The Language Grid for Intercultural Collaboration

Toru Ishida
Department of Social Informatics
Kyoto University
Kyoto 606-8501, Japan
ishida@i.kyoto-u.ac.jp

ABSTRACT
In this paper, we describe the design philosophy of the Language Grid, a new language infrastructure to create multilingual environments for intercultural collaboration. The key idea is to apply a service-oriented approach to wrap various language resources, which are globally distributed, and to combine them to create customized multilingual tools for local activities. We show how we are developing a global community to share language services and how local communities are adopting them for their daily life. Though the Language Grid is available in public, we can use the software to design a private language grid for a specific community. A Wikipedia translation community is described as one example of a multi-language environment.

Keywords
Language Grid, intercultural collaboration, service grid.

1. INTRODUCTION
Since multiple languages are used in various communities in daily life, tools that can effectively support multilingual communication should be provided. A number of online language services already exist such as bilingual dictionaries, machine translators and so on. However, difficulties often arise when people try to use those language services in their intercultural activities; complex contracts, intellectual property rights, and non-standard application interfaces make it difficult for users to apply those language services to their activities. Moreover, we often observe that the success of a multilingual tool in one situation does not guarantee its success in another. For example, web-based machine translation is not useful when buying souvenirs in a foreign market because of its insufficient coverage and poor timeliness.

To develop customized multi-language environments for various situations in various communities, we have implemented the Language Grid; it allows users to freely combine existing language services to develop new services so that users can easily create their own communication tools. The word “grid” is generally defined as “a system or structure for combining distributed resources; an open standard protocol is generally used to create high quality services [5].” Our objective, applying the “grid” concept to ensure the collaboration of language services, has not been tried before.

Since we take a collective intelligence approach in realizing the Language Grid, the platform can grow only through the voluntary efforts of users. The more users provide language services, the more they can utilize the benefits of the services. Thus the platform should be open to encourage users to create and share services. Conceptually, the Language Grid has two main structures: horizontal and vertical. The horizontal grid concerns the combination of existing bilingual dictionaries or machine translation systems for various languages [3]. The vertical grid concerns specific scenarios of intercultural collaboration activities, which require specialized language services including jargon handling. To develop the horizontal grid, we need the global collaboration of universities and research institutes, while for the vertical grid, local communities need to create their own customized multi-language environments.

2. THE LANGUAGE GRID
The Language Grid provides an environment where users can share language services developed by both professionals and end-users in various application fields. Users can wrap language resources as Web services and register them in the service grid. Web service technologies including language service ontology and service composition mechanisms have been developed to enable the collaboration needed among language services [7]. Language service ontology is a technology to define standard language service APIs in a hierarchical way so that end-users are provided with simple interfaces while professionals can access more complex interfaces.

To design an operation model of the Language Grid, we collected the requirements of various stakeholders: university laboratories and research institutes as typical language service providers, and NPO/NGOs and public sectors as typical language service users. Though our current target is language services, we tried to create a model as general as possible that could be applied to various kinds of service grids.

Major stakeholders of a service grid fall into three categories. *Service grid users* are not individuals but groups like research units in universities, and a single group can act as two different stakeholders. Service grid user is called a *service provider* when registering language services in the service grid. Service grid user is called a *service user* when invoking registered language services for intercultural activities. An *associated user* is a service grid user that is licensed to use the service grid under the same agreement. We define a *service grid operator* as a party that receives services from the service grid user and operates and provides services to associated users.

When using the services registered to the service grid, service grid users can allow participants in events or activities organized by the service grid users to utilize the registered services. We define an *application system* as a system that is operated by the service grid user. The application system should allow application system users to indirectly access the service grid without knowing the ID and password of the service grid. For example, in the case of an NPO offering medical interpreter services to foreign patients, the NPO should not enter their service grid ID and password in front of the patients, but embed them in their patient support systems.

To allow hospitals to use the services for anonymous patients, or to allow NPO/NGOs to use the services in their social activities, we need more precise rules. We define "*under client control*" as the status where the application system user is under the control of the service grid user. More specifically, "under client control" is the status where the terminals of the application user are under the control of the service grid user, or where the service grid user is able to identify the application system user who accesses the application system. In each case, the service grid user must grasp the use of the application system at each terminal or by each application system user. Furthermore, the service grid user should have the technical and legal authority to identify and suspend the use of the terminal or the application system user. We also define "*under server control*" as the status where the server used to operate the application system is under the control of the service grid user, while the application system user is not necessarily under the control of the service grid user. In this case, the service grid user must grasp the use of the server of the application system and have the technical and legal authority to suspend the server.

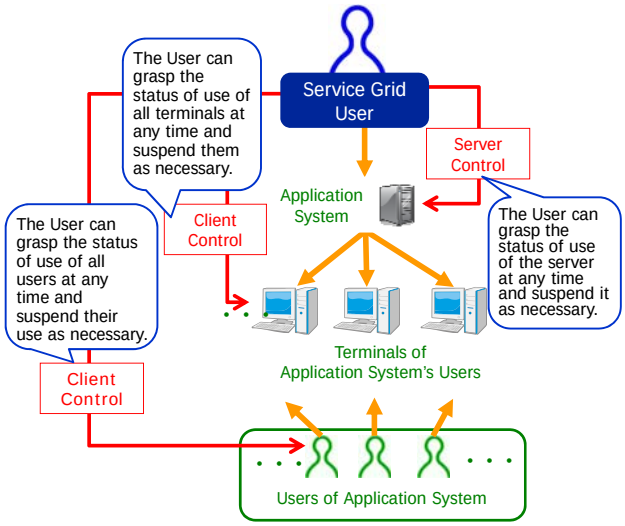

Figure 1: Control of Service Usage in Application Systems

In any case, to guarantee the intellectual property rights of service providers, they are allowed to specify which service users are to be allowed to use the provided services by selecting each user, and to specify the control method of service usage such as "*under client control*" or "*under server control*."

The *Service Manager* was implemented to accelerate the desire of all stakeholders to share information about the service grid. This software supports users to register, delete, and search services, and to monitor and control service access. Monitoring function provides access count and data consumed by each service user. Access control function allows service providers to protect their rights over each service by prohibiting access to the service. Service providers have two choices in publishing their services: *the public mode*, give permission is given to all users and *the members-only mode* to selected users. Using the members-only mode, a service provider who provides a commercial service can permit a service user who already purchased the service or its license to access the service.

Copyright is held by the authors.
Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.

3. GLOBAL COMMUNITY FOR RESOURCE SHARING

After starting operation at Kyoto University in December 2007, the number of participant organizations has steadily increased. So far, 118 groups from 17 countries have signed up and are sharing more than 60 language services. Several companies have also joined to provide their services: Toshiba, Oki, Kodensha, Cross Language, Google and Translution provide their translation services. However, most of the participant organizations are located in Japan; it is not easy for service grid operator in Kyoto to reach users worldwide.

Since we need global collaboration, even to solve language issues in local organizations, we decided to introduce the concept of *affiliated operator* and *affiliated users* to naturally extend the original operation model. More precisely, we define the affiliated operator as an associated user that operates its own service grid using the same agreement, except for its governing law, and affiliated user as a service grid user that is licensed to use the service grid by the affiliated operator under the same agreement, except for its governing law. In this way, service grid operators can be connected in a P2P fashion, and affiliated users can access the services registered to a different service grid through the affiliated operator.

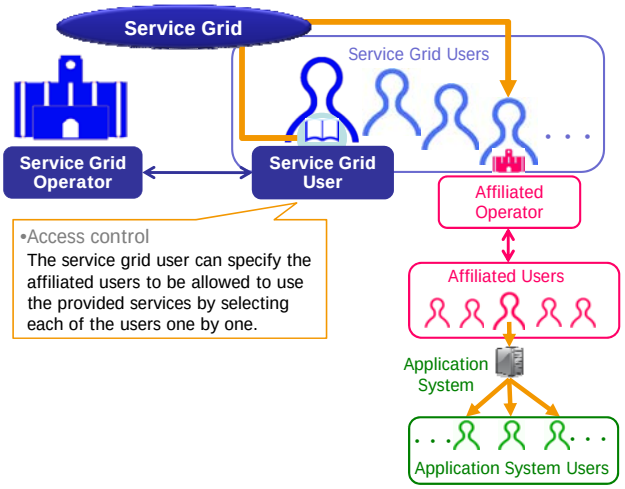

Figure 2: Federated Operation of Service Grids

In the federated operation model, just as in the original operation model, the service providers can specify the affiliated users to be allowed to use the provided services by selecting each user one by one or selecting all affiliated users at once. In both cases, the service provider should understand that the affiliated users will use the provided service under a different governing law. When the service provider does not explicitly allow such use, the affiliated user must not use the provided service.

This model allows two different types of service grid federation. In *homogeneous federation*, service grids in the same domain are connected. In this case, one service grid operator is to be an affiliated operator of another service grid, and vice versa. For example, several language grid operators in different regions can be connected with each other to form a worldwide language infrastructure. In *heterogeneous federation*, on the other hand, service grids in the different domain are connected. In this case,

the relation between service grids can be asymmetric. An e-learning service grid is to be an affiliated operator of a language grid to create a multilingual courseware, but the language grid may not be an affiliated operator of the e-learning service grid if there is no need to use e-learning courseware for creating language services.

4. LOCAL COMMUNITY FOR MULTICULTURAL ACTIVITIES

In multicultural local communities, the languages spoken contain a lot of jargon, which machine translators do not know. For instance, enter the line “You have cleanup duty today” in Japanese and translate it into Korean. The following sentence “오늘은너가청소
당번이야” appears on the screen. Japanese teachers do not understand Korean, so they are unsure if the translation is correct. They perform back translation, i.e., they translate the Korean back into Japanese and the following sentence is displayed “You should clean the classroom today!” This sentence seems a little rude when spoken; however, it may be acceptable, if the line is delivered with a smile. Now, if we translate it into Chinese in the same manner, the following Chinese equivalent “今天你是扫除值日哟” would appear on the screen. When we back translate the Chinese sentence into Japanese, we discover the very strange sentence “Today, you remove something to do your duty.” It appears that the word “掃除当番,” which means duty to clean the classroom, was not registered in the dictionary of this machine translator.

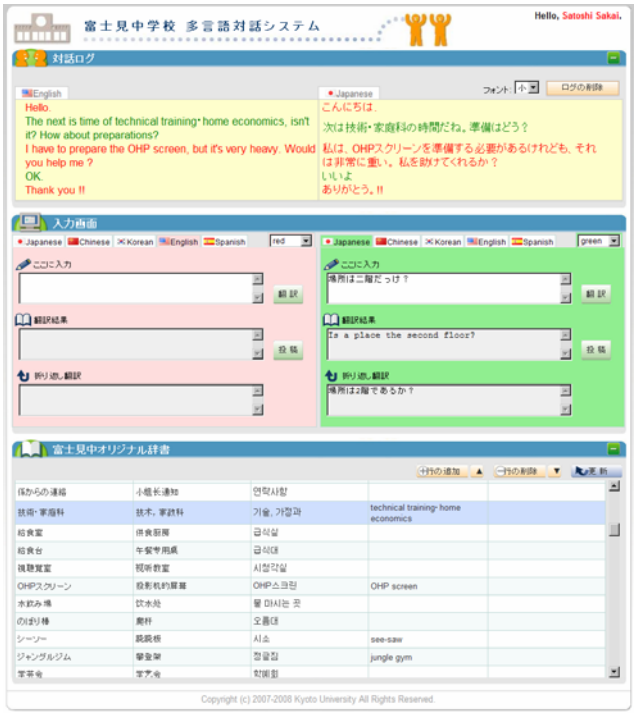

Figure 3: Shared Screen Multilingual Chat

Japan now has an increasing number of students who are non-native Japanese speakers, and most teachers are unable to communicate with the foreign students and their parents. Therefore, we developed the service in which users can chat in a multilingual environment. The support site, called *a shared screen multilingual chat*, was designed specifically for this situation; students, parents and teachers can chat while looking at the same display. As shown in Figure 3, users can input text in their mother tongue, translate the sentence with their own dictionaries, check the back translation, and post it to the log area on top of the page. In addition, to customize machine translations, we need to combine community-specific dictionaries with translators [4], it is necessary to compile a multilingual dictionary of words that are frequently used in schools. Then, we can replace words output by machine translators with the more correct ones in local dictionaries. Users can register terms used in the school into the user dictionary, which makes the translation result more correct. This service provides auto-completion using a glossary for school life provided by local governments. This site was developed in two weeks by three graduate students in Kyoto University. This example shows how quickly we can create a customized multilingual environment by using the registered language services.

There exist situations where we cannot adopt machine translation because of its current quality. A typical case can be found in hospitals. When foreigners, who are not fluent in Japanese, fall ill in Japan, they may be unable to receive adequate medical attention because of their inability to communicate with Japanese doctors. In Kyoto, volunteer interpreters are being dispatched up 1700 times per year. Interpreters are also stationed in several affiliated hospitals. A support system has been developed for communication between foreign outpatients and medical staff at hospital reception desks. After the outpatient answers some questions posed by the system, it replies with the appropriate consultation procedure. Since machine translations are not useful, the system refers to multilingual parallel texts of medical sentences registered to the Language Grid. The parallel-text-collection system has also been developed to accumulate highly accurate parallel text created by volunteer interpreters. This system has been developed and is currently being used at the reception desks of Kyoto City Hospital and Kyoto University Hospital [6].

Figure 4: Multilingual Reception System

Several tools have been prepared to support intercultural activities. Language Grid Playground provides easy access to the Language Grid to try a variety of registered language services through a Web browser. Services in the Playground are categorized as follows. The BASIC services provide interfaces for easy selection and usage of the atomic language services. The ADVANCED services provide composite language services created by combining existing language services. The CUSTOMIZED services provide specific language services to organizations that carry out intercultural activities. Source codes are open to public so as to allow users to learn how to write a program using language services.

Language Grid Toolbox, on the other hand, is a collection of modules to support multilingual communication in a community. Users can install this software into their servers to start services, such as multilingual BBS and multilingual dictionary creation. Since Toolbox is based on the open source environment CMS (Content Management System) XOOPS Cube, Toolbox is also provided as open source software. Therefore, the functions of Toolbox can be extended by developing XOOPS modules to meet the requirements of user communities. Using this Toolbox, we started a project called "Kyoto on the Language Grid" as a successor of Digital City Kyoto [2]. We are creating multilingual BBS and QA Web sites for various communities (shopping streets, temples, hospitals, and schools) in Kyoto to support intercultural collaboration. Basic documents for communities (apartment rental contracts, for example) have been translated in several languages by human experts so as to extract parallel texts and multilingual dictionaries. We use those language resources to train and to customize machine translators. Example-based machine translation [8] has been tested and shown to create highly customized translation results.

5. WIKIPEDIA TRANSLATION COMMUNITY

Even if translation quality is increased by the development of community language resources, we cannot solve all communication problems through translation; we must deepen our knowledge of different cultures to reach an assured mutual understanding. For example, we can translate the Japanese term "cleanup duty" into Portuguese, but it can still puzzle students from Brazil, because there's no such concept in Brazil. As is well known, linkage of one language to another is the first step in understanding, thus we need a system that associates machine translation results with various interpretations of concepts to help us understand different cultures. Wikipedia can be a resource for intercultural collaboration when combined with machine translators, because a large portion of Wikipedia articles are and will be provided in different languages and linked together.

We have started to collaborate with the Wikimedia Foundation to apply the Language Grid to Wikipedia translation. We think Wikipedia articles should be created by humans with the support of machines. Humans can create new articles in their own cultural context, while machines can translate existing articles for reference. Conversely, the translation of Wikipedia articles should be done by machines with the support of humans. This is because machines can provide culturally neutral translations, while humans can improve the quality of machine translations by providing community-specific dictionaries and parallel texts.

Based on the above design policy, we have developed the Language Grid Extension on the MediaWiki platform together with PageDict, a multilingual dictionary for each Wikipedia article, so that various tools on the MediaWiki can easily use language services. For example, LiquidThread, a discussion platform on MediaWiki has been successfully multilingualized to support discussions on translating Wikipedia articles into various languages.

6. CONCLUSION

The Web can be a social machine to strengthen both global and local ties [1]. We are living in multicultural societies, however, where language and cultural barriers hinder intercultural collaboration. This paper explained how the Language Grid, an infrastructure that allows end-users to create new language services for their intercultural collaboration activities, increases the accessibility and usability of online language resources. To this end, language resources including data and programs are wrapped as web services so that users can easily share and combine these web services for creating their own multilingual environment. Using the Language Grid, various kinds of intercultural activities have begun at hospital receptions, local schools, shopping street communities and so on.

ACKNOWLEDGMENTS

We are grateful to the members of NICT Language Grid Project and Kyoto University Language Grid Operation Center. This research was partially supported by Strategic Information and Communications R&D Promotion Programme from Ministry of Internal Affairs and Communications, and a Grant-in-Aid for Scientific Research (A) (21240014, 2009-2011) from Japan Society for the Promotion of Science (JSPS).

REFERENCES

[1] J. Hendler and T. Berners-Lee. From the Semantic Web to Social Machines: A Research Challenge for AI on the World Wide Web. *Artificial Intelligence*, Vol. 174, No. 2, pp. 156-161, 2010.

[2] T. Ishida. Digital City Kyoto: Social Information Infrastructure for Everyday Life. Communications of the ACM, Vol. 45, No. 7, pp. 76-81, 2002.

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

[4] T. Ishida. Intercultural Collaboration Using Machine Translation. *IEEE Internet Computing*, pp. 26–28, 2010.

[5] K. Krauter, R. Buyya and M. Maheswaran. A Taxonomy and Survey of Grid Resource Management Systems for Distributed Computing. *Software-Practice & Experience*, Vol.32, No.2, pp.135-64, 2002.

[6] M. Miyabe, K. Fujii, T. Shigenobu and T. Yoshino. Parallel-Text Based Support System for Intercultural Communication at Medical Receptions. *Intercultural Collaboration*, LNCS 4568, pp. 182-192, 2007.

[7] 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.

[8] H. Somers. Example-Based Machine Translation. *Machine Translation*, Vol.14, No.2, pp.113-157, 1999.