Extending utility of Language Grid through

Language Grid Toolbox

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

in

Software Engineering

Submitted by

Akanksha Upadhyay

(801031003)

Under the supervision of:

Dr. Seema Bawa — **Professor (CSED)**

Dr. Hitashi Lomash — **Assistant Professor (SBSBS)**

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY, PATIALA – 147004

June 2012

Certificate

I hereby certify that the work which is being presented in the thesis entitled, *"Extending utility of Language Grid through Language Grid Toolbox"*, in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Software Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Seema Bawa, Professor (CSED) and Dr. Hitashi Lomash, Assistant Professor (SBSBS).*

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Akanksha Upadhyay

Akanksha Upadhyay

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Seema Bawa 02/JUNE/2012

Dr. Seema Bawa
Supervisor
Professor(CSED)

Hitashi 2/6/12

Dr. Hitashi Lomash
Supervisor
Assistant Professor(SBSBS)

Countersigned by

(Dr. Maninder Singh)
Head
Computer Science and Engineering Department
Thapar University
Patiala

(Dr. S. K. Mohapatra)
Dean (Academic Affairs)
Thapar University
Patiala

Acknowledgement

I would like to express my gratitude and sincere thanks to my supervisor Dr. Seema Bawa, Professor, Computer Science and Engineering Department (CSED) for all her guidance, support and continuous encouragement throughout this thesis work. She has always been motivating and inspiring me to give my best. I am equally thankful to Dr. Hitashi Lomash, Assistant Professor, School of Behavioral Sciences & Business Studies (SBSBS) for sharing her insightful knowledge, providing timely assistance, invaluable time and suggestions which went long way in successful completion of this thesis work.

I would also thank to all of them who have been helpful and were associated with me directly and indirectly throughout the work. Finally, no words to thanks my dear parents and family members whose support and care encouraged me during this workout.

Akanksha Upadhyay
(801031003)

Abstract

This thesis explores Language Grid, a multilingual service platform, to increase the accessibility, customizability and usability of language services. It enables easy registration and sharing of language resources such as online dictionaries, morphological analyzer, and machine translations. It allows users to freely combine existing language services to develop new services so that users can easily create their own communication tools for their community as per their requirements. Many intercultural collaboration tools have been developed using Language Grid such as Language Grid Toolbox, Newsletter and Playground etc. Language Grid Toolbox, intercultural collaboration tools of Language Grid, is a multilingual tool which can be easily accessed through any web browser. It is a support system for the intercultural communication which offers a unique interface and increased usability.

Further thesis analyzes the features of Language Grid for intercultural collaboration and an effort is made to extend Language Grid Toolbox in which currently translation services among 15 languages are configured. Firstly, Toolbox is extended by performing extension and translation of terminology of Language Grid Toolbox by adding a dictionary service to the Language Grid Toolbox site, containing 400 words in English language and Hindi Language by using a dictionary service. Secondly, by enriching Q&A feature of Language Grid toolbox by adding various technical questions and answers to the Q&A Web Interface of the Toolbox by exploring Toolbox site adding questions and answers to the site by following the procedure on the site. These Q&A are verified and validated by the site administrator and are made available to all users whether registered to the site or not.

Table of Content

List of Figures

List of Tables

Chapter 1

Introduction

As multiple languages exist in world as a whole, there must be multiple language barriers. Several measures have been taken to overcome this language barrier by our society. Many multilingual societies generally have lingua francas and its members usually learn more than one language to remove these language barriers, but it is not possible in this way to remove this large existing language barriers. In olden days, since the late 1800s, auxiliary languages, called schematic languages, were traditionally written or constructed by a person or group, to overcome the language barrier. Some examples of traditional auxiliary languages are Esperanto, Ido, and Volapük [1]. Later in the first half of the twentieth century, a second approach to auxiliary languages, called naturalistic languages, emerged in which main focus was on those most widely spoken languages which had many words in common and these words could be developed into a simple language so that people in different countries, speaking these languages would understand this language when they read or heard it. Examples of the newer approach languages are Interlingua, Occidental, and Latino Sine Flexione [1].

The language and cultural barriers and misunderstandings can also get in the way of effective communication and can create complications in the workplace like systemic barriers which can have an adverse affect on certain groups. Taking a live example, a surprisingly large number of immigrants have considerable required education and experience in their native country, but because of their limited English speaking abilities and their cultural differences, their credentials, experience and knowledge may not be recognized. So it is important for the business community, workforce educators, trainers and other human resource to gain a better understanding of language barriers and cultural differences among workers from all backgrounds [2]. Such understanding and sensitivity will help employers increase recruitment, hiring, retention and advancement of immigrant workers and many more.

1.1 Language Grid

Several research groups including NICT, universities, and NTT have also started working to overcome this language barrier by developing a language infrastructure on the Internet called the Language Grid. The development of the Language Grid was started in April 2006, led by Professor Toru Ishida, Department of Social Informatics of Kyoto University. NICT Japan is promoting Language Grid for intercultural collaboration all around the world. The Language Grid is an infrastructure that is built on the top of the Internet. It allows a better understanding of Internet contents written in different languages by different country's people. It allows users to easily develop new language services by combining existing ones to satisfy their needs [3].

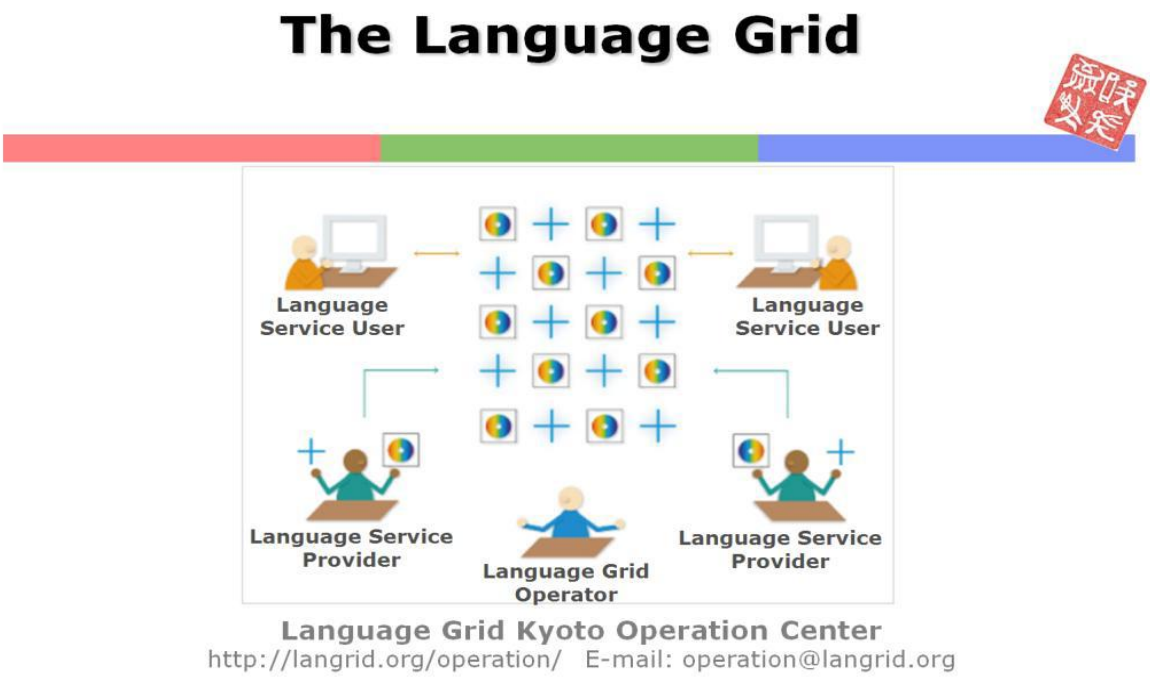

Figure 1.1: Language Grid Operators [4]

Language Grid customizes output of users' language resources and machine translators to produce high-quality translation. It has three types of stakeholders as shown in Figure 1.1. First

is Language Grid Operator, which manages Language Grid Users and controls language and computation resources. Second is Language Resource Provider which provides registered language resources to the Language Grid. And the third is Language Service User which is the Organizations or individuals who use language and computation resources of Language Grid [4].

To overcome language barriers and cultural barriers Language Grid has been developing many collaboration multilingual tools. Among several developed collaboration tools based on the Language Grid infrastructure, the Language Grid Toolbox is an open source communication tool, developed by National institute of Information and Communications Technology.

1.2 Thesis Framework

This section discusses the framework of this thesis. This thesis is organized as follows:

Chapter 2: Literature Review
In this chapter various language barriers and approaches to overcome this language barriers by busing different collaboration tools and their multicultural activities is discussed.

Chapter 3: Problem Formulation
In this chapter main objectives of this thesis are covered.

Chapter 4: Introduction of Language Grid
An introduction of Language Grid is given to easily understand what it is and how it helps to our society.

Chapter 5: Introduction to Language Grid Toolbox
An introduction of Language Grid Toolbox is given to easily understand about its various features and functioning.

Chapter 6: Extension and Translation of Language Grid Toolbox Terminology in Hindi and English
In this chapter a dictionary service is added to the Language Grid Toolbox site containing 400 words in English language and Hindi Language.

Chapter 7: Enriching Q&A Feature of Language Grid toolbox

In this chapter a description of adding various questions and answers to the Q&A Web Interface of the Toolbox by exploring Toolbox site is given.

Chapter 8: Conclusion and Future work

This chapter deals with the importance of Language Grid and Language Grid Toolbox and also tells what else can be done to extend the Toolbox site.

Chapter 2

Literature Review

Language barriers the biggest barrier for intercultural collaboration and to overcome this barrier many languages services are presently available but these Language services are often not accessible, because of their high prices, low customizability and service quality and many more reasons. So to increase the accessibility and usability of language services, language grid is proposed [5]. It takes the service-oriented type of approach by wrapping existing language resources as atomic Web services and enables end users to compose new language resources and services.. Many organizations, institutions, industries and others forming Language Grid Associations are working cooperatively on Language Grid to enhance it.

2.1 Language Grid Association

The Language Grid Association is a loosely coupled organization formed by collaboration among industries, governments, academia, and citizens to advance the technology and application of the Language Grid [30]. This organization consists of various SIGs (Special Interest Groups) whose aim is to accumulate use cases and best practices for Language Grid better usability. Language Grid Association shows number of countries, organizations which are using and helping in the development of the Language Grid and Language grid Toolbox.

2.2 Language Grid Users

With the Language Grid various language resources, through collaboration tools such as Langrid Chat and Langrid Blackboard [13], are developed by NICT, as well as by other many users of Language Grid which are used by many users. Figure 2.1 shows some major users of Language Grid.

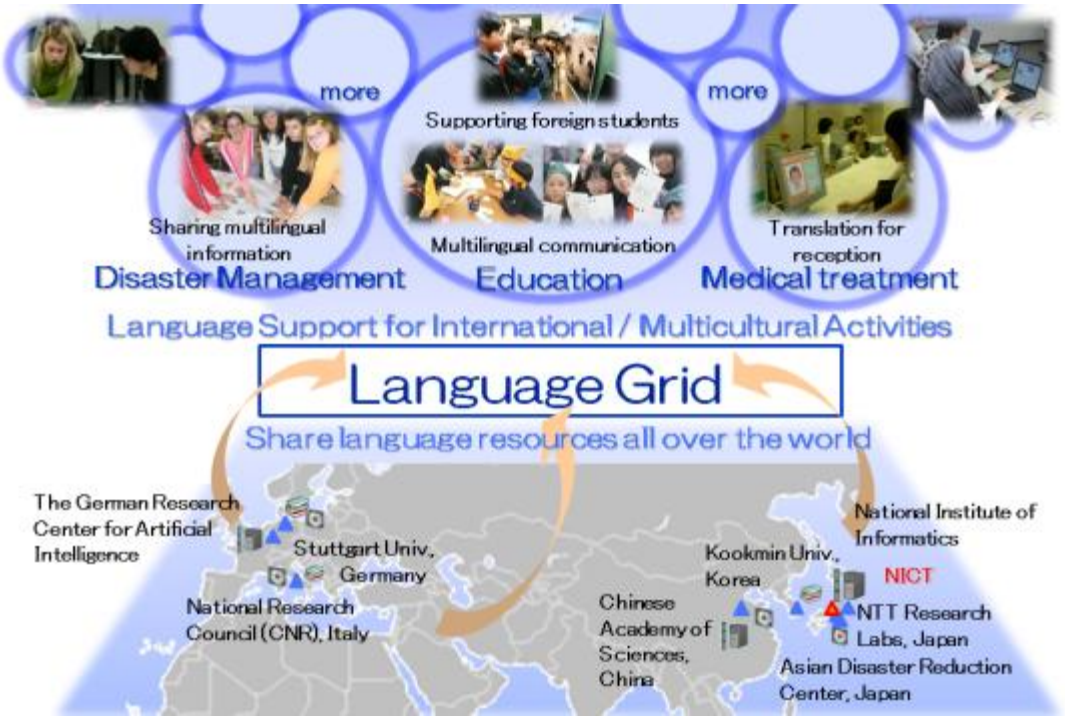

Figure 2.1: Language Grid users [14]

2.3 Collaboration Tools of Language Grid

Among various tools of Language Grid example tools include multilingual chat tool called AnnoChat, collaboration tool called Multilingual NOTA, the Langrid Chat and Langrid Input, and many more. A multilingual tool provided by NICT is used to support communication among the international staff, other tool which is computer mediated multilingual medical communication support system called M3 [6] is a Center for Multicultural Information and Assistance. Kyoto started this medical interpretation program from September 2003 to assist foreign patients. Pangaea creates a "universal playground" for kids all around the world. This organization develops communication packages for kids to feel their "bonds," and use pictograms created by kids for communication. Kyoto Community Broadcast also started an FM station in March 2003 [5].Figure 2.2 shows some of the Intercultural Collaboration Tools.

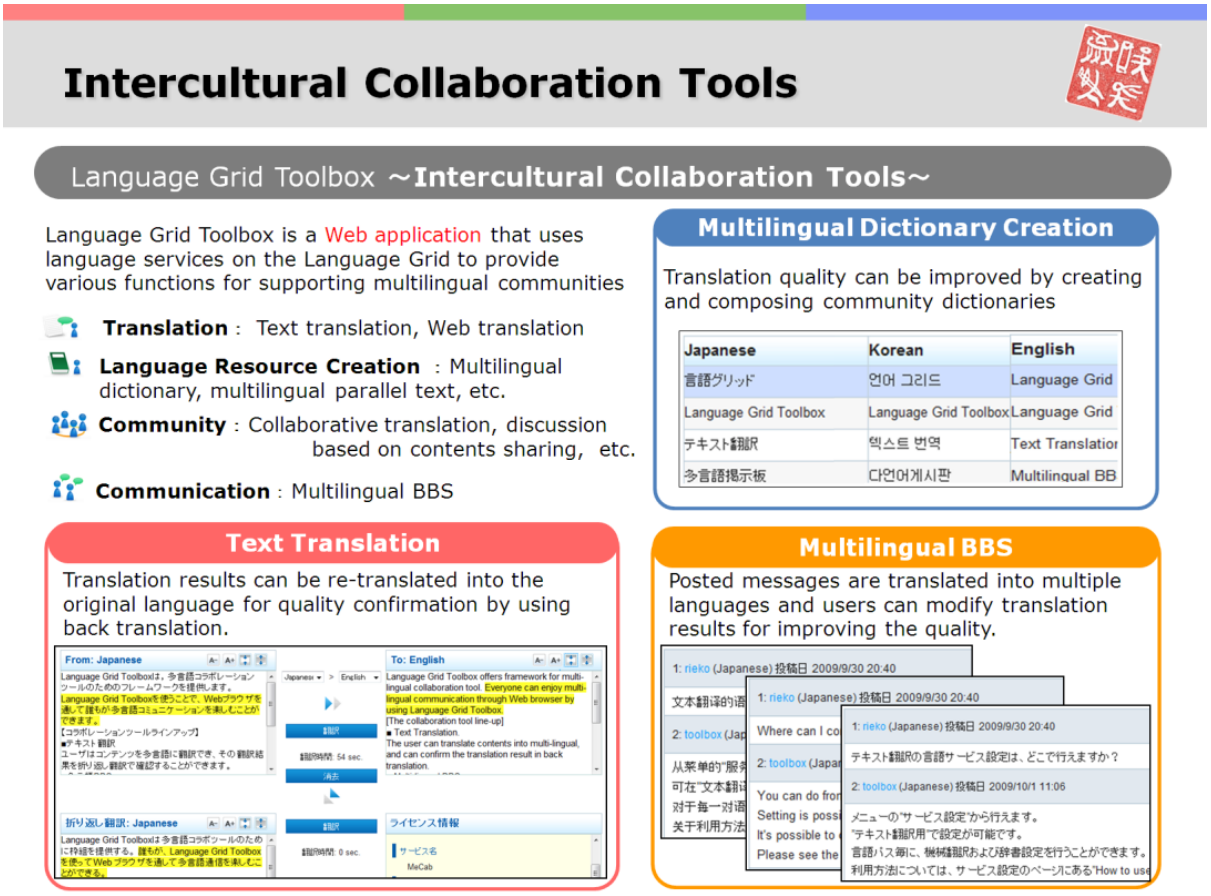

Figure 2.2: Intercultural Collaboration Tool [4]

2.4 Multicultural Local Communities

In multicultural local communities, the languages spoken contain a lot of specific words, which machine translators do not know while translating them. For instance, take an example between English language speaker and Ukrainian language speaker. On entering the line “what are you doing today at 9 am” in English and translate it into Ukrainian as Ukrainian peoples do not understand English, so they are unsure if the translation is correct. They perform back translation, i.e., they translate the Ukrainian back into English and the following sentence is displayed “What to do today at 9 am” This sentence seems to be of different meaning than the original meaning; however, it may be acceptable. Now, if translated into Turkish in the same manner, the following Turkish equivalent “Eğer 9 bugün ne am yapıyorsun” would appear on the screen. When back translated the Turkish sentence into English again, a very strange sentence is observed “I am doing today if 9”.

So it is seen that machine translation works only to a level. Even if translation quality is increased by the development of community specific language resources, all communication

problems through translation is not solved; it require to have knowledge of different cultures to reach an assured mutual understanding, thus a system is required that associates machine translation results with various interpretations of concepts to help in understanding different cultures.

2.5 Languages in Language Grid and Language Grid Toolbox

Total one hundred and forty four organizations from seventeen countries have participated in the Language Grid. The Language Grid is providing total one hundred and six language services [31]. Language Grid Toolbox is working currently on fifteen languages which are as follows:

1. Arabic
2. Chinese
3. English
4. French
5. German
6. Indonesian
7. Italian
8. Japanese
9. Korean
10. Malay
11. Portuguese
12. Russian
13. Spanish
14. Thai
15. Vietnamese

Chapter 3

Problem Formulation

During literature review, following research gaps are observed:

3.1 Research Gaps

i) Combining human and machine translators for supporting multilingual communication involves much difficulty.

ii) The gap between human translators and machine translators remains huge especially for the domain of localization processes that requires high translation quality.

iii) High translation cost of several translation processes, including absolute machine translation processes, absolute human translation processes and translation processes by human and machine translation services.

iv) Machine translators have limitations in translation qualities for translating a specific document with high requirement of qualities in some specific language.

v) Bilingual human translators are not available everywhere for any purpose at any time in the real world and the cost of translations of highly-trained bilingual individuals are always high in both labor and time.

vi) Ambiguity in language observed is an obstacle in multilingual communities.

vii) Difficulty for People from different cultures and language collaborate to meet the needs of a more and more globalized society is very great.

viii) There are still cases where machine translation errors cause misconceptions in multilingual collaboration.

ix) The biggest obstacle for people working in different cultures and countries has always been the language barrier.

3.2 Problem Formulation

This thesis targets to some of the points observed during research gap analysis which includes reducing gap between human translator and machine translator, overcoming some problems of the machine translation by providing free language resources and finding ambiguities in multilingual communities, by Language Grid and Language Grid Toolbox.

3.3 Objectives of Thesis

The objective of the proposed work is:

i) To study and analyze the features of Language Grid for intercultural collaboration.

ii) To study and analyze various features of Language Grid Toolbox for customizing the Toolbox site.

iii) To extend Language dictionary service to include Hindi words to promote Indian culture.

iv) To enhance Language Grid Toolbox by adding more technical questions and answers.

3.4 Methodology

i) a) Study of Language Grid, a multilingual service platform on the Internet.

ii) a) Study of, improving the localization processes by composing human and machine translation services based on the Language Grid.

iii) b) Study of Language Grid Toolbox to identify how to increase quality of translation.

iv) a) A specific document on Language Grid is considered.

v) b) Adding of free resource to the Language Grid Toolbox in the form of a English language to Hindi language dictionary containing 400 words related to Language Grid.

vi) a) Study of some basic features available on the Language Grid Toolbox site through which different people of different culture can interact with each other like Multilingual BBS, Multilingual dictionary etc

vii) a) With the increase of information available on the Internet, a problem with understanding and processing of information in multiple languages

becomes apparent, so study of increasingly important Development of supporting systems for international collaboration.

viii) a) Study of how Language Grid has emerged as one solution to overcome the language barrier both in online and real world multicultural collaboration

Chapter 4

Introduction of Language Grid

In this era of Information, Communication and Technology (ICT) the importance of language has been realized more than ever. Language is a medium of communication, an expression of thought that we require in our daily life and also is the base tool used in information technology. Language is a feature of human civilization which has another dimension too that it acts as a barrier. Identifying the function of language as a barrier, many service providers in the field of IT have attempted to solve the issue by providing language platforms on the internet. Nevertheless, these service providers either because of their high prices, inaccessibility and intellectual property rights have not been able to come up with an optimized solution.

Difficulties often arise while trying to use those language services in their communities, as per the survey, the cost of language services is very high. Machine translation costs around ten thousand USD per language pair per year. Complex contracts, Intellectual Property Rights (IPR) [7], and non-standard application interfaces make it difficult for users to customize language services in support of their activities.

To overcome the issue of inaccessibility and non usability of language services, language grid is proposed which treats existing language services as atomic components and enables users to create new language services by combining appropriate components [8]. It allows users to freely combine existing language services to develop new services so that users can easily create their own communication tools.

With the development of the Internet environments, several machine translation tools have become available on the Web, which provides the opportunity for some multilingual communities to use machine translation for communication [9][10] like Google translation, Lexicon Translations, Atalanta Translation Services etc. However, there exist several problems in machine translation tools available on the Internet.

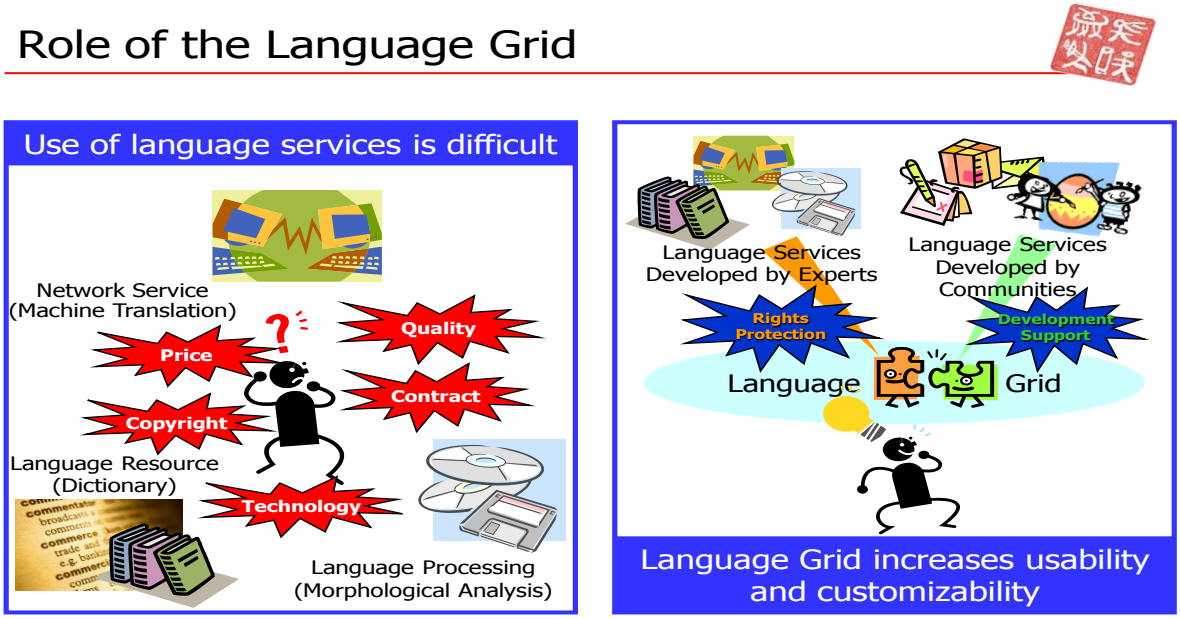

Figure 4.1: Role of Language Grid [11]

The major problem encountered is the low quality of translation regarding specific terms or sentences within communities. Most of the available machine translation tools only provide basic functions of translating sentences, documents or Web pages, and lack necessary grammatical and contextual functions for community users. With such translation tools it is difficult to conduct collaborative work among community users. There is no support for creating customized multilingual environments based on unique community requirements. To overcome such issues in communication and translation, Language Grid as a platform can be looked up. Figure 4.1 shows the graphical comparison of language services available on the internet and language services provided by the Language Grid.

4.1 Definition of Language Grid

The word “grid” of “Language Grid” is generally defined as “a system or structure for combining distributed resources; it is an open standard protocol generally used to create high quality services [12]”. The Language Grid is a middleware with which people can connect and use language resources such as machine translations, morphological analyzers created in the fields of intercultural collaboration. The unique and important aspect of Language Grid is that it cannot exist without the collaboration of Language Grid Users [13]. The interesting point here is that with the addition of every Language Grid users, who provide language and computation

resources, language services and collaboration tools, the grammatical and contextual aspect of the Language Grid is enhanced.

Language Grid is a platform that enables users to share language resources and supports activities for intercultural collaboration. Research institutes and universities which are participating in the Language Grid, provide language resources such as dictionaries, parallel texts, or machine translators for free of cost. While providing these language resources, providers can also specify copyright notices and license information on the profiles of there provided sources and can also set restrictions on users' access to these resources. Some of the organizations that had provided their language resources to the Language Grid are Chinese Academy of Sciences; National Research Council, Italy; Stuttgart University; National Institute of Informatics, Japan; NTT Communication Science Laboratories; and Asian Disaster Reduction Center. These language resources provided by users, NPOs, schools and other non profit sectors have started playing a central role in breaking the language barriers up to a large level. Their activities cover a broad range of fields, including disaster management, education, and medical care etc.

4.2 Technologies of Language Grid

To meet the various goals of Language Grid, different technologies are used by NICT. To start with Semantic Web technologies are developed to enable the collaboration needed among language resources and language processing functions. Language resources include bilingual dictionaries, thesauruses, and corpora, while language processing functions include machine translation, morphological analysis, and paraphrase. The technologies of Language Grid include Language service ontology which is a technology to define language service entries in a standardized way. Collaboration with European research institution advances the standardization of language service ontology. Also, Semantic Web service is a technology is used for Web services via standard methods. The scenario description language developed in Kyoto University will be used to describe interaction among composite services.

Language service ontology and semantic Web services make it possible to combine bilingual dictionaries and machine translations all over the world. For instance, to realize Japanese-German translation, combination of atomic services such as Japanese-English and English-German translation is required [13].

4.3 Language Grid service layer

As of now, brief outline about the Language Grid, its goal and technologies are discussed; the following section would be an attempt to describe the model commonly known as Language Grid Service Layer. Language Grid services consist of four layers which are P2P Grid Infrastructure, Language Resource, Language Services, Intercultural Collaboration Tools.

P2P Grid Infrastructure is the infrastructure which combines multiple servers on the Internet to fulfill users' request. This layer organizes multiple servers, language grid core nodes and language grid service nodes on the Internet to fulfill end users requests [30]. Language Grid users can add their servers to the P2P Grid, and access to usage statistics of their resources [16]. Using the Language Grid Service Manager, Users have access to usage statistics of the language resources and servers they provide.

Language Resource layer provides various language resources as Web services with standardized interface [17]. Users can also add new language resources. To increase usability of the language resources added for language services, standardization of access entry is very important.

Language Services layer consist of Language resources combined by Web service workflows. Various language services will be available, including back translations or specialized translations. Users can add new language services.

Intercultural Collaboration Tools are developed using language services explained above. New tools have been developed by NICT and other universities, and moreover, existing tools have been multilingualised. All four layers are clearly shown in Figure 4.2.

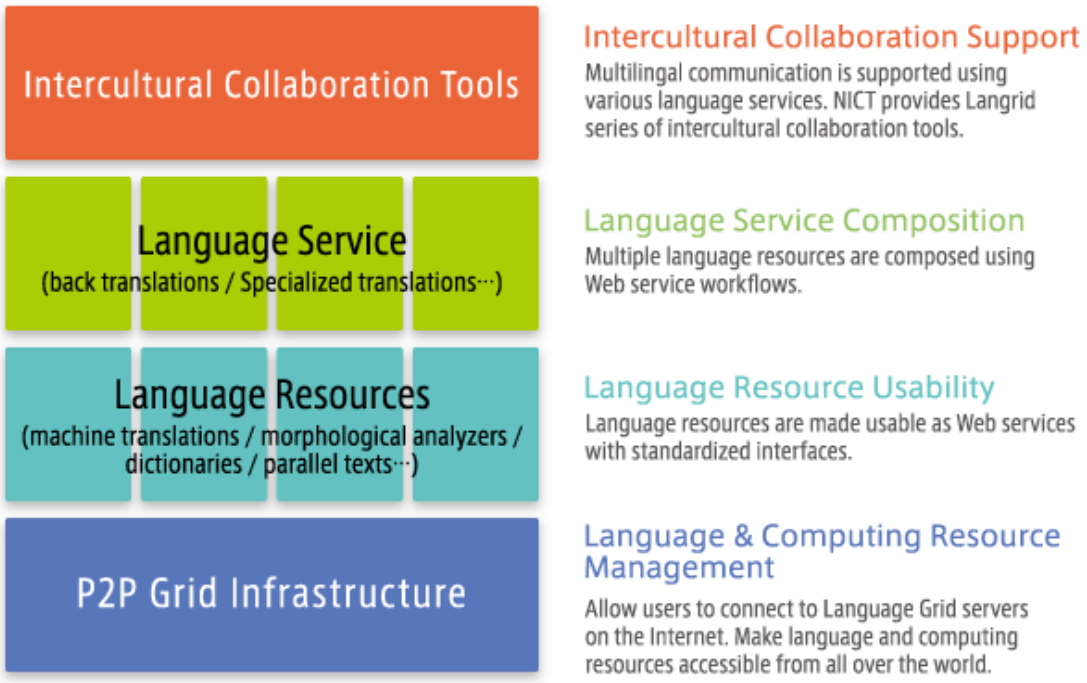

Figure 4.2: Language Grid service layers [18]

4.4 Application of language grid using toolbox

While using Toolbox, one distinctive feature was observed, in the collaborative translation tool on the website "http://langrid-tool.nict.go.jp/toolbox/", using this site when a language is translated into another language and when back translated into original one, the meaning of the context is completely changed**.** Figure 4.3 and Figure 4.4 describe the translation of text written in a) English to Arabic and back translation in English and b) English to Japanese and back translation using Language Grid Toolbox. It is observed that when Japanese context is back translated to English language, original meaning of the context is totally changed. However when back translated from Arabic language, context remains the same. This shows that as compared between Japanese and Arabic more work is needed for Japanese language for translation services. Similarly the amount of work needed to correctly translate an English text to any other language available in the toolbox varies a lot. The issue here is that the basis of Language Grid is to provide the language services with the best translation. It can be said that existing language services do not fulfill all requirements of translation from one language to other language; more work is required to optimize the result of translation services. This point shows that there is an area which needs to be further researched for better translations and more optimized results.

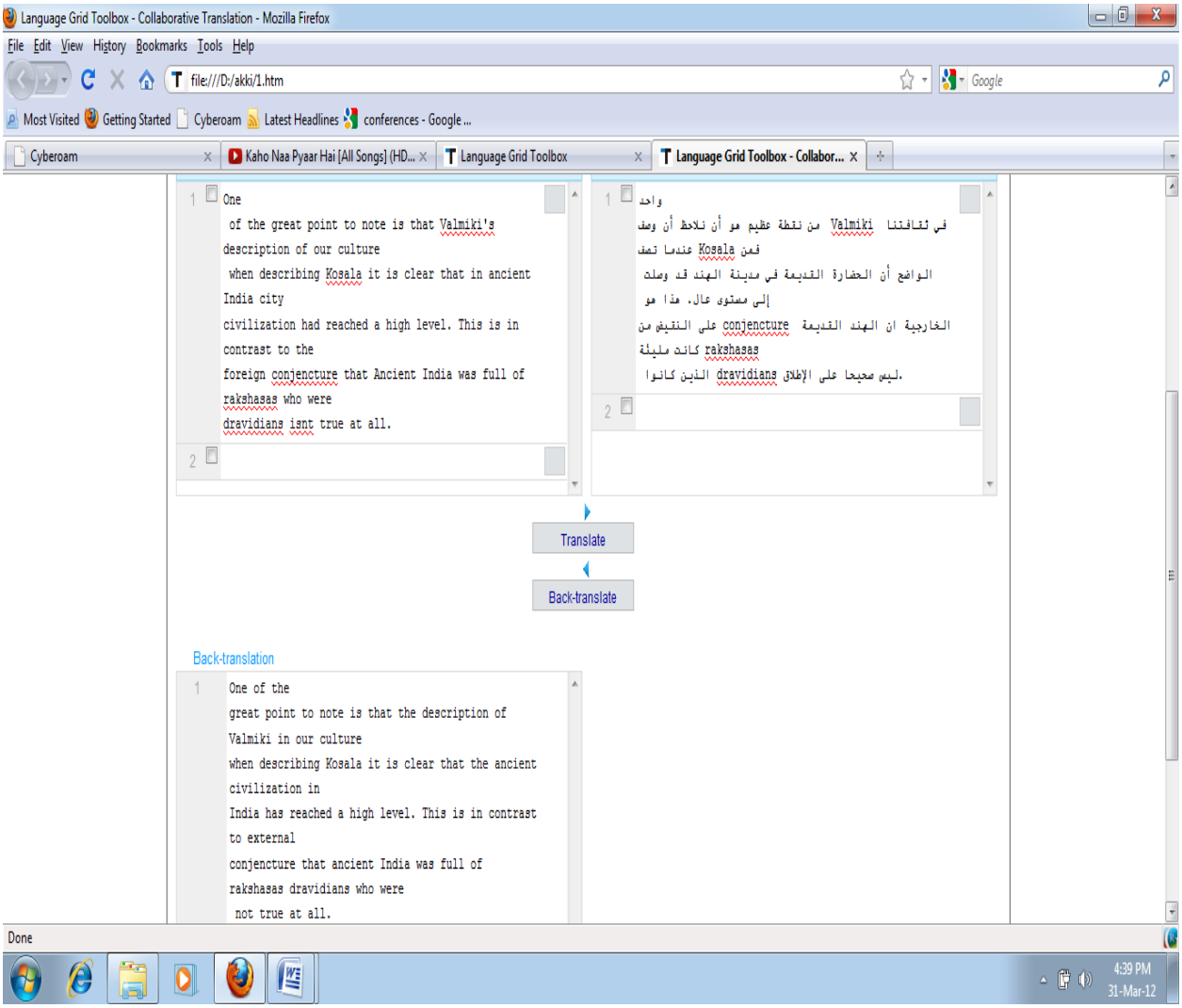

Figure 4.3: Translation from English to Arabic and back translation

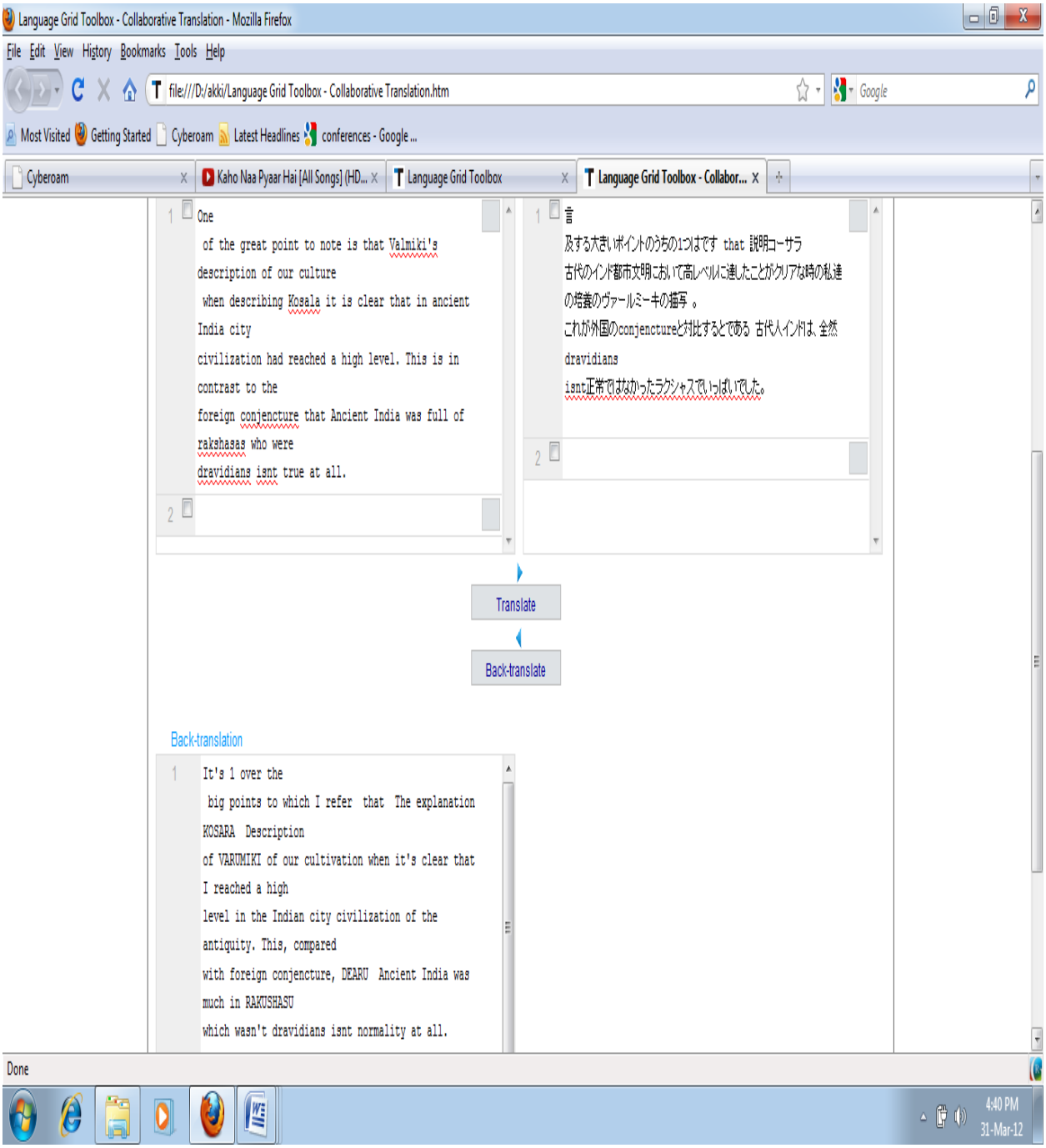

Figure 4.4: translation form English to Japanese and back translation

Original paragraph and back translated paragraph from Arabic and Japanese are shown in Table 4.1 and Table 4.2.

Table 4.1: Back translation from Arabic language to English language

ORIGINAL PARAGRAPH	BACKTRANSLATED PARAGRAPH FROM ARABIC
One of the greatest point to note is that Valmiki's Description of our culture when describing Kosala it Is clear that in ancient India city civilization had reached a high level. This is in contrast to the foreign conjenctures that Ancient India was full of rakshasas who were Dravidians isn't true at all.	One of the greatest point to note is that the description valmiki in our culture when describing Kosala it is clear that the ancient civilization in India has a high level. This is in contrast to external conjencture that ancient India was full of rakshasa Dravidians who were not true at all.

Table 4.2: Back translation from Japanese language to English language

ORIGINAL PARAGRAPH	BACKTRANSLATED PARAGRAPH FROM JAPANESE
One of the great point to note is that Valmiki's Description of our culture when describing Kosala it Is clear that in ancient India city civilization had reached a high level. This is in contrast to the foreign conjenctures that Ancient India was full of rakshasas who were Dravidians isn't true at all.	It's 1 over the big points to which I refer that the explanation KOSARA description of VARUMIKI of our cultivation when it's clear that I reached a high level in the Indian city civilization of the antiquity. This compared with foreign conjencture, DEARU Ancient India was much in RAKUSHASU which wasn't Dravidians isn't normality at all.

From the above discussion, it can be concluded that there are still problems such as miscommunications often occur when a machine translation is used because of nonexistence of machine translate which can translate original message into the other language perfectly

including the nuances of it. In, order to prevent the mistakes in communications by using a machine translation Language Grid can be used [19].

In the framework of the Language grid new services are constructed called language services for a multilingual community which requires some multilingual supports. Language services are realized by combining various language resources researched, developed and collected around the world. When building language services the Language Grid does not modify any language resources themselves but makes it easy to construct new language services by combining them appropriately. The Language Grid aims the realization of easy support to multilingual communication by constructing a language infrastructure on which such things is possible. That is to say, creating new language services necessary to multilingual communication on the Language Grid realizes easy multilingual communication support. Through Language Grid two kinds of software are introduced, the service grid server software and the Language Grid Toolbox, and code for both are available via open source licenses [20].

Chapter 5

Introduction of Language grid Toolbox

Originally, the Language Grid project was initiated as a reaction to the current situation in language services, currently quite a few languages services exist, however, they often have less usability and no flexibility to add new words, to customize domain specific expressions or sentences, and to combine several services. The Language Grid Toolbox shown in figure 5.1 facilitates multi-language-mediated collaboration.

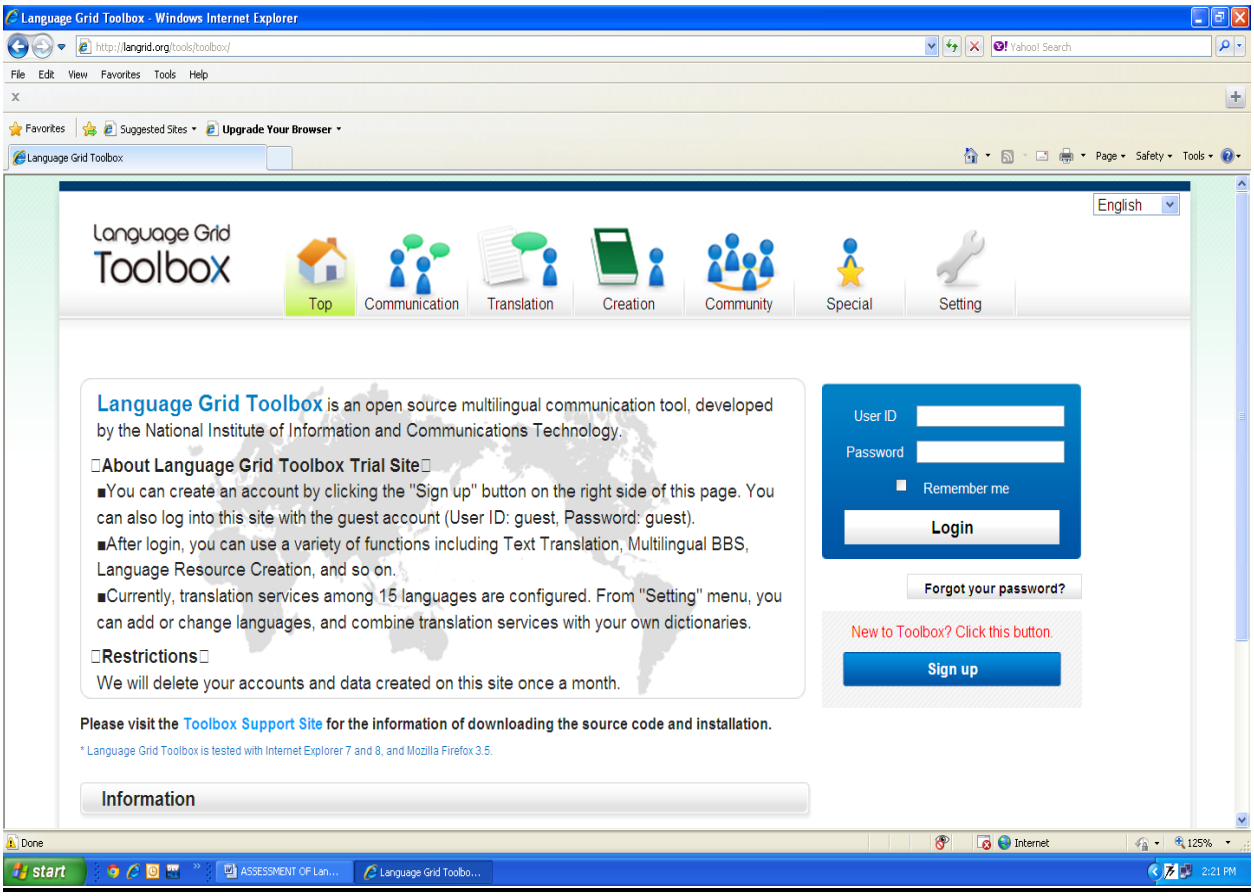

Figure 5.1: Language Grid Toolbox [21]

Language Grid Toolbox is a series of multilingual communication tools which use the Language Grid to enhance the multicultural and multilingual activities. Toolbox is present as open source software, which can be easily extended by developing modules by user communities to meet different requirements for intercultural collaboration and can be enhanced to increase its usability. There are two types of Toolbox modules in the Toolbox, basic modules that implement the basic functions for multilingual collaboration, and community modules that provide special functions based on the community requirements [24].

Language Grid Toolbox is a collection of various modules to support multilingual communication in a community. Users can install this Toolbox software on their servers to offer services, such as multilingual Bulletin Board System (BBS), multilingual dictionary creation, Q&A interface, and reception service and so on. As Toolbox is provided as open source software, the functions of toolbox can be extended to meet requirements of user communities. Furthermore by using registered language services, existing communication tool can introduce multilingual function easily and with increased usability.

The number of online machine translation tools continues to expand, so the importance of utilizing machine translation in multilingual communities is also increasing. Yet various problems exist when using the existing machine translation tools with intercultural communication in a multilingual community like translation of the community-specific terms or sentences within communities is always of low quality, machine translation tools lack a view of improving these low-quality translation of multilingual community's activities, and furthermore, they do not provide the means for customization based on the requirements unique to a community. To address these issues, Language Grid Toolbox is developed. Language Grid Toolbox aims to support the intercultural collaboration using the Language Grid and provides several functions such as creation of community-specific dictionaries combined with a machine translator and multilingual BBS, where translated language can be corrected collectively by community members. Moreover, since Toolbox is developed as open source software and provides APIs of basic functions, customized functions for each community can easily be developed. Several customized communication tools extended from Toolbox's basic modules have already been implemented by universities and local governments [22].

5.1 Architecture of Language Grid Toolbox

Figure 5.2 shown below illustrates the architecture of XOOPS-based Community Software. In the figure, the architecture is divided into three layers. XOOPS Core is responsible for the fundamental functionalities such as user management, authentication and relies on the original code of XOOPS. Toolbox APIs provide basic functions of multilingual communication tools. They are used to support development of multilingual communication tools on Toolbox. Toolbox Modules are applications for end users. Since requirements of multilingual communication vary from community to community, they are designed or customized to meet requirements of each community [23].

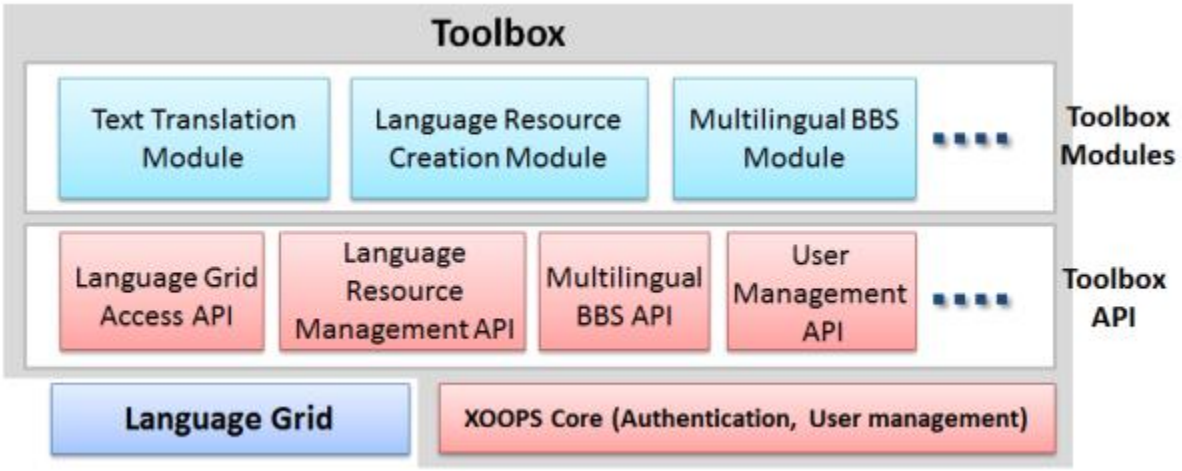

Figure 5.2: Architecture of Toolbox [24]

5.2 Features of Language Grid Toolbox

Language Grid Toolbox provides a series of multilingual collaboration tools. With Language Grid Toolbox, anyone can enjoy multilingual communication through any web browser present. Toolbox provides basic functions for multilingual communication including BBS, text translation, and dictionary and so on.

5.3 Environment of Toolbox

Language Grid Toolbox is developed based on XOOPS Cube. The environments for Toolbox are shown as follows:

- OS: Linux / Windows / Mac OS.
- Web server: Apache is recommended.
- PHP: PHP 5.3. php-mbstring, php-pear, php-soap, php-dom (not contained in default configuration of some distributions, such as Fedora Core) are required.
- Database:MySQL. Versions later than MySQL4.1 are recommended.
- Postfix: Required for sending passwords to users via E-mail.

Toolbox is tested in the following environment:

- Apache 2.2.3
- Php 5.3
- MySQL 5.0.77
- postfix 2.3.3-2.1.el5_2

Following machine specification is recommended:

- CPU: more than 1GHz
- Disk space: more than 120GB
- Memory: more than 1GB [25]

5.4 Design Concept of Language Grid Toolbox

Toolbox is based on the open source environment CMS (Contents Management System) XOOPS Cube so the functions of Toolbox can be extended by developing XOOPS modules to meet the requirements of any community. The source code of Language Grid Toolbox is opened as GPL. Now before going further, it is better to understand what XOOPS is.

5.5 XOOPS

XOOPS is a free, open content management systems (CMS), written in PHP. It uses a modular architecture allowing users to customize, update their websites. XOOPS is released under the terms of the GNU General Public License (GPL) and is free to use, modify and redistribute. XOOPS is an acronym of "extensible Object Oriented Portal System". It aims to serve as a web framework for use by any sites, through the installation of modules. For example, a small XOOPS installation can be used as a personal web log or journal, but this can be expanded upon and customized, such as users might add the appropriate modules. Constantly updated and improved, it draws praises from reviewers [26].

5.6 XOOPS Cube

XOOPS Cube is an Open Source Web Application Platform, empowering everyone to create dynamic and content rich websites with ease. It is highly compatible with XOOPS, allowing users to implement and enjoy their XOOPS modules, while using the new and improved Object Oriented Web Application Platform - XOOPS Cube. XOOPS Cube is the ideal tool for developing community websites, intranet portals, corporate websites, web blogs, or whatever is needed. Concept of XOOPS Cube is Simple, Secure, Scalable (3S) [27].

5.7 Versions of toolbox

Following are the versions of Toolbox:

- Toolbox Ver 3.0 (2010/4): The development of Toolbox is managed by the open source project from April 2010.
- Toolbox Ver 2.3 (2010/03): In this version three new features were added in Toolbox which are: Q&A Web interface, Collaborative translation, Translation task management
- Toolbox Ver 2.2 (2010/02): New feature added in this version was Discussion
- Toolbox Ver2.1 (2010/1): New main function: Language resources (Q&A/Glossary text),
- Toolbox Ver2.0 (2009/11): Text translation, Multilingual BBS , Language resources (Dictionary / Parallel text), Service settings , User management [28].

Currently the following two versions of Toolbox are available:

I. XOOPS-based Community Software: This version of Toolbox aims at supporting communication in a multilingual community. This version is simply called "Toolbox". It is developed on the open source CMS (Contents Management System) XOOPS Cube, this version is also provided as open source software. The functions can be extended by developing XOOPS modules to meet the requirements of any specific community.

II. Language Grid Extension for Media Wiki: Language Grid Extension for Media Wiki enables the other extensions of Media Wiki to invoke translation services on the Language Grid. After the Language Grid Extension for Media Wiki is installed, two new tabs will appear in each article, "Setting" and "Page Dictionary". In the "Setting" tab, end users can select translators and dictionaries; in the "Page Dictionary" tab, end users can create a dictionary to be combined with a translator [23].

5.8 Basic modules of Toolbox

Language Grid Toolbox is designed for the purpose of supporting community users to create their own multilingual community sites; it provides the basic functions based on the most important requirements in multilingual communities including the language resource creation, translation, communication and collaboration functions and many more.

A. Language Resource Creation*:* Creating and sharinglanguage resources are very important for communities to conduct multilingual communication and collaboration for improving Quality of machine translation services. To provide easy customization environments for communities to use the Language Grid, Language Grid Toolbox provides the language resource creation as basic function that enables community users to create and share dictionaries, parallel texts, Q&As, glossaries, web, translation templates.

B. Communication: Multilingual communication is another important requirement when creating multilingual community sites. Language Grid Toolbox provides a basic communication called multilingual BBS that enables community users to communicate with each other with their mother tongues; since the contents of BBS are translated multi-lingual by language services on the Language Grid. Like the basic translation function above, community users can also customize the usage of translation services.

C. Translation: Translation feature of toolbox includes text translation and web translation. Text Translation enables to translate text and estimate the translation results by referring to the back-translation results whereas Web Translation enables to translate Web pages.

D. Community: Community feature of Toolbox include User list, File sharing, task management, collaborative translation, discussion. Profile can be seen by using User list feature of Toolbox. File sharing enables users to share files with other users. A user can upload his/her file to a category which is created by the administrator. Any user can download the uploaded file. With Task Management function, document translation tasks are managed. By Collaborative Translation function, users can collaborate on document translation by recording the task's status. A variety of translation support functions are also available such as Progress of translation tasks can be recorded in the collaborative translation function. The status of each unit in the document for translation can be set as "not modified", "in progress" or "finished". By sharing the progress status of the translation tasks among community users, the whole tasks can be appropriately assigned to multiple users for collaborative work and the last feature of Community is Discussion which provides the discussion function among users by sharing contents such as images, maps and so on.

E. Special: Special module of the Toolbox contain reception, with the Reception function, the Q&As stored in the Toolbox can be displayed on the computers at the reception counter in stores. The reception function is developed to support communication using multilingual Q&A registered in Toolbox, which is always used to support customers in the front desk at shops. Customers can select the Q&A category and questions to find answers. Moreover, the reception function is composed with the glossary registered in Toolbox. When the words registered in the glossary are included in answers, the meaning of the words can be displayed by hovering the mouse cursor. Another feature of Special module is Q&A web interface The Q&A Web interface function is developed by combining the multilingual Q&A creation function and the multilingual BBS function. By using the Q&A Web interface function, Toolbox community users can collaborate to create Q&A contents for users outside the community. The Q&A Web interface site enables users that do not belong the Toolbox community to search Q&As in Toolbox and post new questions. When a new question is posted, it will be automatically registered as Q&A in Toolbox, and added as a new topic in the multilingual BBS. Answers will be created based on the discussion among Toolbox users. Created answers can be searched by the Q&A Web

interface if they are added into Q&As in Toolbox. And other are showroom BBS and showroom discussion.

F. Setting: This module includes all setting features of Toolbox site.

All the modules of the Language Grid Toolbox site can be customized according to the user's requirement. In next two chapters an effort has been made to enhance the Toolbox by adding a dictionary service from English language to Hindi language and by adding some technical questions and answers to the Q&A Web Interface module of the toolbox.

5.9 Goal of Language Grid Toolbox

Although the Language Grid offers a lot of language services, an application is needed to support users in a multilingual community and the Language Grid Toolbox is that application, a communication tool for intercultural collaboration. The goal of Toolbox is to provide customizability to support a wide variety of collaboration in multilingual communities. Customization for multilingual communities is classified into "Customization of language resources" for improving translation quality and "Customization of functions" for facilitating collaboration. Toolbox focuses on supporting collaboration in a multilingual community. The former is achieved by a function which allows community members to accumulate translations of particular terms and expressions of the community. The translation quality generally declines if the input sentences contain particular terms and expressions of a community because a machine translator for general purposes cannot correctly translate such terms and expressions. Therefore we improve translation quality by combining the translator with dictionaries which contain terms and expressions accumulated in a community.

Chapter 6

Extension and Translation of Language Grid Toolbox Terminology in Hindi and English

6.1 Resource addition in language Grid Toolbox

Language Grid Toolbox (also referred to as Toolbox) provides basic functions like language resource creation, text translation, multilingual BBS and Web creation and many more that are required for multilingual communities. All the functions are realized by using language services on the Language Grid, such as machine translation services, dictionary services, and morphological analysis services and so on. By using the basic function of creating language resources like dictionaries in Toolbox, community specific terms can be registered in dictionaries, which can be used to combine with atomic translation services to improve the translation quality. In multilingual BBS, messages can be exchanged among community users and translated into other languages. Web creation function enables users to register translation result of Web contents as templates and share those templates among community users. Toolbox is developed as open source software and provides APIs for basic functions. By this means, customized functions for communities can be easily developed. Several customized functions have already been implemented by universities and local government and embedded into Toolbox. In this chapter a dictionary service from English language to Hindi language is added to the Language Grid Toolbox.

6.2 Creating a dictionary service in Toolbox

The Dictionary function enables to create and manage dictionaries for the purpose of improving translation quality. A dictionary is created by clicking the "Add" button as shown in figure 6.1. The created empty resource can be updated by clicking "Add record" button, "Delete record" button, "Add language" button and "Delete language" button. Double-clicking on an empty cell will prompt a text field for inputting the item.

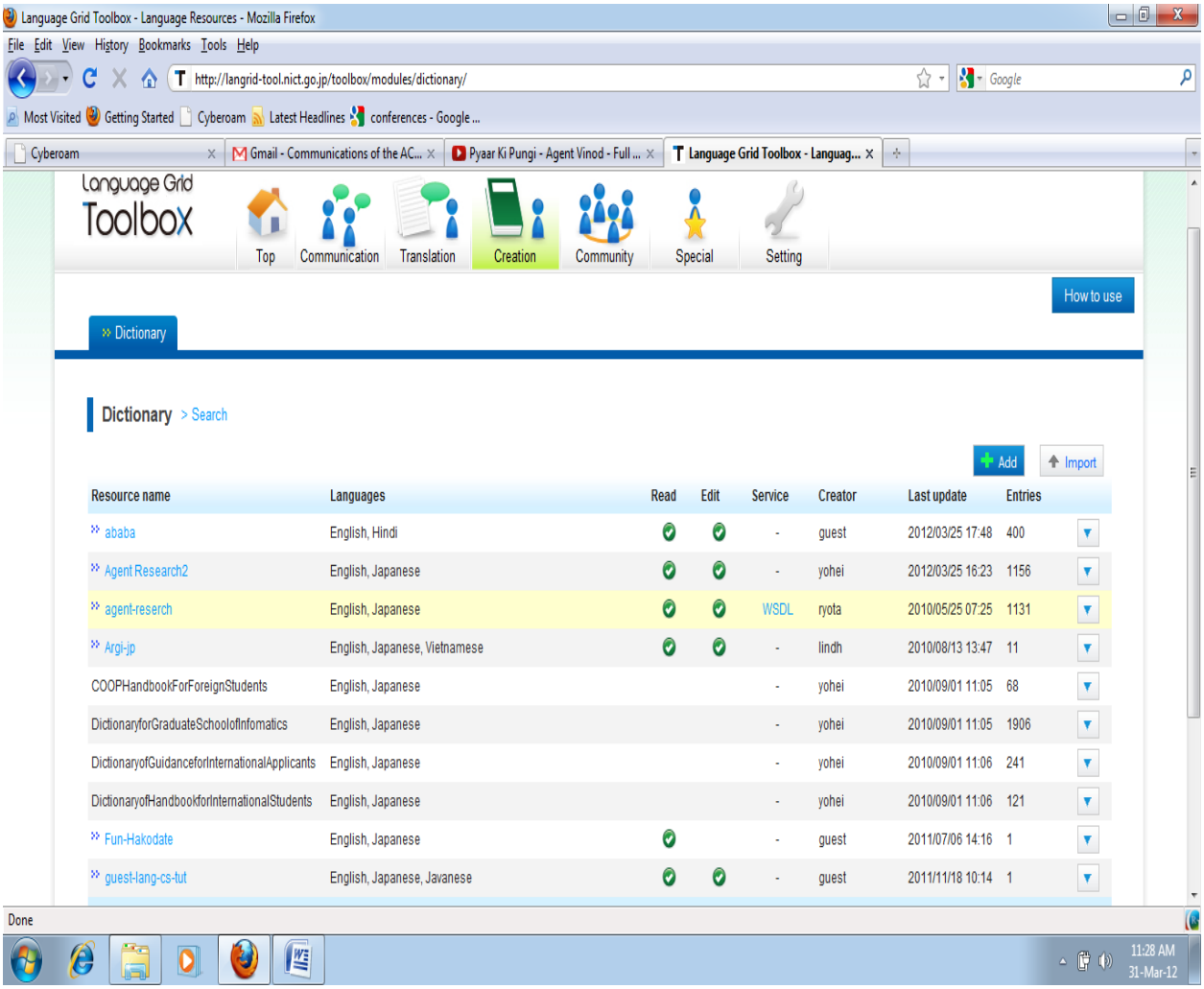

Figure 6.1: Creating a Dictionary

These words are added in the dictionary to Language Grid Toolbox website "http://langrid.org/tools/toolbox/modules/dictionary/" with the help of English to Hindi language dictionary. A dictionary from English to Hindi language of 400 words is added in the Toolbox containing words related to Language Grid and Language Grid Toolbox. Following figure 6.2 and 6.3 shows some of the words of dictionary service added to the Language Grid Toolbox site.

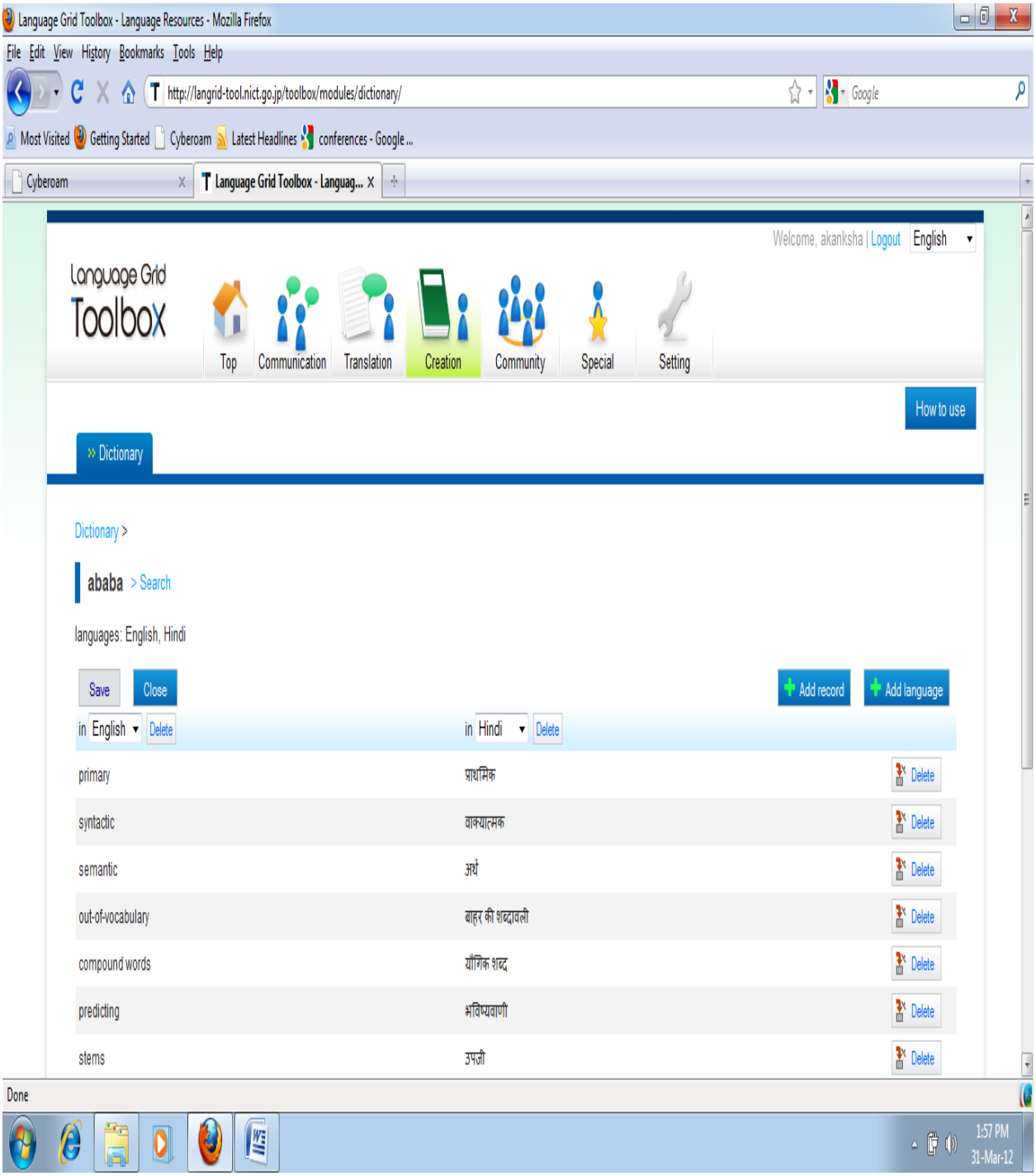

Figure 6.2: Extending Dictionary Service of Language Grid Toolbox for Hindi

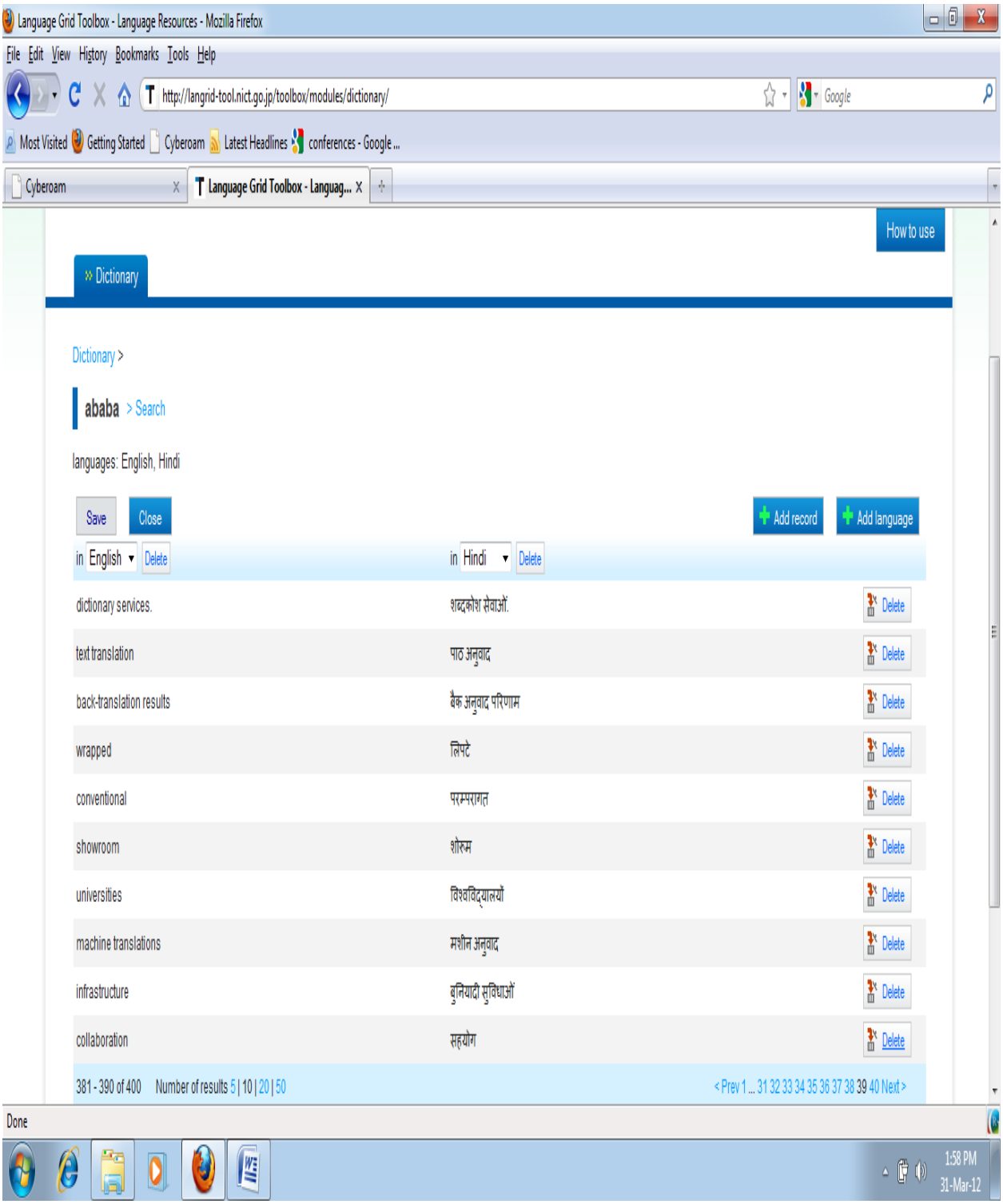

Figure 6.3: Extending Dictionary Service of Language Grid Toolbox for Hindi

Some words added in the Dictionary service of the Toolbox are:

1. Syntactic
2. Semantic
3. Customization
4. User profile
5. Task result
6. Toolbox community
7. Text translation
8. Conventional
9. Machine translation
10. Collaboration
11. Dictionary service
12. Parallel text
13. Linguistic
14. Trial site
15. Heuristic
16. Nuances
17. Thesauruses
18. Paraphrase
19. Association
20. Grammatical
21. Intellectual property rights
22. Page Dictionary
23. Bilinguals
24. Authentication
25. Morphological

Chapter 7

Enriching Q&A feature of Language Grid toolbox

7.1 Enriching Q&A Web Interface of website

The basic aim of the chapter is to explore the Language Grid Toolbox and based on this exploration demonstration as how to add Q&A Web Interface feature to the Toolbox, by enriching its Q&A Web Interface feature by adding the basic conception understanding regarding the Language Grid Toolbox for a new user. The Q&A Web Interface enables to search Q&As, to post new questions, and most important to post answers in the Toolbox. The unique quality of Q&A Web Interface function of Toolbox is that posting new questions in Toolbox do not even require the users to create their own Toolbox account but posted questions are first displayed to login users only. The questions are displayed to all users after an answer is posted by a login user.

7.2 Posting Question on Language Grid Toolbox

To post a new question, click on “Post a question” button on the top page of Q&A Web Interface as shown in figure 7.1. Language of question can be chosen from the drop down list box below the tab. A question which has no answer is not displayed if the user is not logged in the Toolbox. The questions will also be added into the Toolbox another feature Q&As.

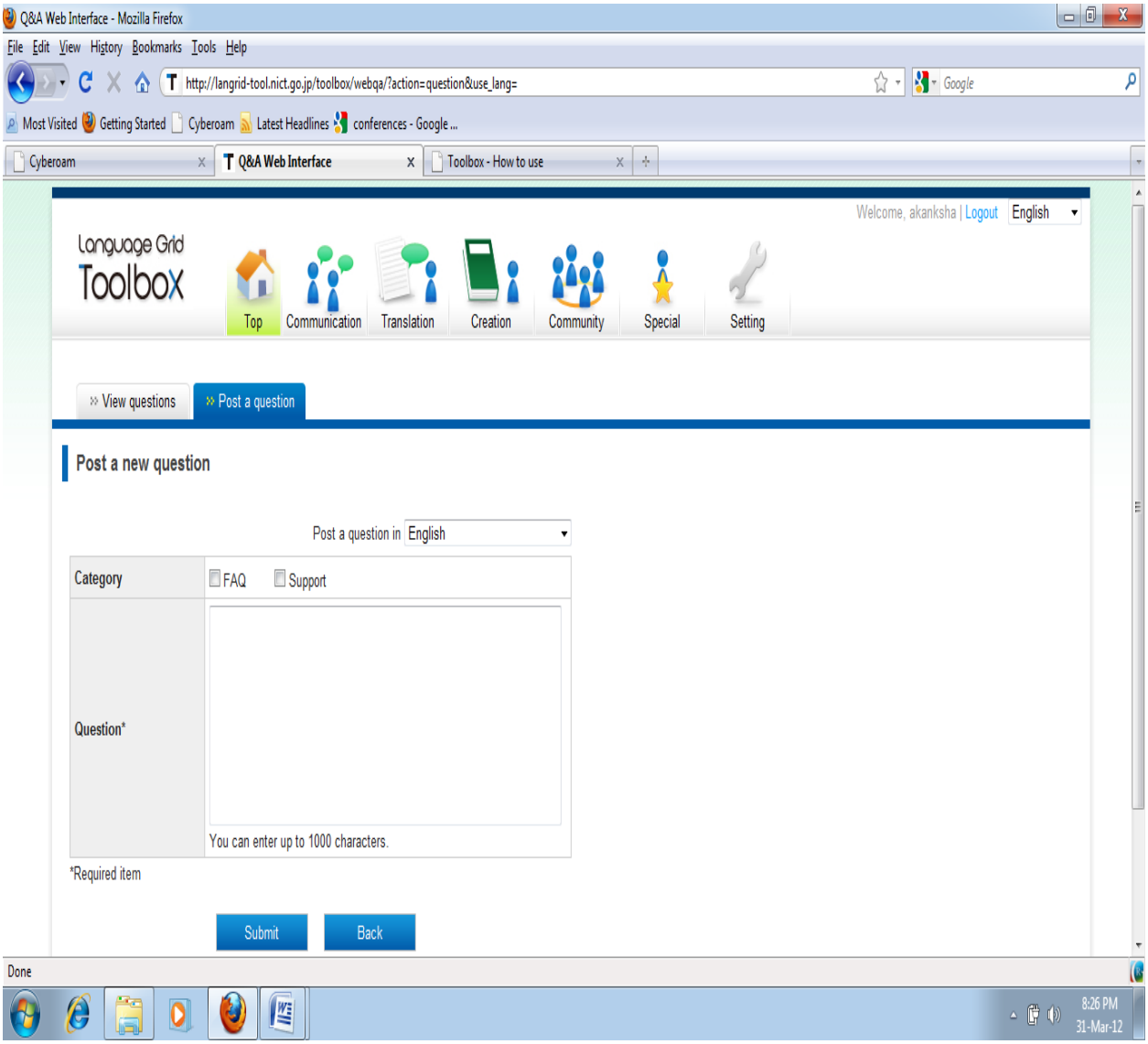

Figure 7.1: How to post question

7.3 Posting Answers on Language Grid Toolbox

A question can be viewed by clicking a snippet on the question list. If the user is logged in, an answer form is displayed to post an answer at a time. Answer can be posted by filling out the form and clicking Submit button as shown in figure 7.2. The question which corresponds to the posted answer becomes visible to all users.

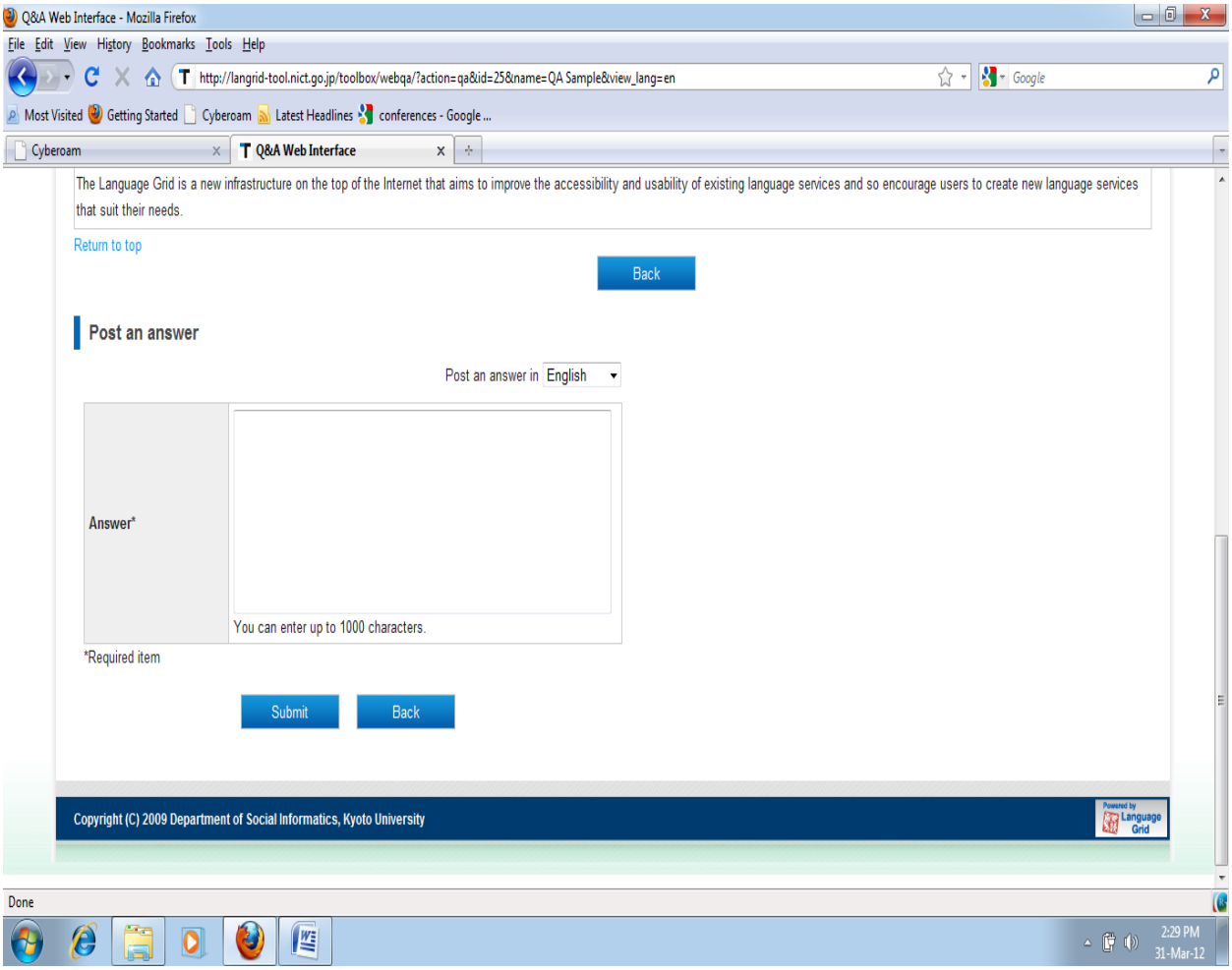

Figure 7.2: How to post an answer

7.4 Searching Q&A on Language Grid Toolbox

Q&As in Toolbox can be searched by entering the keywords in the search textbox and clicking the "Search" button. Language of Q&A can be chosen according to the user choice from drop down list box below the tab. The list of questions that include user entered keywords are shown after searching. Answers to the question can be seen by clicking the snippet of the question. As shown in Figure 7.3 keyword which is entered is "authentication", after searching resulting two questions are displayed on the screen whose answers can be seen after clicking on the question.

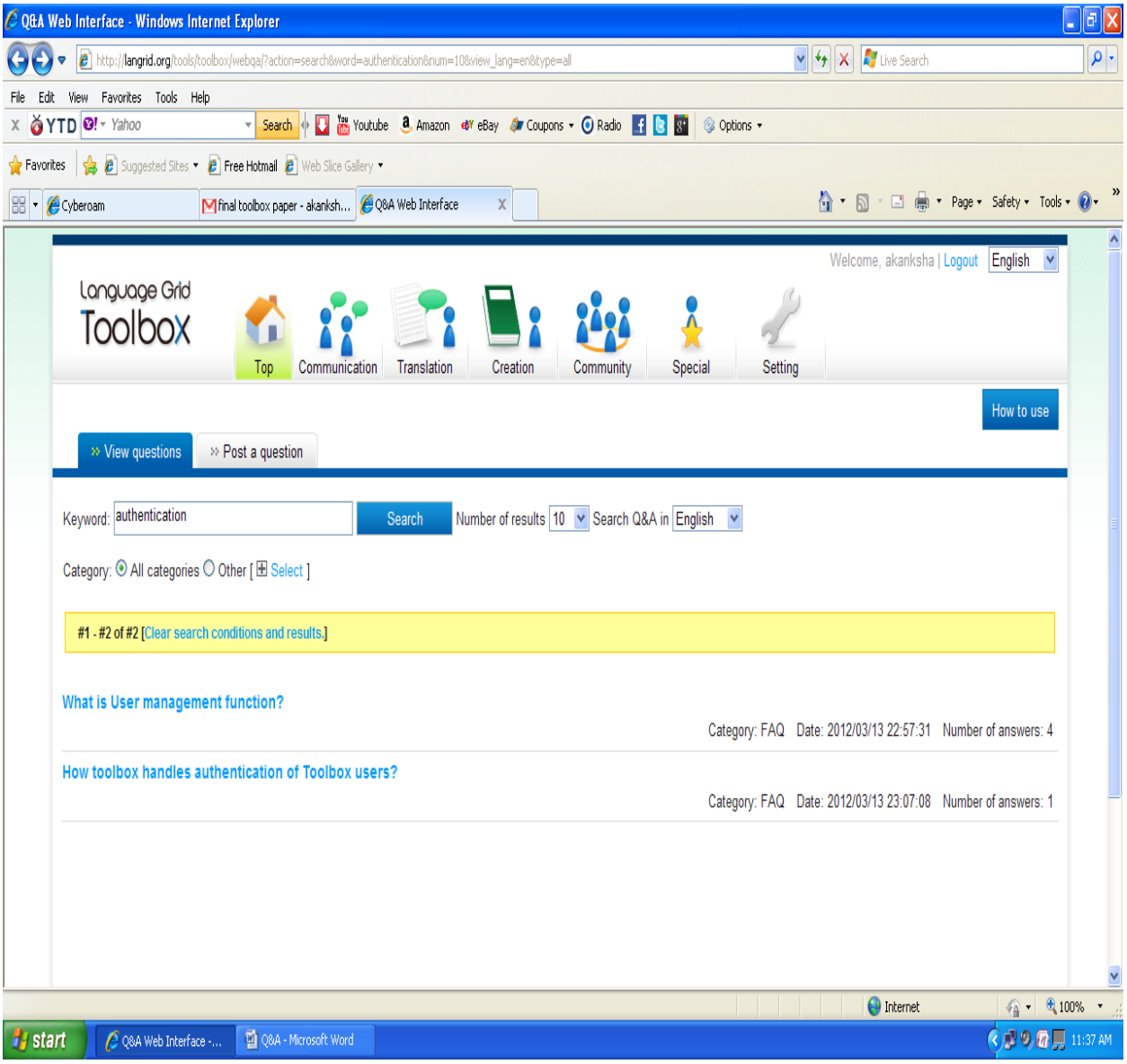

Figure 7.3: Searching Q&A

Some examples of the questions and answers which are posted in the Language Grid Toolbox site are shown below in Table 7.1. These all questions and answers are also added in toolbox another feature Q&A and reception feature of Toolbox.

Table 7.1: Questions and Answers posted in Language Grid Toolbox

I. What are the basic functions of language grid toolbox?

a) Language Grid Toolbox enables easy creation of multilingual community sites and customized multilingual environments for communities.
b) Toolbox has the basic function of creating language resources like dictionaries, which are used to combine with atomic translation services to improve the translation quality
c) Toolbox is developed as open source software and provides many basic APIs that can be used for communication, customized functions for communities can be easily developed.
d) Toolbox is a Web application which provides support tools for multilingual communication. Toolbox was developed by extending XOOPS, an open source CMS (Contents Management System).

II. What is the use of communication BBS feature in language grid toolbox?

a) BBS that enables community users to communicate with each other with their mother tongues

b) The contents of BBS are translated multilingual by language services on the Language Grid.

c) By multilingual BBS users can modify the machine translation results and improve the quality of multilingual contents.

d) Multilingual BBS is a BBS that enables users to post messages in different languages and translate them into other languages by using machine translation services and dictionary services.

III. What is the use of text translation feature in language grid toolbox?

a) Text Translation enables you to translate text and estimate the translation results by referring to the back-translation results
b) Users can extend the basic text translation function to collaborative translation function, the multilingual BBS function to multilingual discussion function, and so on.

c) 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; the user can get the translation of the entered word by using the dictionary service wrapped from the dictionary data.

d) 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.

IV. What is parallel text?

a) A parallel text is a text placed alongside its translation or translations

b) Parallel text alignment is the identification of the corresponding sentences in both halves of the parallel text. The Loeb Classical Library and the Clay Sanskrit Library are two examples of dual-language series of texts.

c) Large collections of parallel texts are called parallel corpora.

d) Alignments of parallel corpora at sentence level are prerequisite for many areas of linguistic research. During translation, sentences can be split, merged, deleted, inserted or reordered by the translator. This makes alignment a non-trivial task.

V. What are other different types of collaboration tools have been developed using the language services provided by the Language Grid?

a) Playground,

b) Newsletter

VI. How can toolbox be extended?

a) Toolbox can be extended to meet the requirements of use communities. Furthermore by using registered language services, existing communication tools can introduce multilingual function easily. like popular collaboration tool including liquid threads and nota have been successfully multilingualized.

b) Language Grid Toolbox is a series of multilingual communication tools which use the

Language Grid. Currently the following two versions of Toolbox are available.

c) XOOPS-based Community Software- This version of Toolbox aims at supporting communication in a multilingual community (This version is simply called "Toolbox"). Developed based on the open source CMS (Contents Management System) XOOPS Cube, this version is also provided as open source software. The functions can be extended by developing XOOPS modules to meet the requirements of each community.

d) Language Grid Extension for Media Wiki-Language Grid Extension for MediaWiki enables other extensions of MediaWiki to invoke translation services on the Language Grid. After the Language Grid Extension for MediaWiki is installed, two new tabs will appear in each article: "Setting" and "Page Dictionary". In the "Setting" tab, end users can select translators and dictionaries; in the "Page Dictionary" tab, end users can create a dictionary to be combined with a translator.

VII. What is User management function?

a) User Management function allows Toolbox modules to manage members of a multilingual community.

b) It also handles authentication of Toolbox users.

c) Shows the list of members of a community to let them know the state of the community and information of other members.

d) User profiles are managed by this function and can be defined for each community. Community members can search for other members based on the profiles.

VIII. What is toolbox community module?

a) Community modules use functions of basic modules through APIs, and implement community specific functions.

b) In community modules, implementation cost can be significantly reduced since the implementation mainly focuses the part of community specific functions.

c) Usage of APIs from multiple basic modules enables easy implementation of community modules that have composite functions.

d) Community modules also provide APIs so that new modules can be built even above

them.

IX. What is toolbox APIs?

a) Language Resource Management API

b) User Management API

c) Language Grid Access API

d) Text Translation API

e) Multilingual BBS API

Figure 7.4 and 7.5 shows Q&A posted in the Toolbox site as shown below.

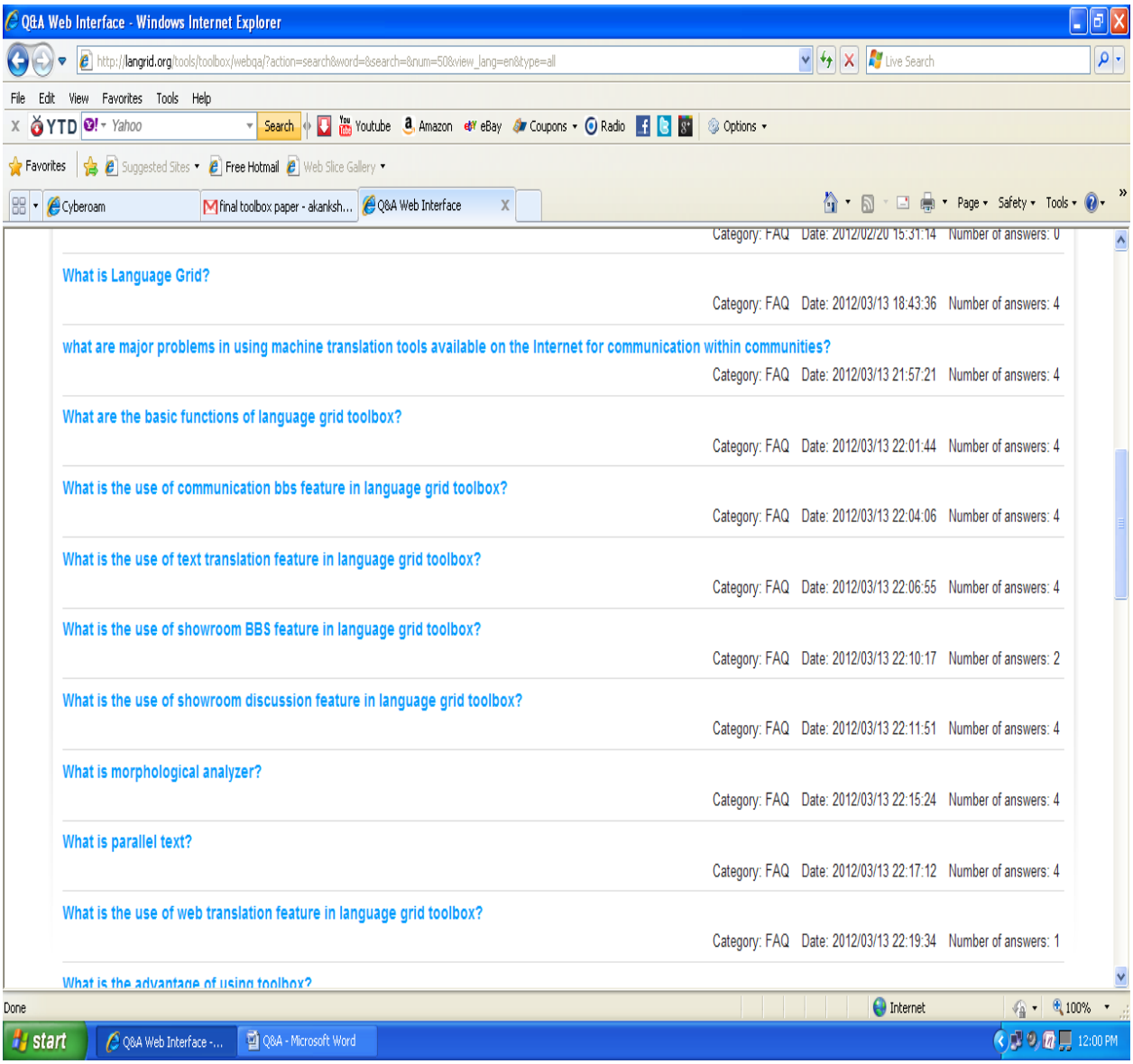

Figure 7.4: Questions posted in Q&A web interface

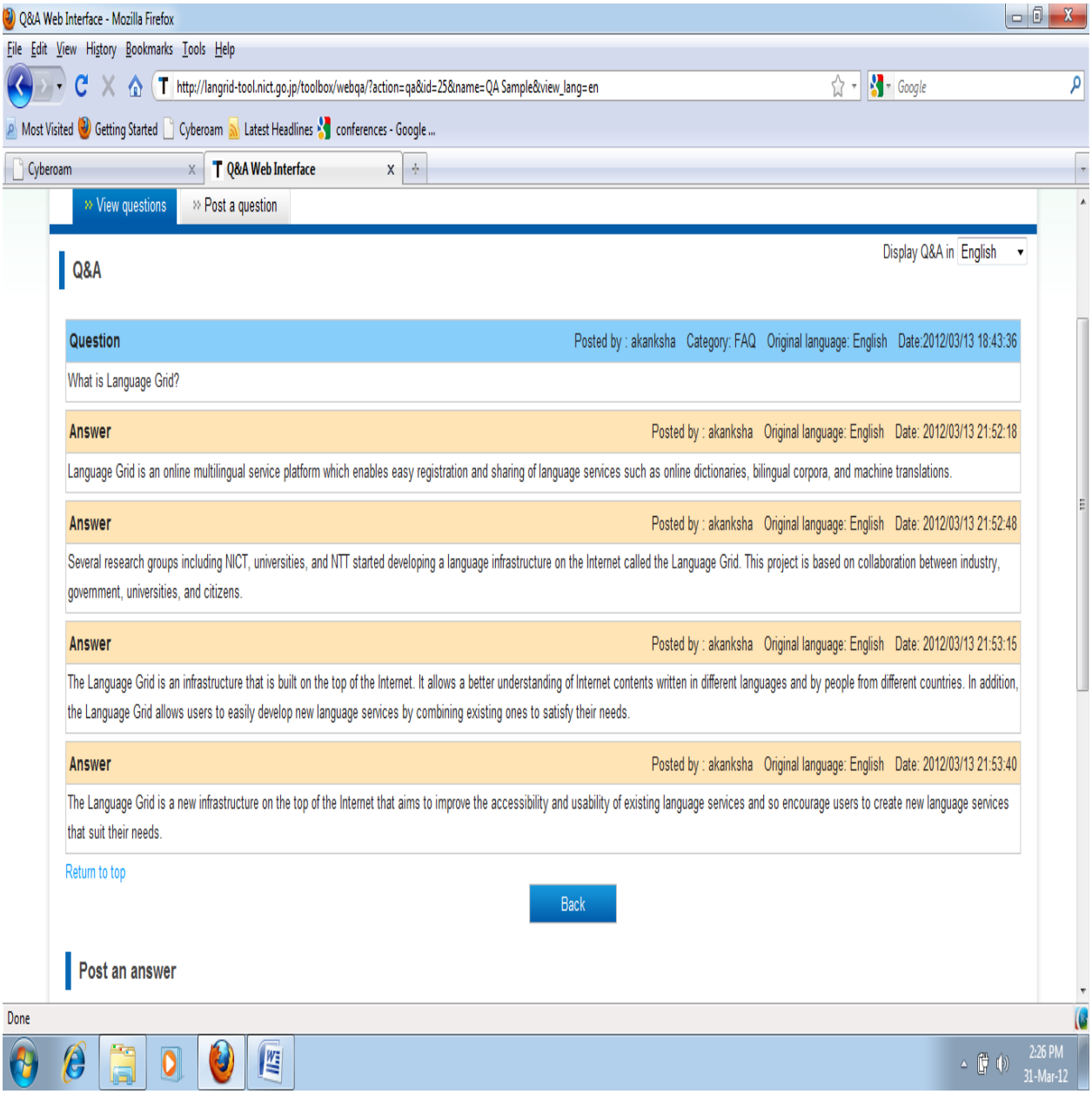

Figure 7.5: Answers posted in Q&A web interface site

7.5 Reception Feature

To start displaying a reception window, click on the "Start" button in the line of the appropriate Q&As. First, the language for a customer which is the main language and a receptionist which is a sub language is to be specified as shown in Figure 7.6. A window will open as from which appropriate category is to be selected. After selecting the category, all questions of that category will be listed.

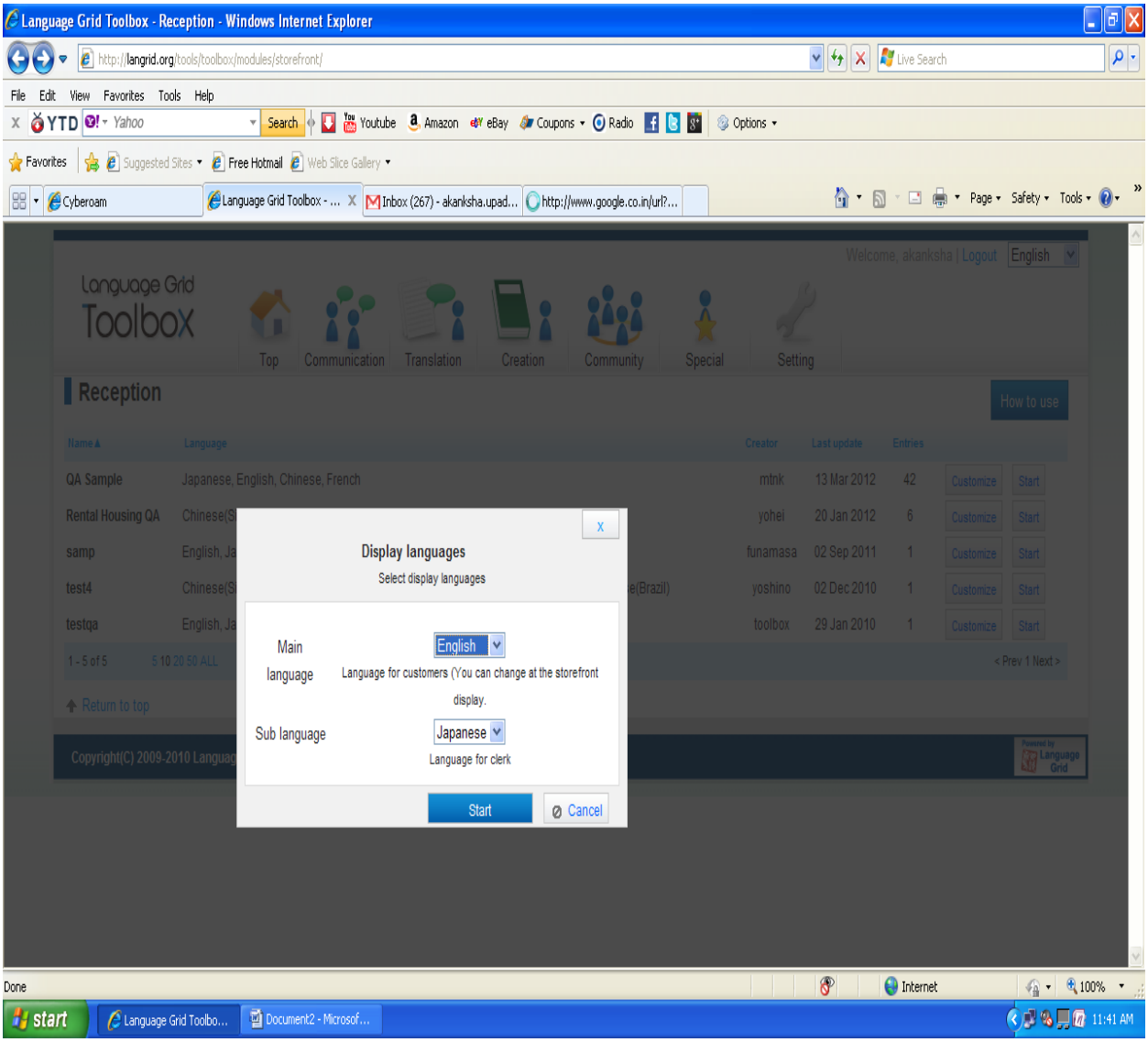

Figure 7.6: Reception window to choose main and sub language

To see the answers of a particular question, a click is required on that particular question. Answer is shown in both languages that are in main language as well as in sub language as shown in figure 7.7. When images or maps are linked to the answer, they will be displayed automatically on the right side of the window. When the answer includes terms in the linked glossary, the terms will be displayed in blue letters.

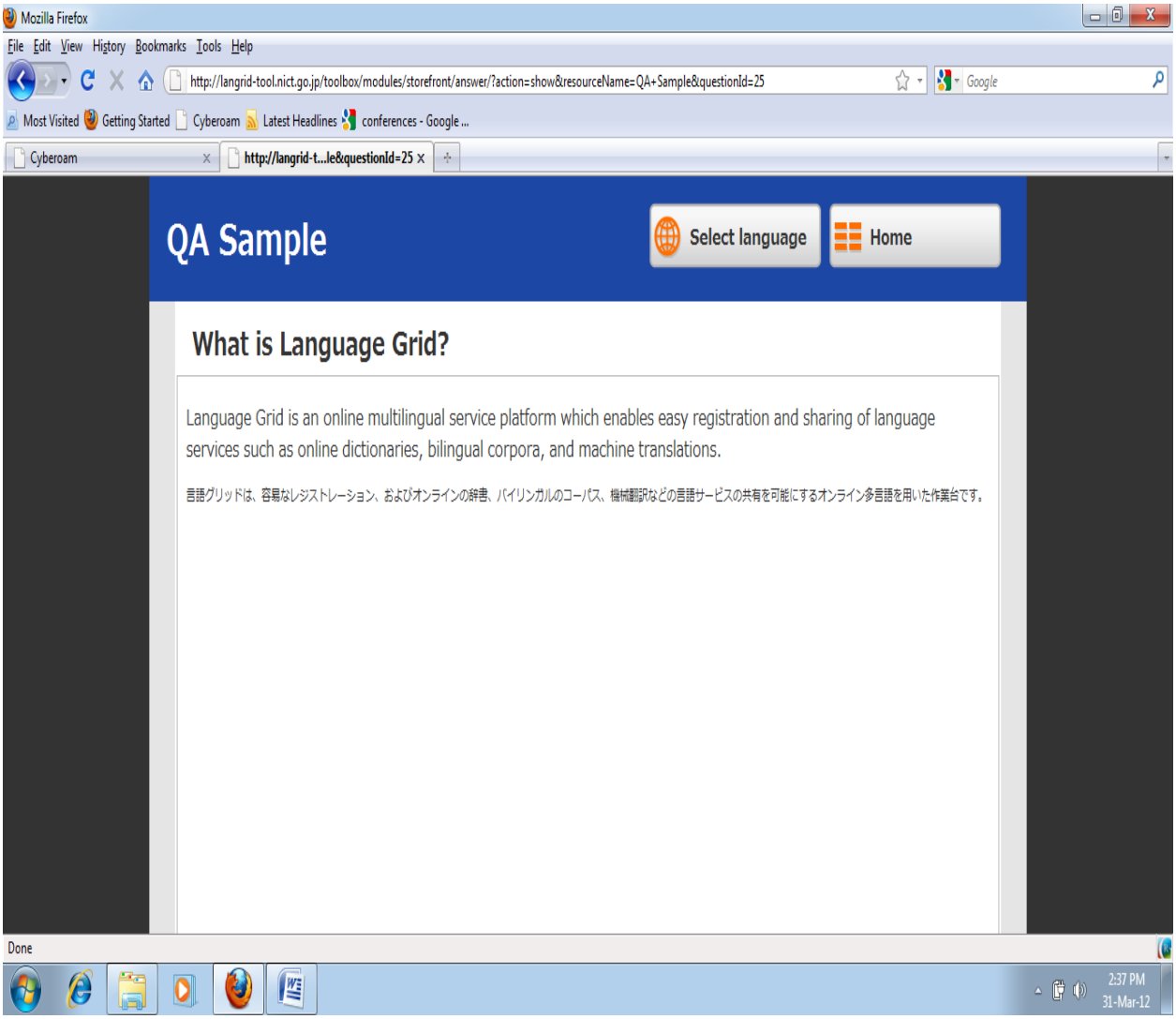

Figure 7.7: Answer in main language and sub language on Reception window

7.6 Viability of Q&A on Google

If the added questions and answers of researchers is of one of high usability in respect to Toolbox, the result displayed would show researcher Q&A in the Google search. On searching in Google any query related to questions added in Q&A Web Interface, answers which are posted in Q&A web interface comes as result of the asked query in the Google. For example, if a query “reception feature in language grid toolbox” is typed in the Google, resulted options of this query also contain the answers which were posted in the Q&A Web Interface during fulfilling the main objective of this paper as shown in Figure 7.8.

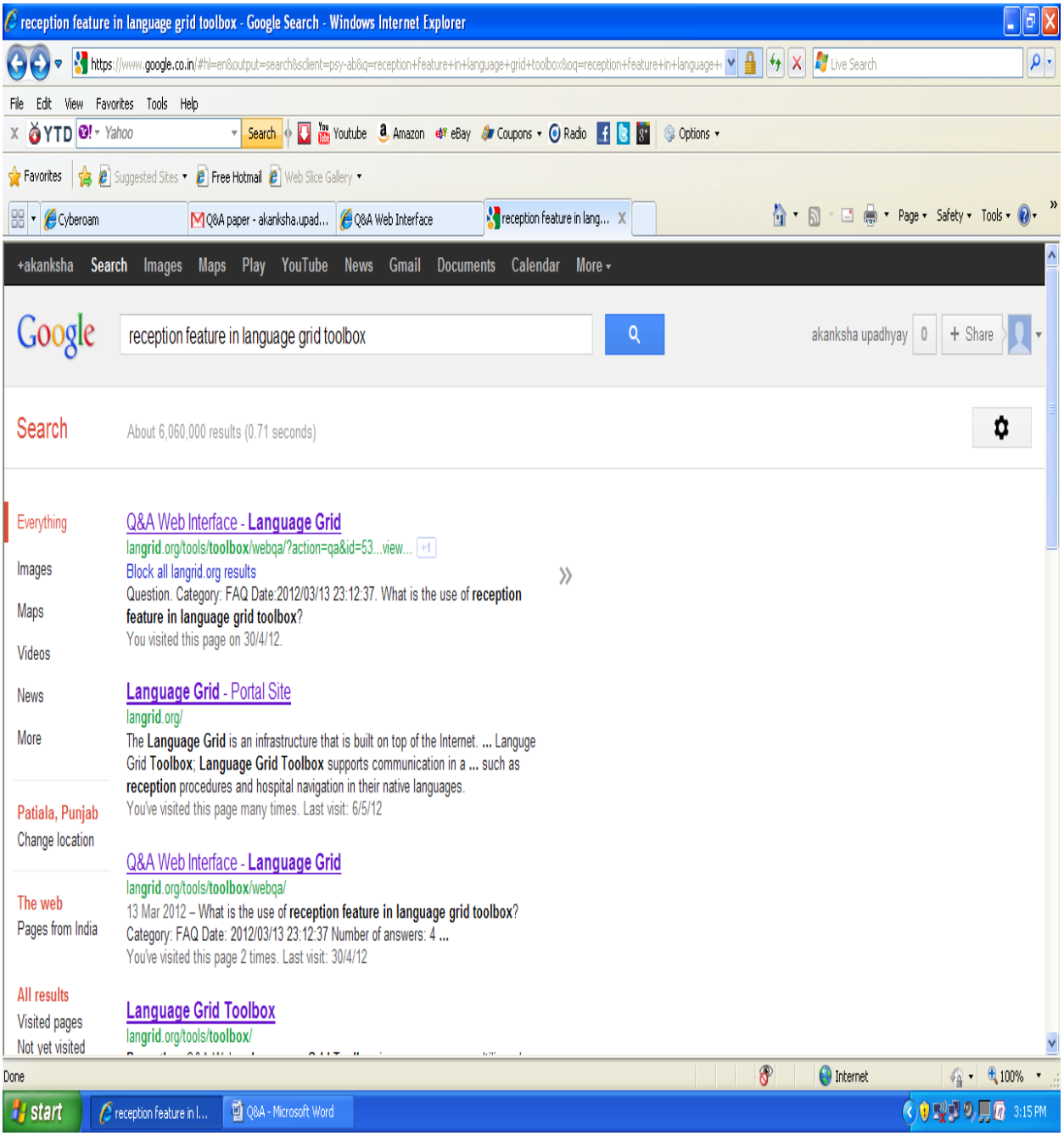

Figure 7.8: Q&A web interface related query in Google

On clicking the first option of the results retrieved by Google, same answers are listed which were added in the Q&A web interface of the Toolbox for the question related to same context as shown in Figure 7.9.

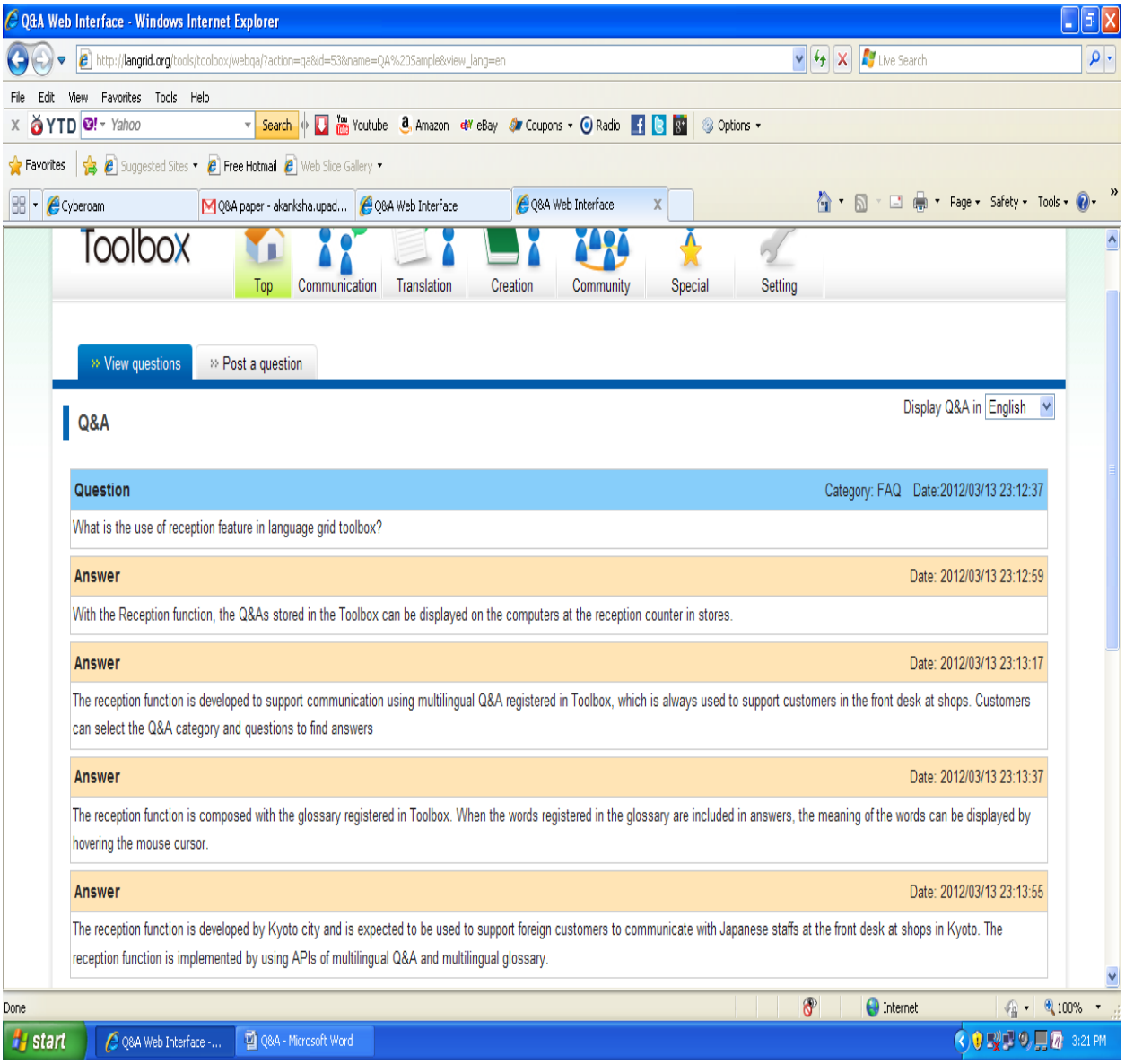

Figure 7.9: Answers related to the question posted in Google

Chapter 8

Conclusion and Future Scope

In this thesis an exploration is done on the Language Grid and Language Grid Toolbox and it is found that these platforms are very much useful in multilingual collaboration among communities.

Language Grid and Language Grid Toolbox are infrastructure that allows end users to use language services for their intercultural collaboration activities. In lieu of adding and enriching services to Language Grid Toolbox interface, an effort is made in this thesis to add a dictionary service in English language to Hindi Language, keeping in preference the Hindi language specific community.

To provide basic conception understanding regarding "Language Grid Toolbox", various questions and answers are added to the Q&A Web Interface of the Toolbox, so that users understand what is Language Grid Toolbox and how it functions and can take advantage of Toolbox to increase multilingual communication among various multilingual communities. These Q&A are implemented on the site by following the procedure available on the site. The Q&As are verified and validated by the administrator of the site and are made available to all the users who want to know and explore about the Language Grid and Language Grid Toolbox.

Future Work

A lots of work need to be done in this field as by adding parallel text, web, glossary and translation template to the creation feature of Language Grid Toolbox or by adding community specific user list, file sharing, task management, collaborative translation and discussion to the Toolbox site. In future any researcher can extend features of Language Grid Toolbox site as according to his needs or his community requirements to overcome the language barrier.

References

[1] “Language barrier,” http://en.wikipedia.org/wiki/Language_barrier,Jan 5, 2012.

[2] “Charlotte works,” http://www.charlotteworks.com/clbpositionpaper.PDF,Jan 5, 2012.

[3] T.Ishida, A. Nadamoto, Y. Murakam, R. Inaba, T. Shigenobu, S. Matsubara, H. Hattori, Y. Kubot, T. Nakaguchi, E. Tsunokaw, A Non-Profit Operation Model for the Language Grid.

[4] “The Language Grid”http://langrid.org/file/TheLanguageGrid-en.pdf, Feb 13, 2012.

[5] T.Ishida,Y. Murakami, D.Lin, The Language Grid: Service-Oriented Approach to Sharing Language Resources, 2006.

[6] “M3: Computer-Mediated Multilingual Medical Communication Support System,” http://www.langrid.org/association/m3support/indexe.html, Feb 13, 2012.

[7] T. Ishida, The Language Grid for Intercultural Collaboration, Web science Conference, NC, USA, 2010.

[8] T. Takasaki. PictNet: Semantic Infrastructure for Pictogram Communication, International WordNet Conference (GWC-06), 2006.

[9] R. Inaba, Y. Murakami, A. Nadamoto and T. Ishida, Multilingual Communication Support Using the Language Grid. International Workshop on Intercultural Collaboration (IWIC-07), pp. 118-132, 2007.

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

[11] T.Ishida “The Language Grid for Intercultural Collaboratio,” http://uptv.univ-poitiers.fr/web/data/pieces/ps0213160245.pdf, Feb 15, 2012.

[12] 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.

[13] S. Sakai, M. Gotou , M. Tanaka, R. Inaba , Y. Murakami, T. Yoshino, Y. Hayashi, Y. Kitamura, Y. Mori, T. Takasaki, Y. Nayab , A. Shigeno, S. Matsubara and T. Ishada, Language Grid Association: Action Research on supporting the Multicultural Society.

[14] "Language Grid: Service Collective Intelligence for Sharing Language Resource," http://langrid.nict.go.jp/en/project.html, Jan 5, 2012.

[15] T. Ishida, Language Grid: An Infrastructure for Intercultural Collaboration, IEEE, 2006.

[16] T.Ishida, The Language Grid: Service-Oriented Collective Intelligence for Language, Springer.

[17] N. Calzolari, A. Zampolli, and A. Lenci, "Towards a Standard for a Multilingual Lexical Entry: The EAGLES/ISLE Initiative", CICLing, pp. 264-279, 2002.

[18] Y. Murakam, D.Lin, M. Tanaka, T. Nakaguchi, T, Ishida, Language Service Management with the Language Grid.

[19] T.Ishida, "Application of Example-Based Machine Translation to Multilingual Community," Department of Social Informatics Graduate School of Informatics Kyoto University, Feb 2, 2010.

[20] "Translation in the 21 centuary"

http://www.ai.soc.i.kyoto-u.ac.jp/publications/thesis/M_H21_morimoto-satoshi.pdf.

[21] http://langrid.org/tools/toolbox/.

[22] M.Tanaka, R.Inaba, A. Nadamoto, T. Shigenobu, Intercultural Collaboration tools based on Language Grid, 2007, Volume 4568.

[23] "Service Grid Open Source Project Community Site," http://servicegrid.net/oss-project/en/toolbox.html, Feb 20, 2012.

[24] T. Ishida, M. Tanaka, Y. Murakami, D. Lin, Language Grid Toolbox: Open Source Multi-language Community Site, IEEE.

[25] "Toolbox Installation,"

http://servicegrid.net/oss-project/document/index.php/Toolbox_Installation, March 27, 2012.

[26] “XOOPS,” http://en.wikipedia.org/wiki/XOOPS, March 31, 2012.

[27] “Xoopscode-modules,” http://code.google.com/p/xoopscube-modules/, March 31, 2012.

[28] “Toolbox Shedules,” http://langrid.nict.go.jp/langrid-toolbox-wiki-en/index.php?Release%20schedule, March 31, 2012.

[29]T. Ishida, M. Tanaka, Y. Murakami, D. Lin, Language Grid Toolbox: Open Source Multi-language Community Site, IEEE.

[30]”Language Grid Association” http://www.langrid.org/association/indexe.html, Feb 20, 2012.

[31] “Language Grid,” http://langrid.org/en/index.html, Feb20, 2012

Papers Accepted/Communicated

1. Akanksha Upadhyay, Seema Bawa, Hitashi Lomash, *Enriching Q&A feature of Language Grid toolbox* at "Learning , Culture and Social Interaction".**(Communicated)**

2. Akanksha Upadhyay, Seema Bawa, Hitashi Lomash, *Extension and Translation of Terminology of Language Grid Toolbox in Hindi and English language* at "Applied Computing and Informatics". **(Communicated)**

3. Akanksha Upadhyay, Hitashi Lomash, Seema Bawa, *Language Grid For Intercultural Collaboration* at "Journal of Computer And System Science". **(Communicated)**