Accessing Dynamic Web Page in Users Language

Manoj Kumar Sharma[1], Pradipta Kumar Saha[2], Sayan Sarcar[3], Soumalya Ghosh[4] and Debasis Samanta[5]

IIT Kharagpur, Kharagpur - 721302, West Bengal

Email: [1]manojs@sit.iitkgp.ernet.in, [2]psaha@sit.iitkgp.ernet.in, [3]sayans@sit.iitkgp.ernet.in, [4]sghosh@sit.iitkgp.ernet.in, [5]dsamanta@sit.iitkgp.ernet.in

Abstract—In recent years, there is a rapid advancement in Information and Communication Technology (ICT). However, the explosive growth of ICT and its many applications in education, health, agriculture etc. are confined to a limited number of privileged people who have both language and digital literacy. At present the repositories in Internet are mainly in English, as a consequence users unfamiliar to English are not able to get benefits from Internet. Although many enterprises like Google have addressed this problem by providing translation engines but they have their own limitations. One major limitation is that translation engines fail to translate the dynamic content of the web pages which are written in English in web server database. We address the problem in this work and propose a user friendly interface mechanism through which a user can interact to any web services in Internet. We illustrate the access of Indian Railway Passenger Reservation System and interaction with Wikipedia English Website signifying the efficacy of the proposed mechanism as two case studies.

I. Keywords

Information and communication technology, Human computer interaction, Internet, Information retrieval, Ubiquitous computing.

II. INTRODUCTION

According to UNESCO report [1], about 64% of population of the globe is English illiterate. The percentage is more in developing countries like India, China, and Pakistan etc. where nearly 76% people are English illiterate. It is obvious that these people are deprived from the advantages of Internet because majority of web pages (nearly 45%) are in English ([2], [3]. Hence, there is a need to bridge the digital divide exists since beginning of IT revolution i.e. the last decade of the previous century. Let's discuss about challenges ahead to access the Internet repository. In present scenario, users who are familiar with Indian languages and less conversant with English face difficulties in accessing the web services. The traditional web services (like Indian Railways) generate dynamic web page [4] with respect to query given by the user in English language as they maintain their database in the same language. Thus, the services are lacking of support to user's query in Indian languages and as a result, unable to produce dynamic web pages in any language excluding English. It may be also noted that many translation engines ([5]–[17]) which convert web pages from English to Indian languages, have addressed the problem for static content of a web page. For dynamic web page content, the success rate is very poor. This specific challenge has been addressed in this work. We propose [18] a mechanism, called "Two-way Interaction" which enables a user to interact with the dynamic web pages in user's mother language only and the results returned during interactions are displayed in same language.

III. Proposed Methodology

In this section, we discuss our proposed approach to interact with a web service in user's language. Let user's language be L. A framework of our approach is shown in Fig.1. The framework consists of two major components: RTR and IHDD, which are discussed in the subsections below.

The RTR (Retrieve, Translation and Render) module searches the Internet for the web page with respect to request given by the user. After getting the requested page, it translates the retrieved web page into language L and renders the resulting web page in proper manner to the client machine. This module consists of three sub modules namely Retrieve, Translation and Render. The Retrieve sub module retrieves a web page as requested by the user and separates the content of web page (by using HTML parser) into HTML tags and English text. Proper indexing of links in the page is done and maintained using a table called Index Table. The Index Table handles the layout and links present in the original web page. The Translate sub module takes the extracted English text from the Retrieve sub module and converts the text to the language L. The Render sub module fully furnishes the web page in language L. Once text is converted, it merges the content in language L and set HTML tags using Index Table. In fact, the Render sub module recreates the web page in user's language having same look as the original web page.

The IHDD (Input Handler and Data Dispatcher) module is responsible for converting users input from language L to English and forward them to original web server. The module has been subdivided into two sub modules namely Input Handler and Data Dispatcher. The Input Handler first extracts the input data from web page in language L and then converts the data from that language to English. The module then invokes the Data Dispatcher which performs the authentication needed to access the web site and finally regenerates the query to be posted to original web server in English.

The working of the module is described as follows. User enters the URL of a web page what he wants to get serviced. The proposed interface then invokes the Retrieve sub module within RTR, (Step 1). After that, Retrieve module searches the specified web page in the Internet and result is returned back to the same module (Step 2 & 3). Once the web page
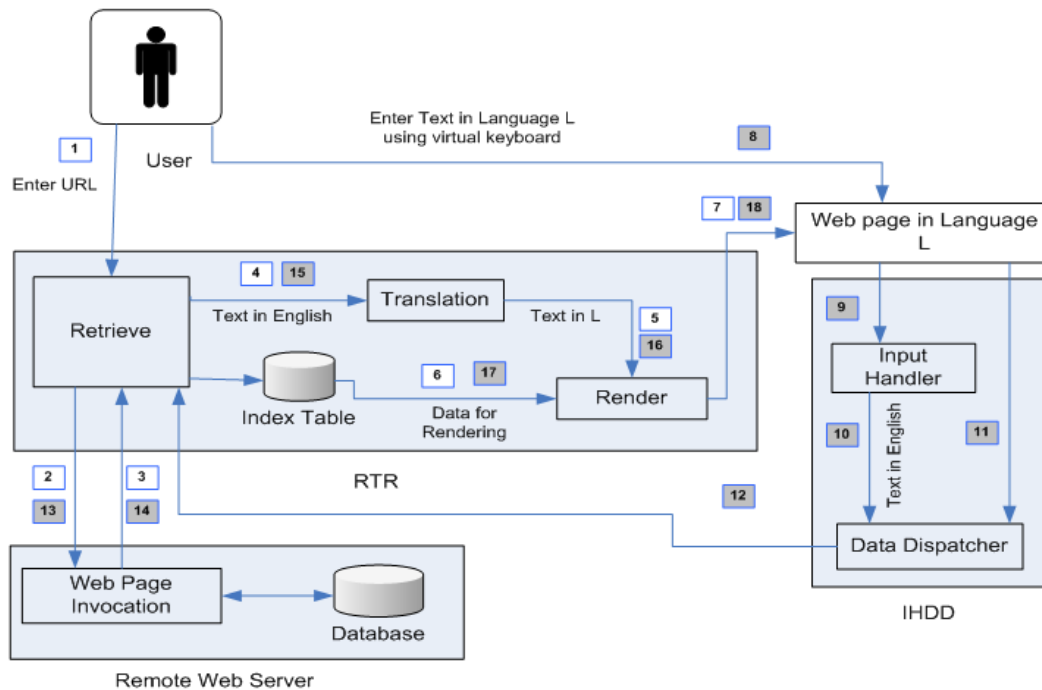
Fig. 1. Framework of Two way Interaction

is fetched, the Retrieve module separates the HTML tags and English text from the web page. The separated English content is further sent to Translation module (Step 4) for converting to language L and stores the link information in Index Table for proper maintenance of the web page (Step 6). After the completion of text conversion, Render module takes the input from Index Table and Translation module to merge the content of English text and HTML tag and generates the virtual web page in user's language (Steps 5, 6 and 7). This completes the first phase of converting the requested web page in language L.

In second phase, user gives input in his language by filling forms which appears on the virtual web page in L with the help of virtual keyboard [19] (Step 8). Input Handler then extracts the text entered by user (Step 9) and calls Translation module to convert the text from language L to English. After that, this generated English text and the virtual web pages are sent to Data Dispatcher module (Steps 10 and 11) which regenerates the query in English, handles all the authentication needed to post the web page in Internet and then invokes the Retrieve module (Step 12). Now the task of Retrieve module is to process the request and invoke the remote server which accesses the database (database contents are in English). The result returned from the remote server has finally been stored within the Retrieve module (Step 13 and 14). The Retrieve module then sends the result to the Translation (Step 15) and Render (Step 16 and 17) modules which generate web page in language L. This gives illusion to user that the result which is displayed on the virtual web page fetched exclusively in user's language instead of English.

IV. CASE STUDY

In this section, we illustrate our proposed mechanism [18] with two case studies.

A. Indian Railway Passenger Reservation System

We consider one of the popular web services in India, "Indian Railway Passenger Reservation System (IR-PRS)" [20]. Our proposed mechanism provides an interface between the user and IR-PRS in Internet maintained by Indian Railways [20] (CRIS). The interface gives a flavor to the user that the entire IR-PRS web page is written in his language. It also provides facility to the user for giving input and getting output in his language in his language. In this study, we explain the accessing the IR-PRS in Hindi language (spoken majority of people in India). It is not necessarily limited to Hindi. In fact it can be applied to any language, of course with their corresponding translation scheme in needed. The original service provided by Indian railway will be beneficial for user familiar to English language. The corresponding web page of Indian railway written in English language is shown in Fig: 2(a). The converted web page in Hindi language is shown is Fig: 2(b).

When user gives source and destination name of stations [21] say "Kharagpur" and "Delhi" to the original web page of IR-PRS the results returned by IR-PRS in English is shown in Fig: 2(c). With our proposed Interface, this original web page is first translated to Hindi using machine translation systems and displayed the same but in Hindi. The users are now allowed to enter the text directly in Hindi Unicode [22] format using virtual keyboard for Hindi. After entering the

(a) Original content of IR-PRS

(b) Entering input in Hindi to virtual web page of IR-PRS

(c) Result returned by IRP-RS

(d) Displaying the result in Hindi

Fig. 2.   Accessing Dynamic Web Page Of Indian Railway in Hindi Language

source (खड़गपुर) and destination (दिल्ली), the user submits the form to Hindi web page for processing as shown in Fig: 2(b). The result returned by IR-PRS is retrieved, translated, rendered and finally displayed in Hindi. The web page shows the detail of the train name with number in Hindi language. One can also check the availability and train time table in the same language (see Fig: 2(d)).

B. Wikipedia English website

Searching for the word (लकड़ी) in English Wikipedia "http://en.wikipedia.org/wiki/" using Google website translation service, will return only those page written in Unicode.

As the word (लकड़ी) doesn't exist in English Wikipedia it will not return any result but when the word "wood" is given it will display the english pages containing the word "wood".
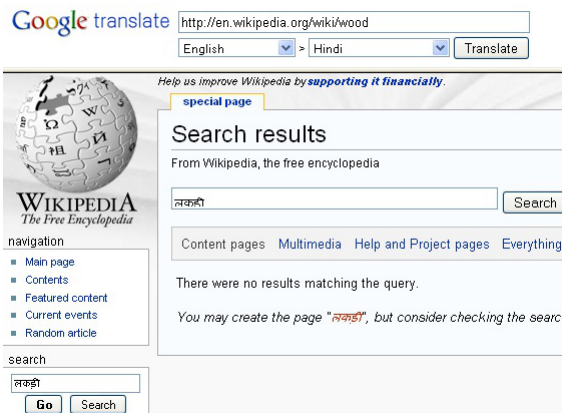

Fig. 3.   Result when user gives input in Hindi under Google translation for word (लकड़ी)

Fig. 4.   Result when user gives input under Google translation for word (wood)
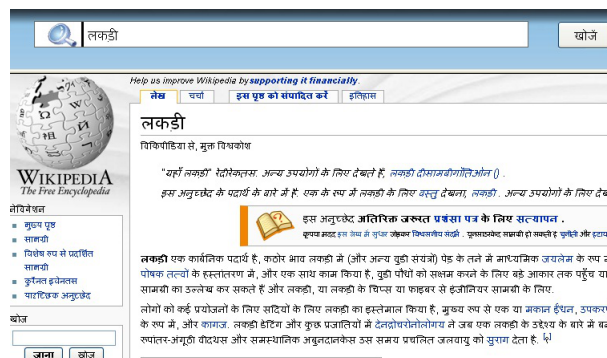

Fig. 5.   Result when user gives input in Hindi in our service for word (लकड़ी)

V. Discussion

In this work, we have addressed the problem facing by English illiterate people to access Internet where majority of

the pages are stored in English. Our work is a contemporary and first of its kind. The work also addresses many limitations, which Google is yet to address. Google has web page translation scheme called Google Translate [23]. This translation scheme is partial and more importantly one way. As an example, to translate IR-PRS web site using Google's translation, user must give the station name in English. If user gives the input in English, site searches the train between the given stations and resulting web page is rendered to user in English only (as Google looses the control to translate [24]). When user gives the source and destination name in Hindi Unicode like source name (खड़गपुर) and destination name (दिल्ली), Google gives unauthorized invocation error and unable to translate the page, neither in English nor in Hindi.

We have tested our mechanism with more than 35 popular web sites with two way interactions and results are error free and satisfactory. We have tested our approach with Hindi and Bengali and it is applicable to any language provided that accurate translation scheme is available in that language.

Accessing Dynamic Web Page in Users Language plays a great role, as it will minimize the gap between the Internet and user due to language barrier ([18]). It helps the users to access Internet in their day to day life without worrying much about the language web page is originally written. This service provides the Internet content to millions of people who might not have good capability to read the web content in English.

REFERENCES

[1] "Unesco, international literacy statistics a review of concepts, methodology and current data," http://unesdoc.unesco.org/images/0016/001 628/162808e.pdf (Last accessed on March, 2010).

[2] "India world's second largest english speaking country," http://tesol-india.ac.in/EnglishTeachingIndustry/india-worlds-second-largest-english-speaking-country (Last accessed on April, 2010).

[3] "Languages and cultures on the internet study 2007," http://dtil.unilat.org/LI/2007/ro/resultados_ro.htm (Last accessed on July, 2010).

[4] "Wikipedia, dynamic web page," http://en.wikipedia.org/wiki/ Dynamic_web_page (Last accessed on April, 2010).

[5] "Google translate," http://translate.google.com (Last accessed on August, 2010).

[6] "Translate a block of text," http://in.babelfish.yahoo.com (Last accessed on August, 2010).

[7] "World lingo, free website translator," http://www.worldlingo.com/en/websites/url_translator.html (Last accessed on August, 2010).

[8] "Free text translation," http://www.freetranslation.com (Last accessed on August, 2010).

[9] "Free automatic translators, machine translations comparison tests," http://www.humanitas-international.org/newstran/more-trans.htm (Last accessed on August, 2010).

[10] "Language translation, translate phrase or word," http://www.translation.langenberg.com (Last accessed on August, 2010).

[11] "Websites translator, translate your website from english," http://www.websitestranslator.com (Last accessed on August, 2010).

[12] "Taming the beast, web page language translation," http://www.tamingthebeast.net/articles6/page-language-translation.htm (Last accessed on August, 2010).

[13] "Ability, website translation and localization," http://www.localization-translation.com/translation-localization-services/web-sites-localization.html (Last accessed on August, 2010).

[14] "Babylon, english to hindi translation," http://translation.babylon.com/English/to-Hindi (Last accessed on August, 2010).

[15] R. Sinha, R. Jain, and A. Jain, "Translation from english to indian languages: Anglabharti approach," in *Proceedings of Symposium on Translation Support Systems STRANS2001, February,15-17, Kanpur, India,* 2001.

[16] "An english to hindi machine-aided translation system," http://mt-archive.info/MTS-2003-Sinha.pdf (Last accessed on August, 2010).

[17] "Wikipedia, comparison of machine translation applications," http://en.wikipedia.org/wiki/Comparison_of_machine_translation_ applications (Last accessed on June, 2010).

[18] "Bharati project," http://www.nid.iitkgp.ernet.in/Bharati/ (Last accessed on August, 2010).

[19] "Wikipedia, virtual keyboard," http://en.wikipedia.org/wiki/Virtual_ keyboard (Last accessed on May, 2010).

[20] "Indian railways passenger reservation enquiry," http://www.indianrail.gov.in (Last accessed on September, 2010).

[21] "Wikipedia, list of railway stations in india," http://en.wikipedia.org/wiki/List_of_railway_stations_in_India (Last accessed on August, 2010).

[22] "Unicode consortium," http://unicode.org (Last accessed on March, 2010).

[23] "Google translate, frequently ask question," http://www.google.com/intl/en/help/faq_translation.html (Last accessed on July, 2010).

[24] "Google translate, indian railway page translated by google," http://translate.google.com/translate?hl=en&sl=en&tl=hi&u =http://www.indianrail.gov.in/inet_srcdest_names.html (Last accessed on August, 2010).