The Language Grid

- The Language Grid combines users’ language resources and machine translators to produce high-quality translation that is customized to each field.
- The Language Grid, a software that provides a multilingual service infrastructure, was developed and released as an open source software by the Language Grid Project of the National Institute of Information and Communications Technology (NICT) from April 2006.
- Using the open source code, universities, research institutes, and nonprofit organizations (NPOs) can operate the Language Grid.
- The Department of Social Informatics at Kyoto University has started operation of the Language Grid for nonprofit/research purposes from December 2007.
- Users have established the Language Grid Association and have started activities for intercultural collaboration, using the language resources registered by users around the world.
- Three types of stakeholders exist in the Language Grid:
  - Language Grid Operator: The operator manages Language Grid Users and controls language resources and computation resources.
  - Language Resource Provider: The provider registers language resources such as machine translations, morphological analyzers, specialized dictionaries, and parallel texts to the Language Grid.
  - Language Service User: Organizations or individuals who use language and computation resources.
Role of the Language Grid

- As an example, let us explain how the Language Grid is operated. Currently this Language Grid is operated by Kyoto University for nonprofit/research use.

- Research institutes and universities participating in the Language Grid are providing language resources such as dictionaries, parallel texts, or machine translators for free. When providing language resources, providers can specify copyright notices and license information on the profiles of the resources. Providers can also set restrictions on users’ access to these resources.

- The organizations that provide their language resources to the Language Grid include the following: Chinese Academy of Sciences; National Research Council, Italy; Stuttgart University; National Institute of Informatics, Japan; NTT Communication Science Laboratories; and Asian Disaster Reduction Center.

- Using language resources provided by users, NPOs, schools and other nonprofit sectors have started playing a central role in breaking the language barriers. Their activities cover a broad range of fields, including disaster management, education, and medical care.

- As described above, the Language Grid is a platform that enables users to share language resources and supports activities for intercultural collaboration.
From Language Resources to Language Services

- Language resources like dictionaries and machine translation are always provided as data with CD/DVD medias. The users should install and consider the maintenance of the language resources, while the providers should deal with the issues of intellectual property.

- The Language Grid enables sharing of language services to improve the usability and accessibility of language resources. In the Language Grid, language resources are wrapped as Web services with standard interfaces. By this means, users can use the language services around the world.

- By wrapping language resources as Web services, the users can compose different language services. Therefore, the providers can provide services while protecting intellectual property by controlling the access of their language resources from users.

- For example, the user can get the translation of the entered word by using the dictionary service wrapped from the dictionary data.

- The user can get the translation result of the entered sentences by using the machine translation service which is wrapped from the machine translation software.

- Moreover, human translators can also be wrapped as human translation services to get higher translation quality than machine translation services.
Customizing machine translation

The major difference between machine translation on the Language Grid and a conventional translation system on the Internet is that users can themselves improve the quality of translation by using the Language Grid.

First, users can use the registered parallel texts while translating. As an example, let us consider the parallel texts by Kawasaki City Board of Education, registered in the Language Grid. When a user enters a sentence, examples with meanings similar to the entered sentence will appear automatically. If the user can find a sentence that suitably conveys the meaning that he/she intended, he/she could obtain an accurate translation result by using parallel texts.

If the user is unable to find the intended expression, machine translation could be executed. In this case, a dictionary registered by the user also helps to improve the quality of translation. For example, if the dictionary of school terms is registered, it will improve the result of machine translation.

In machine translation, English is often a hub language. For instance, the process of a translation from Japanese to Portuguese as follows: translation from Japanese to English and then English to Portuguese. The Language Grid realizes the composite services, such as the multi-hop translation explained above.

These features can be realized by the Language Grid, where combinations of various language services can be provided.
### Service Layers of the Language Grid

**Intercultural Collaboration Support**
- Multilingual communication is supported using various language services. NICT provides Language Grid collaboration tools.

**Language Service Composition**
- Multiple language resources are composed using Web service workflows.

**Language Resource Usability**
- Language resources are made usable as Web services with standardized interfaces.

**Language & Computation Resource Management**
- Allow users to connect to Language Grid servers on the Internet. Make language and computation resources accessible from all over the world.

#### Service Layers of the Language Grid

- **Language Grid services consist of four layers:**
  - **P2P Grid Infrastructure**
    - The infrastructure combines multiple servers on the Internet to fulfill users’ request. Users can add their servers to the P2P grid. Using the Language Grid Service Manager, users have access to usage statistics of the language resources and servers they provide.
  - **Language Resources**
    - Various language resources will be provided as Web services with standardized interface. Users can add new language resources.
  - **Language Services**
    - Language resources can be combined by Web service workflows. Various language service will be available, including back translations or specialized translations. Users can add new language services.
  - **Intercultural Collaboration Tools**
    - Collaboration tools are developed using language services explained above. New tools have been developed by NICT and universities, and moreover, existing tools have been multilingualized.
P2P Grid Infrastructure

- The P2P Grid Infrastructure is aimed at connecting servers around the world and combining language resources on the Internet. The P2P Grid consists of two kinds of servers (core nodes and service nodes).

- **Language Grid Core Node**
  - Core nodes manage all registered language services. They provide search and composition of atomic language services. Core nodes invoke other constituent language services based on Web service workflow.
  - Registered information of language resources is shared among all core nodes. The same services are equally available, regardless of which core node users access. The core nodes also control access to language resources.

- **Language Grid Service Node**
  - In the service nodes, language resources are deployed as Web services. The server provides various atomic services.
Language Resources

- To develop the Language Grid, we need to make language resources on the Internet available as Web services. In other words, we need to wrap language resources.
- Language service ontology is required for wrapping language resources. This standardizes interfaces of machine translations or dictionaries. The Language Grid Project is developing language service ontology in cooperation with DFKI in Germany.
- A wrapping manual based on the tentative language service ontology is available at Language Grid Association Web site (http://www.langrid.org/association/).
- The following language resources are currently available:
  - Morphological analyzers: Japanese, Chinese, Korean, English, German, Spanish, French, Italian, Dutch, Russian, Bulgarian
  - Bilingual Dictionaries: Life science terms (Japanese, English), Disaster management terms (Japanese, Chinese, Korean, English, French, Spanish), Academic terms (Japanese, English)
The following language services are available from the Language Grid:
- Atomic service: A Web service corresponds to each language resource. For example, bilingual dictionaries, parallel texts, morphological analyzers.
- Composite service: An advanced service described by a workflow for composing atomic services. For example, specialized translations, back translations.

The Language Grid provides multilingual services for intercultural collaboration.
- For example, specialized translation can be achieved using a composite service: several atomic services, such as machine translations, morphological analyzers, and specialized dictionaries are combined to create specialized translation.
- In the Language Grid, WS-BPEL is used to describe workflows which are interpreted and executed by a BPEL engine, which sequentially invokes Web services in the workflow.
**Intercultural Collaboration Tools**

Language Grid Toolbox is a **Web application** that uses language services on the Language Grid to provide various functions for supporting multilingual communities.

- **Translation**: Text translation, Web translation
- **Language Resource Creation**: Multilingual dictionary, multilingual parallel text, etc.
- **Community**: Collaborative translation, discussion based on contents sharing, etc.
- **Communication**: Multilingual BBS

**Text Translation**

Translation results can be re-translated into the original language for quality confirmation by using back translation.

**Multilingual Dictionary Creation**

Translation quality can be improved by creating and composing community dictionaries.

<table>
<thead>
<tr>
<th>Language</th>
<th>Japanese</th>
<th>Korean</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Grid Toolbox</td>
<td>Language Grid Toolbox</td>
<td>Text Translation</td>
<td></td>
</tr>
<tr>
<td>네이키다봉</td>
<td>네이키다봉</td>
<td>네이키다봉</td>
<td></td>
</tr>
<tr>
<td>Multilingual Dictionary</td>
<td>Multilingual Dictionary</td>
<td>Multilingual Dictionary</td>
<td></td>
</tr>
</tbody>
</table>

**Multilingual BBS**

Posted messages are translated into multiple languages and users can modify translation results for improving the quality.


- You can do it.
- Setting is on.
- It's possible to.
- Please use the.

**Intercultural Collaboration Tools**

- NICT has been developing the Language Grid Toolbox as intercultural collaboration tools, which can be easily accessed through a Web browser (Trial site: http://langgrid.org/tools/toolbox/). Language Grid Toolbox is developed in the framework of Service Grid Open Source Project. Anybody can participate in the development in the open source project community site (http://servicegrid.net/oss-project).

- Language Grid Toolbox provides a series of intercultural collaboration tools for supporting multilingual communities. Typical functions are as follows.

  - **Text Translation**: users can translate contents multilingually and confirm the translation result with back-translation. Users can also click any part of a sentence to highlight it and its translation/back-translation results.
  
  - **Multilingual BBS**: contents of the BBS are translated multilingually. Users can post messages using their native languages and modify the machine translation results to improve the quality of multilingual contents.
  
  - **Multilingual Dictionary Creation**: community members can create their multilingual dictionaries, and combine them with machine translators to improve the translation quality. Besides multilingual dictionaries, users can also create multilingual parallel texts, Q&As, glossaries and so on.
Federated Operation

- NECTEC (Thailand National Electronics and Computer Technology Center) launched Bangkok Operation Center from January 2011. The operation center covers south east Asia. Kyoto Operation Center started the federated operation of the Language Grid with Bangkok Operation Center.

- The federated operation enables users of both operation centers to share 106 services (51 languages) on Kyoto Operation Center and 20 services (13 languages) on Bangkok Operation Center. A user of one operation center can use services on another one.

- On the Language Grid of Bangkok Operation Center, various Asian language resources are available. The language resources include concept dictionaries (Indonesian, Korean, Sinhala, Thai, Nepali, Japanese, Hindi, Vietnamese, Bengali, Burmese, Mongol, Laotian), machine translators (English-Thai, Thai-Laotian), morphological analyzers (Thai), speech synthesis (Thai).

- The infrastructure software of the Language Grid is available as open source software. Since the agreement complies with the laws of the country where the Language Grid is operated, it is easy to internationally expand the federated operation.

- We will register various Asian language services and collaborate with language resource projects by EU and NSF (National Science Foundation) in the United States for international expansion of the Language Grid.