts for Foreign Pationts," The Transactions of the Institute of Electronics, Information and Communication Engineers D, Vol. J92-D, No. 6, pp. 708–718, 2009.
17 Y. Ikeda, Y. Yoshioka, and Y. Kitamura, "Intercultural Collaboration Support System Using Disaster Safety Map and Machine Translation," Culture and Computing, Lecture Notes in Computer Science 6259, Springer, 100–112, 2010.

(Accepted June 14, 2012)

***MURAKAMI Yohei**, Ph.D.*
*Senior Researcher, Information Services Platform Laboratory, Universal Communication Research Institute*
*Services Computing, Multi-Agent System*