Interacting with Dynamic Web Portals in Local Language

Ch.JyosthnaDevi ,Santosh Kumar shah ,J.Devi Durga, K.Ruth Ramya,N.Vahini,Gyan pandy

Abstract- Today and Tomorrow, the most advanced technology is Information Technology (IT). But, the applications of IT are being limited to only those who are familiar in English. It is not useful for those who are unknown in digital literacy. In an increasingly interconnected world, the interactions among devices, systems, and people are growing rapidly. At present the repositories in Internet are mainly in English, as a consequence users unfamiliar in 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 Andhra Pradesh State Road Transport Corporation System and interaction with Wikipedia English Website signifying the efficacy of the proposed mechanism as two case studies.

Index Terms - Human computer interaction, Information technology, Internet, Information retrieval, Ubiquitous computing, Virtual Keyboard.

1. INTRODUCTION

India is a highly multilingual country with eighteen constitutionally recognized languages and several hundred dialects & other living languages. Even though, English is understood by less than 3% of Indian population, it continues to be the de-facto link language for administration, education and business. Hindi, which is official language of the country, is used by more than 800 million people. 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 users languages and less familiar with English face difficulties in accessing the web services. The traditional web services (like Wikipedia search engine and many other) 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 through remote web server.

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 observe that many translation engines like google translate([5]–[11]) 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 [13] a mechanism, called "Twohanded Interaction" which enables a user to interact with the dynamic web pages in user's required language only and the results returned during interactions are displayed in same language.

2. Standard architecture for web- based applications

This system is designed based on a traditional three-tier architecture used by many web applications. Three-tier architecture includes a presentation layer, business rules/ logic layer, and the data layer. The three-tier architecture is shown in Figure 1.

The three-tier architecture is generally used when an effective distributed client/server design is needed that provides

- increased performance
- ✤ flexibility
- ✤ maintainability
- * reusability and

[•] Santosh kumar shah is currently pursuing Bachelors degree program in computer science and Engineeringin KLCE, India, E-mail:shah.santoshcse@gmail.com

J.Devi Durga lakshmi is currently pursing Bachelors degree program in computer science and Engineering KLCE,India, Email: devi.jangala@gmail.com

[•] Ch.Jyosthna Devi, Asst Prof, Dept Of CSE, K L University

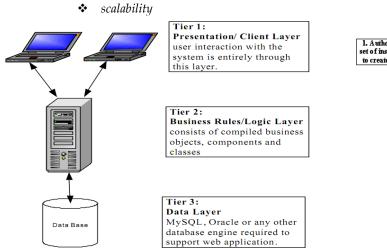


Figure:1 Standard architecture for web- based applicationS

3. Implementation Technologies

By typing a URL (Uniform Resource Locator) into the address box of the browser the communication between a browser and a web server is started. Each conversation consists of two pieces:

Dynamic web page

In a dynamic web page content (text, images, fields, etc.) on the web page can change, in response to different contexts or conditions. There are two ways to create this kind of web pages:

1. Using client-side scripting to change interface behaviors within a specific web page

2. Using server-side scripting to change the sequence of the web pages or web content

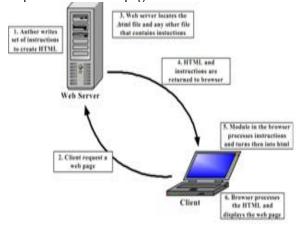


Figure:2 Client-Side Dynamic web page

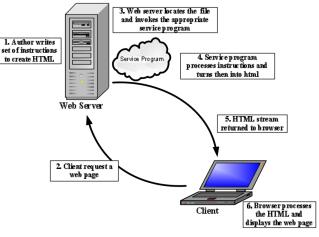


Figure:3 Server-Side Dynamic web page

4. Google Translation

Google Translate is a free, web-based and statistically-based machine translation service provided by Google. It enables to translate section of text, document or webpage, from one language to another.

4.1 Google Translate API

The Google Translate API lets websites and programs integrate with Google Translate programmatically. During the project development phase Google provides two versions of API. The version 2 is the latest available version of the Google Translate API. We decided to use version 2 within the project. After that point whenever we refer to translate API in this report, we mention version 2. One needs a Google account to use the translation service. Because, the Translate API requires the use of an API key and it can only be received from Google APIs console. There are two ways to invoke the API: Using REST directly or using REST from JavaScript (This does not require server-side coding.). JSON is used by Google as a data format.

4.2 Limitations

When one uses the Google API, it is required to accept the terms and conditions. Here are the most important limits for our project:

1. Every request text to be translated can be up to maximum of 5000 characters long.

2. Daily limit is 100.000 characters per API key.

3.Continuous translation requests successively results in Suspected Terms of Service Abuse .

4. Batch requests are against Terms of Service.

5. Tricia

Tricia Platform is an Open Source platform for developing dynamic web applications. It is built in Java and realizes the model-view-controller pattern, which consists of the following components:

- 1. An abstraction for defining control flow.
- 2. A templating language
- 3. An object/relational persistence mapping

It has to be emphasized that what has been described so far is called Tricia Platform, since it is a generic platform for building arbitrary dynamic web applications. Tricia Platform is implemented in the Eclipse project Toro. In addition to the pure infrastructure, the toro project already comes with the basic asset types Person and Group, which are required by almost any dynamic web application. Additionally, Toro comes with an existing basic layout. On the other hand, there are concrete plugins built on top of the Tricia platform and the basic Tricia plugin. Combining these plugins results in the Tricia Application.

A typical configuration of Tricia consists of the plugins Toro, File, Wiki, and Blog. The distinction between the platform and the application is similar to Ruby on Rails as a platform and applications like Basecamp or Backpack, which are built using Ruby on Rails.

5.1 Existing System

The present internationalization system in Tricia is not fully tested and does not perform its full intended functionality. It is a partially implemented system. The internalization features present in Tricia are:

1. Language Handler:

The Language Handler handles the change of language parameter as triggered by the user action. The Language Handler identifies the user chosen language from the url parameter and sets in to the session variable. The session variable is set when the user clicks his/her preferred language on the Tricia page. The languages are identified with the flags as icons.

2. Translator Configurator:

The Language flags on the Tricia pages can be configured through the Translator configurator tool of Tricia Platform. This tool provides away to add the desired languages (according to the multiple languages required to be supported by the Tricia). A language can be added through the configurator tool.

- The Internationalization plugin for Tricia has to
- 1. Identify the texts appropriate for translation.
- 2. Store and handle the transalted text.

3. Update the Tricia Web page according to the language chosen by the user.

4. Regularly update the translations.

5.2 Tricia Technical Design:

After we discovered what to be translated, we designed our translation system. It is depicted in Figure 6. We try to design the system component based. As we mentioned before, static texts are not stored in database in the Tricia. However, we decided to save translated texts into database because of three reasons: Performance issues, Google Translate API limitations and memory consumption. If we think translation as a kind of computation, we can say that we use optimization technique so-called memoiza-tion18. We translate the texts once and store them. Whenever we need a translation of a text, we look up from database table. On the other hand, this method reduces API calls and helps

us not to exceed API limits. The last but not least is memory consumption. The static texts were stored in the files and when Tricia started to run, the texts were copied into memory in order to benefit of speed of memory operations. This was acceptable for one language but what happens when Tricia supports many languages? This increases memory consumption factor of supported number of languages. Scalability of Tricia was undermined due to the internalization. Storing the static texts into database enables us to overcome all these problems.

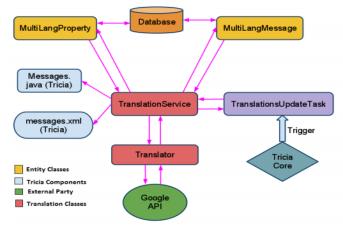


Figure: 4 overview of Translation system

6. System Overview

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.5. 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 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 (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.

International Journal of Scientific & Engineering Research, Volume 2, Issue 12, December-2011 5 ISSN 2229-5518

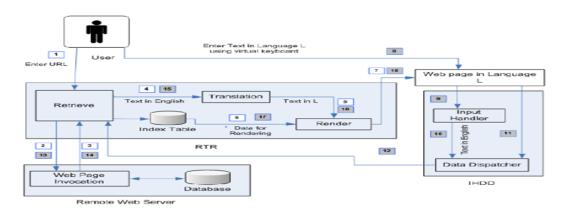


Figure:5 Framework of two-way interaction

7. CASE STUDY

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

7.1 Andhra Pradesh State Road Transport Corporation System

We consider one of the popular web services in India, "Andhra Pradesh State Road Transport Corporation System (APSRTC)". Our proposed mechanism provides an interface between the user and APSRTC in Internet maintained by Andhra Transport .The interface gives a flavor to the user that the entire APSRTC 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 APSRTC" in Hindi language (National language of 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.



Fig:1 Accessing Dynamic web page of Andhra transport in English language



Fig:2 Accessing Dynamic web page of Andhra transport in Hindi language

7.2. Wikipedia English website

Searching for the word in English Wikipedia "http://en.wikipedia.org/wiki/" using Google website translation service, will return only those page written in Unicode. As the word doesn't exist in English Wikipedia it will not return any result but when the word "Agriculture" is given it will display the English pages containing the word "Agriculture".

कृषि Search	
Recommended Sites	
🖬 Exators 🧐 Enclares	
tan 📷 Annar	
Maplepages •	
wzarti Czywa	
Fig.2 Willing die Cooreb Fra	
Fig:3 Wikipedia Search Eng	gine
N	6 Logini de
Aricle Discussion Reat Ed	View history Search
EDIA Registration for WikiConference India 2011, Numbai is now open. Click here to register.	
un consect un	913
Agriculture	
ent From likipeda, he tree enclospeda s	
Agriculture (also called farming or husbandry) is the cutivation of animals, plants, fungi and other life forms for food, fiber, and other products used to susta	
testa Agriculture was the key implement in the rise of sedentary human chilization, whereby farming of domesticated species created bood surpluses that nutred th development of civilization. The study of agriculture is known as agricultural science. Agriculture is also observed in certain species of ant and terming. ²⁰⁰⁹ but	
speaking refers to furnan activities.	generally
The history of apriculture dates back thousands of years, and its development has been driven and defined by greatly different climates, cultures, and technolo	
eda	
	loped Agriculture
Portal However, all tamming generally releas on techniques to expand and mantam the lands suitable for masing domesticated species. For plants, this usually require form of ingation, although there are methods of dryland farming; restoral herding on rangeland is still the most common means of raising herstock. In the deve	
total former, all torming senselly release in technologies to expand and martine the lastic suitable for using domesticated spores: the parts, this usually requere form dringstone, although there are methods of dripland financy particul hereing or complexitor is shift the meter common mease drinking limits (being the work), rubustical approximation and in page-scale moticulture backcome the dominant spaties of motion filmings.	inable General
Portal However, all tamming generally releas on techniques to expand and mantam the lands suitable for masing domesticated species. For plants, this usually require form of ingation, although there are methods of dryland farming; restoral herding on rangeland is still the most common means of raising herstock. In the deve	inable General Aptusiness - Aptuduralisti
ont were, at himsing prevaily easi is interting and is equited an intertain the basis search for any dimensional point in the subary any dimensional point. In the subary point is the subary dimensional point is the time contract, in the deviation of the subary dimensional point is the subary increased yields from clusters, but is the subary increased yields from clusters.	inable General Aptualness Approximations Approxy - Animal husband Extension Extension
port lower, at himsgrees, yes in lockspace to expect an extract to sale usate for any heread point, the usail years ben displaced by the en ended of displacements grantering and usate and the sale of the sale of the sale of the lower displaced by the ended of the sale of the sale applicate by any sale sales grantering sales and the sales are sales and the sales are sales and the sales are sales applicate by any sales and the sales are sales are sales and the sales are sales are sales are sales and the sales are sales ar	inable General Agrossiness - Agricultural scie Agrocomy - Animal hostend Betwee Extensive faming ig farming Factory faming - Farm - Free n
port ones, at himsgreesity eris in inclusion is expect an increase the sease scale for any dimensional point. In source, increase than in displace, that do by the are mothed of displacements providers in all the recent means of assignments. In the de- ory of which is grandwater to they scale recordure to scale the dimension providers in all the recent is any scale in the same special in grandwater or spece special providers in the scale increase the dimension providers in all the recent providers in the same state of the scale increases and increases and the same state of the same increases of the theory of the scale increases of the theory of the scale increases of the theory of the scale increases of the scale increases and any scale increases of the trackics, specification of the scale increases and and increases providers in the scale increases and and increases and the increases, and the increases of the increases of the increases, and the increases of the increases o	inable General Agrossiness - Agricultural scie Agrocomy - Animal hostend Betwee Extensive faming ig farming Factory faming - Farm - Free n
port lower, all himsing generally real is indicing and begins that match the basis search for any dimensional points. The stand region is the fund in right stand by the are indiced of high thimsing stand begins and the most carbon manue of assis places. It has each indicated in the stand stand by the stand begins and the stand begins and the most carbon begins and the stand stand and stand by the stand by the stand in the stand by the stand by the stand by the st	intable General Aptosities - Aptochristick Approxy - Annali tuskies ig farming r biotestic spriciture to biotestic spriciture biotestic family - Personality
prot lower, al himtogreeckij wie is lochtigen to logist al nettrant to kass usaek trangs direkticatel point i rydgen, the usail greech ben direktij wie is lochtigen ein erhold of usative mission start level and angelos is all the rock minima and assiss places. It he de und rudke usative greech wie is lochtigen ein beneden eine bener strend minima and assiss places. It he de und rudke usative greech einer beneden einer beneden einer beneden einer beneden einer beneden einer beneden var de und rudke usative greech einer bene beneden einer beneden einer beneden einer beneden einer beneden var de und rudke usative greech einer bene beneden einer beneden einer beneden einer beneden var de nachte direktige al derage and regelos bene beneden einer beneden einer beneden und einer beneden und einer beneden her einer beneden greech einer beneden einer beneden her einer beneden greech einer beneden einer beneden einer beneden einer beneden einer beneden einer beneden her einer beneden greech einer beneden ein beneden einer beneden einer beneden einer beneden	Intable General Aptoximes - Aptoximition Aptoximes - Aptoximition Aptoximes - Aptoximition Betwee Entry Entry Entry Interview Entry Interview Entry Organic Remoy - Persecutive Seatematic systems
port ones, al himograesa jesis in inclusione is expect and increation the bias search for angl detectable point. In fusion, priving, this usually prevent the search in the de- term of increase. The development of the search increases and the search increases and the search in the de- origination of increases and the search increases and the developed and produces in the sharph consequences in the search increases and the developed and produces and the brady paper into brady, paper into brady and paper into brady paper into brady paper into brady, paper into brady, paper into brady, paper into brady and paper into brady, paper into brady, paper into brady, paper i	intelle General Apthobases - Apthobases - Apthobases Aptromy - Acmail hotand pig lammag Factory Samay - Fann - Freen Intensis family Organic temps - Armania Stationale approache
prot lower, al himtogreeckij wie is lochtigen to logist al nettrant to kass usaek trangs direkticatel point i rydgen, the usail greech ben direktij wie is lochtigen ein erhold of usative mission start level and angelos is all the rock minima and assiss places. It he de und rudke usative greech wie is lochtigen ein beneden eine bener strend minima and assiss places. It he de und rudke usative greech einer beneden einer beneden einer beneden einer beneden einer beneden einer beneden var de und rudke usative greech einer bene beneden einer beneden einer beneden einer beneden einer beneden var de und rudke usative greech einer bene beneden einer beneden einer beneden einer beneden var de nachte direktige al derage and regelos bene beneden einer beneden einer beneden und einer beneden und einer beneden her einer beneden greech einer beneden einer beneden her einer beneden greech einer beneden einer beneden einer beneden einer beneden einer beneden einer beneden her einer beneden greech einer beneden ein beneden einer beneden einer beneden einer beneden	Inside Control Applications - Applications - Applications of the Name Applications - Applications of the Application - Applications of Applications - Applications - Applic
prote prote there at theme generally rear is inclusionable to equate at memory the status the term and density the status of terms that and them if regional terms in the special executive terms status integration at all the executive terms and a dang terms. In the den- sity of terms and a dang terms that the special executive terms is the status memory at the special general terms and any channel or general status of terms and terms is the status of terms in terms and terms in terms and dang terms and any channel or general status of terms and terms and terms is the status of terms and terms in terms and terms in terms any channel or general status of terms and terms and terms in the status of terms and terms in the status of terms and terms in the channel endpoint and the status of terms and terms and terms in the status of the terms and terms in the status of terms and terms in the terms and terms and terms and terms and terms and terms in the status of the terms and terms in the status of terms and terms and terms in the status of terms and terms a	Intelle General Agtisuitees - Apolations of Agtisuitees - Apolations of Compose - Annie Nateden Compose Inneg - Fairer Free or Internet Journey - Fairer Free or Internet Journey - Fairer Inneg Oppose Inneg - Fairer Statement apolation Statement apolation Statement apolation Adda Baboy History of Spottane Anda Apolaterial Foodation
point lower, at himsing penary level is explored in explored and transmit the basis sector for any dimensional point. In the outpression of the penary dimensional point is penary dimensional point. In the openary and the sector be done and the outpression of the rescue the done any dimensional point. In the openary and the sector be done and the outpression of the rescue the done and the penary dimensional point of the sector be done and the outpression of the rescue the done and the penary done and the sector be done and the done and the done a	Institle Constitutions of Applications - Applicatio

Fig:4 Result when the user gives input In English Word



Fig:5 Result when user gives input in Hindi under Google translation for word कृषि.

4. 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 . This translation scheme is partial and more importantly one way. As an example, to translate APSRTC web site using Google's translation, user must give the station name in English. If user gives the input in English, site searches the BUS between the given stations and resulting web page is rendered to user in English only (as Google looses the control to translate). When user gives the source and destination name in Hindi Unicode like source name and destination name, 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.

Interacting Dynamic Web Portal in local Language plays a great role, as it will minimize the gap between the Internet and user due to language barrier . 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, methodologyand current data," <u>http://unesdoc.unesco.org/images/0016/001</u> 628/162808e.pdf.

[2] "India world's second largest english speaking country," <u>http://tesolindia</u>.

International Journal of Scientific & Engineering Research, Volume 2, Issue 12, December-2011 7 ISSN 2229-5518

c.in/EnglishTeachingIndustry/india-worlds-

second-largestenglish- speaking-country

[3] "Languages and cultures on the internet study 2007," http://dtil.unilat.org/LI/2007/ro/resultados ro.html.

[4] "Wikipedia, dynamic web page," http://en.wikipedia.org/wiki/ Dynamic web page[5]"Googletranslate,"http://translate.google.com

[6] "Translate a block of text," http://in.babelfish.yahoo.com

[7] "World lingo, free website translator," http://www.worldlingo.com/en/websites/url translator.html

[8]"Freetexttranslation,"http://www.freetranslation .com

[9] "Free automatic translators, machine translations comparison tests," <u>http://www.humanitas-</u>

international.org/newstran/more-trans.htm

[10] "Language translation, translate phrase or word," http://www.translation.langenberg.com

[**11**] "Websites translator, translate your website from english," http://www.websitestranslator.com 2010).

[**12**] "Taming the beast, web page language translation,"http://www.tamingthebeast.net/article s6/page-language-translation.html

[13] "Ability, website translation and localization," http://www.localization-

translation.com/translation-ocalizationservices/ web-sites-localization.html

[14] "Accessing Dynamic web pages in user Language"http://ieeexplore.ieee.org/xpl/freeabs_all .jsp?arnumber=5783859