International Journal of Smart Business and Technology
Vol.3, No.2, (2015), pp. 1-16
http://dx.doi.org/10.21742/ijsbt.2015.3.2.01

XLIFF: Multilingual Translation Memory Management among Divergent Language Families

Guo Jiangzhen[1], Priyanka Pawar[2], Pratik Ardhapurkar[3], Priyanka Jain[4], Anuradha Lele[5], Ajai Kumar[6] and Hemant Darbari[7]

[1]*Department of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, China*
[2,3,4,5,6,7]*Centre for Development of Advanced Computing, Pune, India*
[1]*gjzltx001@163.com,* [2]*priyankap@cdac.in,* [3]*apratik@cdac.in,* [4]*priyankaj@cdac.in,*
[5]*lele@cdac.in,* [6]*ajai@cdac.in,* [7]*darbari@cdac.in*

Abstract

For localization, the only textual output is not sufficient the need of Machine Translation unless it is in a usable format. In the Indian scenario, localization as an industry has not been recognized yet which had led to a lack of Language Standards leading to varied translation quality. Localization is the process of adapting a product or service to a particular language, culture, and desired local "look-and-feel" Machine Translation is one of the most important activities under localization but it is not complete unless it is adapted by end user in the desired manner. In this paper, we are introducing format retention utility using "XLIFF" as an important tool to English to Indian Language Machine Translation. Machine Translation (MT) is the technique of translating source text of input language into the target language text. This process uses bilingual data set along with other language assets to frame language and phrase models which are used while translating text in machine translation. Here, Machine Translation System (MTS) that uses the Tree Adjoining Grammar (TAG) is considered. Here, the uniqueness and complexity of the task have been discussed. In this paper, we are proposing a design and architecture to support the system along with experiments, results and future aspects. It is closely related to the long-term vision of enabling code to support local, regional, language, or culturally related preferences.

Keywords: *Natural language processing, Machine translation, Indian language, Localization, Standardization, Format extraction and format retention, Rich text format, HTML, XLS and global XLIFF, Machine translation, Natural language processing, Tree adjoining grammar, Translation memory*

1. Introduction

Localization (shortened to "l10n") is the process of taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold. This process becomes effective as document production has been organized and automated with the use of XMLized form. For better adaptability by end-user, localization companies use a huge variety of file type formats Word, Xls, PDF, and many more. These file formats are capable to maintain formats like - font face, font color, font size, indentation, alignments, justification, header and footer to name a few. Along with

Article history:
Received (August 6, 2015), Review Result (October 12, 2015), Accepted (November 14, 2015)

Print ISSN: 2288-8969, eISSN: 2207-516X IJSBT

this. maintaining this format after machine translation of the document is the purpose behind our research.

The motivation behind our work is to examine and implement the essential factor of a complete localization process of Machine Translation into different language families. Looking into the Indian environment, we are discussing the English to Indian Language [E-IL] Machine Translation System. To overcome the language barriers, it utilizes the potentiality of a computing system for Machine Translation [MT]. Indian languages belong majorly to Aryan and Dravidian languages which are flexible free order in Subject-Object-Verb [SOV] syntactic order whereas the English Language has syntactic order Subject-Verb-Object [SVO].

The universal bid for translation is increasing perceptibly in the knowledgeable age without enough gain in the count of translation professionals. The research and development in Machine Translation extend to develop at a rapid rate. This has been achieved as the computational exercises have become dominant along with the internet which stimulates the multilingual and global community widely. This has accelerated productivity and the distribution rate of translations. Machine Translation System (MTS) is based on the Tree Adjoining Grammar (TAG) [19]. Translation Memory (TM) is a popular technology among translation practitioners since it has come to the market in the 1990s [27].

Being belongs to the different grammatical family, retention of file format after translation from source to target is a challenging task as format information of source token in input text is relocated into translated text. Here, XLIFF [8] offers a standard format to accomplish these exchanges. It provides a mechanism to generate localized files from XML format as required and panoramically endorse the collaborator specific data formats. XML is meant in such a way that it is easily understood by humans, and at the same time, it is machine readable [28]. It is a tool-independent and standardized file format that can be used for MTS.

This paper describes the conversion of documents from the source file to XLIFF and back to the original format. This paper gives a technical overview of format retention using XLIFF, proposed system design, its different modules and interaction among these. Section-II gives a study on related work in the area. Section-III describes file formats, difficulty in its retention in E-IL System and XLIFF as a solution. Section-IV proposes a working model and its architecture design to support format extraction and rebuilding in Machine Translation. Section V presents experiments and results on it. The last section concludes the paper with a discussion on the importance, limitation and future work for localization.

2. Literature review

In today’s business world, customers from all over the world regardless of the size of the company or organization stipulate a system to have support for multiple languages. Many organizations and development teams work in globalization and multilingual support, which enables products and applications to be used in many different languages and countries. It manages standards on different aspects of information management and technology topics. Machine Translation in the desired format plays an important role to support this environment.

The Pan American Health Organization (reHo) aid the exercises of Machine Translation since 1980 [20]. By processing the required text to be translated by the systems at any time into smaller chunks a better output quality and performance can be achieved [21]. To have machine translation, a rule based approach was taken. The translation rules used by the system are motivated by linguistic considerations. For this, a tree generating system called tree adjoining grammar (TAG) was studied which was introduced by Joshi et al. (1975) and Joshi (1985) [22]. In the TAG algorithm, insertion of the absolute and unbounded quantity of stuff in the existing tree structure is granted by using adjunction operation [23].

In India, translations from English to Indian Languages is a crucial need that promotes the development of Machine Translation System [MTS]. In this paper, MTS is implemented using TAG formalism such that it parses source sentences to generate a target structure that ultimately can get the target sentence. A TAG is a combination of elementary trees that is further bifurcated into initial and auxiliary trees. These trees undergo adjunction and substitution which build derived trees. Based on Part Of Speech (POS) category sequence of a source sentence, derivation structure is produced [24]. The outcome of this approach structure alters for Indian languages. The structure of Indian languages conflict grammatically with English language grammar, it becomes critical to rebuilt a file in the target language.

The translation system can also be forced to store this translated output in the form of translation memory so that it can be reused for the same translations performed latterly. This is a renowned technique but it can be achieved via various ways to full-fledge it. This technique used in MTS is known as Translation Memory (TM). Translation Memory (TM) technology, has originated in the 1970s and headed on in the 1980s. The development has swung into a significant monetary entity since the 1990s [26]. The Translation Memory schemes as a backbone to the multilingual content processing industry.

The need for file formats has been subsisted to have long term retention including documentation, wide adoption, transparency, self containment and usage within the archival community. The Internationalization Tag Set5 (ITS) [11] a localization standard released by W3C for localization engineers to support the internationalization and localization of data and documents. The RWS Group came up with a need for a file format for machine translation systems where localized data is supposed to get translated [13]. At the end of 2000, several companies based in Ireland of localization groups had decided to try and find some common solution to their problems. XLIFF was established by professionals based in Dublin, Ireland as an informal group of localization and globalization. Today SAP [7], SDL [8] and Microsoft[4] have also joined their efforts to get a hierarchical format, recursive representation of interchange file data and a format that could handle variant translations and machine translation. As updated by [15], the Data Definition group was thus brought into existence and they started to work on this issue.

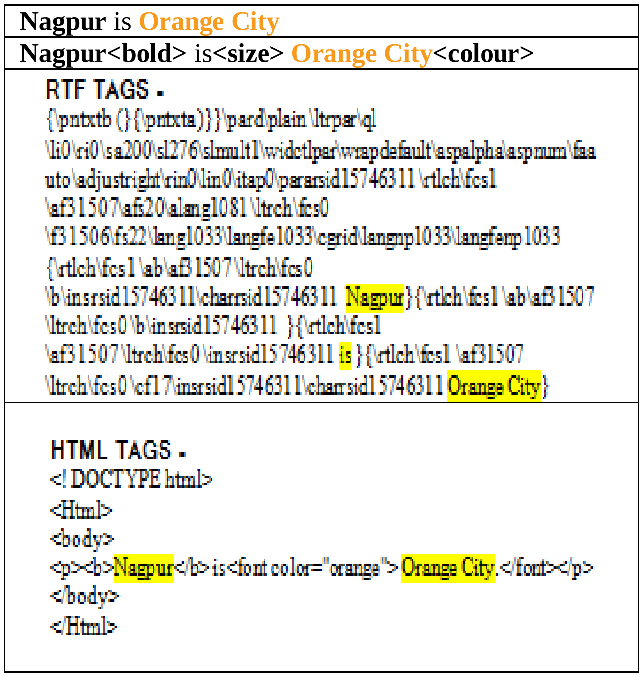
The result of their efforts is XLIFF [9], the XML Localization Interchange File Format. It is a framework that combines many localization standards in the multilingual information framework which carries a large amount of metadata. Oracle Corporation [6] uses XLIFF on Business Intelligence Publisher which provides translations which are for a reporting solution. Similarly, SharePoint Server [3] uses XLIFF to transport information about a file and its contents from SharePoint Server to a human translator. ITS2XLIFF(v 0.6) conversion tool was developed to generate up to date XLIFF files (v 1.2) from XML files for which W3C ITS rules [16]. A survey[1] has been presented on the use of XLIFF in the Localization Industry and Academia. Today XLIFF version 1.0 is used in production which was published in 2002 beginning. It was being used to provide more loss-less data interchange between tools and simplify its extensibility.

The XLIFF version 1.2 [18] (approved as a Specification in February 2008) specifies the XML Localization Interchange File Format (XLIFF) where localizable data is stored and carried from one step of the localization process to the other while allowing interoperability between tools. XLIFF represents a format that can is efficiently handled and modified by translators in MT. A well-formed document in XML such as an XLIFF file can be displayed easily by most web browsers [29]. The XLIFF standards are extended in their structure at several points which aims just to provide interoperability between different tools [18].

3. File format and XLIFF

Nowadays, Major organizations essentially have data in their predefined and standard format. The only textual translation is not sufficing the need of MT as it is incomplete unless it is in a desired usable format. A task is accomplished if it fulfills these pre-requisites because the rapid technology change has made us available with many file formats. For facilitating all the phases of the localization process, the file format serves as a container for data. This externalized data needs to be interchanged between localization tools and software services providers in order. Here, we are discussing few file formats which are widely used for general documentation. XLIFF can rebuild most of the documents. But document formats in translated layout require fixes to some stage. XML, RTF and HTML are the formats that usually require very few adjustments after translation. Fig. 1 shows RTF and HTML file formats with different tags used for basic formatting.

Rich Text Format [17] generally known as RTF is a proprietary document file format developed by Microsoft Corporation[4] and supported by many word processors. It can be considered a universal word-processing format as it can be used for cross-platform document interchange. However, unlike plain text, it retains basic formatting information, like font sizes and styles and also includes many other formattings like bold, italic, underline, colors, etc. It is beneficial in terms of usability, comprehension, beautification, daily purpose official document. The HTML[2] tags are not visible on the browser, but it structures the website to interpret the content of the page in a required format.



Nagpur is **Orange City**

Nagpur<bold> is**<size>** **Orange City<colour>**

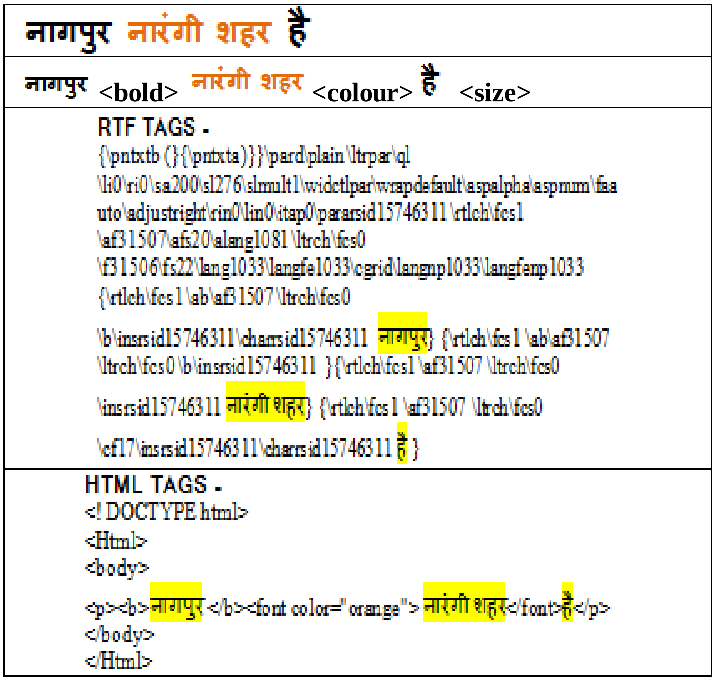
RTF TAGS -

```
{\pntxtb (}{\pntxta)}}\pard\plain \ltrpar\ql
\li0\ri0\sa200\sl276\slmult1\widctlpar\wrapdefault\aspalpha\aspnum\faa
uto\adjustright\rin0\lin0\itap0\pararsid15746311 \rtlch\fcs1
\af31507\afs20\alang1081 \ltrch\fcs0
\f31506\fs22\lang1033\langfe1033\cgrid\langnp1033\langfenp1033
{\rtlch\fcs1 \ab\af31507 \ltrch\fcs0
\b\insrsid15746311\charrsid15746311 Nagpur}{\rtlch\fcs1 \ab\af31507
\ltrch\fcs0 \b\insrsid15746311 }{\rtlch\fcs1
\af31507 \ltrch\fcs0 \insrsid15746311 is }{\rtlch\fcs1 \af31507
\ltrch\fcs0 \cf17\insrsid15746311\charrsid15746311 Orange City}
```

HTML TAGS -

```
<! DOCTYPE html>
<Html>
<body>
<p><b>Nagpur</b> is<font color="orange"> Orange City.</font></p>
</body>
</Html>
```

Figure 1. Formatting tags in RTF and HTML files



नागपुर नारंगी शहर है

नागपुर <bold> नारंगी शहर <colour> है <size>

RTF TAGS -

```
{\pntxtb (}{\pntxta)}}\pard\plain \ltrpar\ql
\li0\ri0\sa200\sl276\slmult1\widctlpar\wrapdefault\aspalpha\aspnum\faa
uto\adjustright\rin0\lin0\itap0\pararsid15746311 \rtlch\fcs1
\af31507\afs20\alang1081 \ltrch\fcs0
\f31506\fs22\lang1033\langfe1033\cgrid\langnp1033\langfenp1033
{\rtlch\fcs1 \ab\af31507 \ltrch\fcs0
\b\insrsid15746311\charrsid15746311 नागपुर} {\rtlch\fcs1 \ab\af31507
\ltrch\fcs0 \b\insrsid15746311 }{\rtlch\fcs1 \af31507 \ltrch\fcs0
\insrsid15746311 नारंगी शहर} {\rtlch\fcs1 \af31507 \ltrch\fcs0
\cf17\insrsid15746311\charrsid15746311 है }
```

HTML TAGS -

```
<! DOCTYPE html>
<Html>
<body>
<p><b>नागपुर </b><font color="orange"> नारंगी शहर</font>है</p>
</body>
</Html>
```

Figure 2. Rebuilt formatted tags with translated output

Here, the source sentence text is formatted in various tags as per the user requirement. To make a complete usable transformation of the source text into the target text, translation output should retain the format in translated target text which was present in the source.

Fig. 2 shows RTF and HTML files after rebuilding with translated output text. Here as we can see the word order of lexicons in target output is shuffled from that of the source sentence. This maintainability gets more complicated when it comes to the retention of this translated text which belongs to different grammar families, English and Indian Languages in our case. Fig. 3 shows how translation shuffles text which makes work intricate.

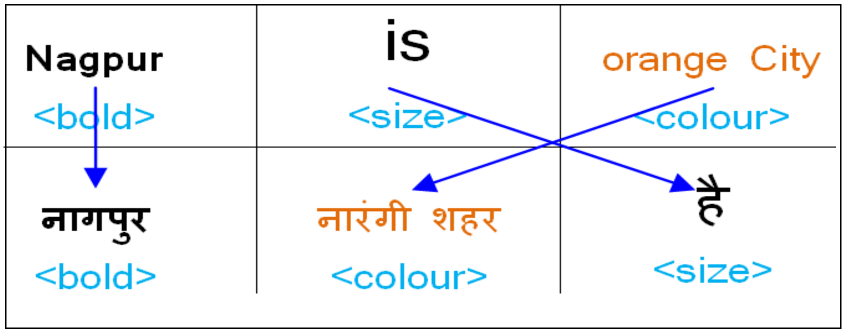


Figure 3. Translation shuffles in the Indian language

A careful planning strategy can overcome this and ensure that you can meet your operational requirements to retain translated documents in original formats. We need a standard file format that aims to hide the complexity of source file formats from translators and tool providers. Here, we are proposing XLIFF as a solution. The purpose of XLIFF along with its different components [5] is to define through XML terminologies, a mechanism that has an extensible specification for the interchange of localization information. At this level, it has become mandatory to carry the localizable data without losing information from one step to other across participants. XLIFF provides competence to capture the mark up localizable data and exchange and use information with different processes or phases without loss of that information. It allows the development of tools that are compatible with the implementer's own proprietary data formats and workflow requirements. In Multilingual Machine Translation System for Indian Languages, the input text specified differs from the translated text immensely. The MTS based on TAG [19] parser and generator trees state dependencies between nodes of trees. In TAG formalism there is a derivation tree for each derivation acquired. This tree contains nodes for the initial and auxiliary trees and edges for adjunctions and substitutions taken place in the process. Thus this constrained mathematical formalism supports the development of lexicalized grammar using these sets of trees. In such kind of system, the need for segregation of translatable text becomes imperative. XLIFF contributes highly to storing the contents, i.e. the translations are captured rather than the formalization.

4. Proposed system

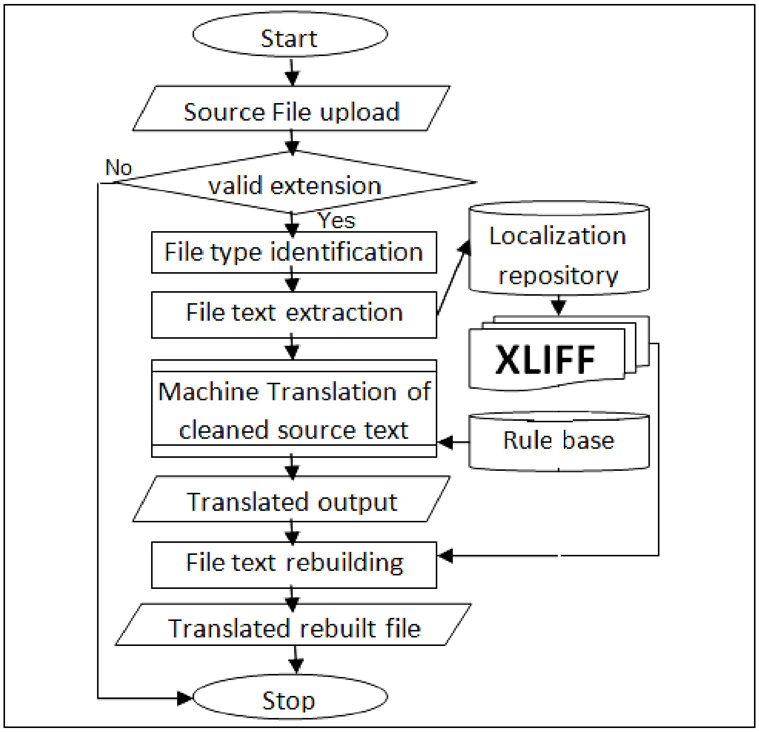


Figure 4. Flowchart for XLIFF format in MT system

This proposed system specifies a tool, which is independent and standardized for localization and supporting its process. Apart from having a Machine Translation component in it, it has two main parts as Format Extraction and Format Rebuilding. The main process of this system is divided into three steps: input format extraction, machine translation and rebuilding of the output format.

Here the process of machine translation is carried out by a rule based approach which consists of a collection of rules called grammatical rules, translated lexicons and software programs that will practice these rules. Rules play a vital role in distinct phases of the translation mechanism. This includes syntactic practicing, semantic analysis and contextual practicing of the translatable language. These phases can be measured in many ways but rule based approach can deeply analyze syntax and semantic levels which is mandatory in Indian language grammars. The translation is done by pattern matching of rules. Language understanding is carried out based on knowledge and reasoning. To compute the logical form, grammatical rules are used and further these rules are used to undergo parsing and generation processes using the elementary tree sets and appropriate lexicons from language dictionaries respectively. The rules are independent of language so this set of patterns can be applicable for various Indian languages. This ultimately leads to translated output which is generated in user-specified language. This translated output is used for the rebuilding of the source file which is converted to XLIFF to rebuilt it in the required format. To accelerate the activity of the online translation system process we used Translation Memory (TM), which is a database of the Input source text and its corresponding Output target text. Machine Translation is a sluggish and complex task that is very time-consuming so speeding up such a system becomes essential. Thus we introduced TM which executes at the backend in this MT system. This TM stores the source text in the form of sentence ID which is a mathematical calculation of ASCII values of characters including summation of language ID and Domain ID which are again ASCII values of respective language and domain where the user is working for translation. A lookup is performed in Translation Memory (TM) while every input takes place. This avoids the re-parsing and re-generation of every already generated sentence. This TM match allows us to use similar sentences again and again without going through all the generation modules. A large file is taken as an input to MTS to cover many TAG parsing and generation of sentences to get the outputs of respective inputs for storing in TM. Translation Memory (TM) plays an important role in the MTS system as the processing of each sentence even on repetition requires more memory and time.

Fig. 4 demonstrates the flowchart for our tool of format extraction and retention process using XLIFF in Machine Translation System from English to Indian Languages. Users can upload data in a proprietary format with a valid extension. To render translation in the expected target Indian language user need to select the target language for translating the source text. File type identification is done to keep the original file format and source localization related data is extracted from the file and given for translation.

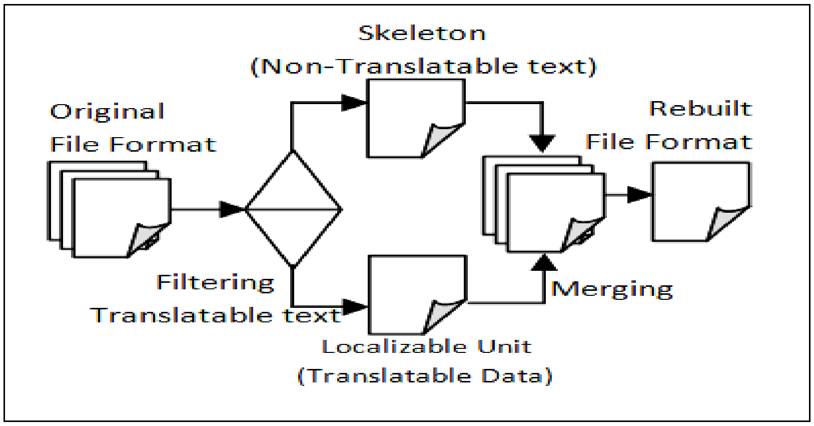
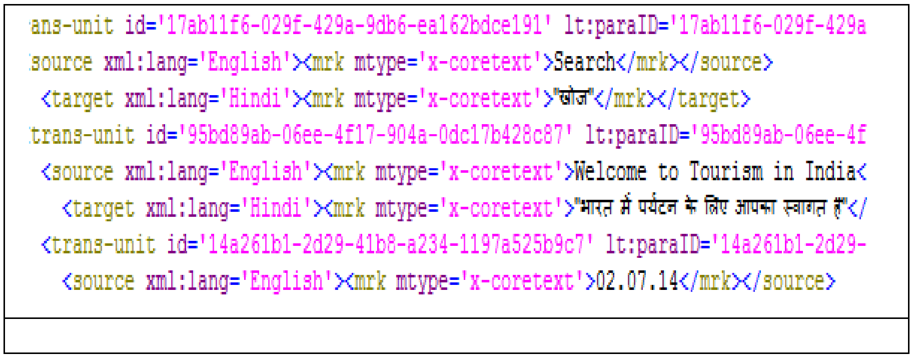


Figure 5. Architecture of XLIFF format in MT system

The XLIFF file initially contains just the source text and after processing, the translation for each segment is added to the file. We extract the various segments into an XLIFF file and put unique identifiers as placeholders. XLIFF provides rich semantic tags and can be used in many different ways, but still provide a consistent structure that is conducive to interoperability between tools. Fig. 5 illustrates the architectural view of XLIFF format extraction and rebuilding in a machine translation system.

In the filtering process, the Input Format Extraction module converts the source document to plain text. In this process, it extracts all the relevant information like file headers, formatting tags and extra spaces and stores for future retention purposes. This filtered translatable text is then stored in an XMLized document called XLIFF_Doc as shown in Fig. 6. XLIFF_Doc provides a standard format and all important translation tools support to accomplish these exchanges. It provides a mechanism to generate localized files from an XML format, as presented in [12]. The <trans-unit> tag contains a <source> and <target> element. This translatable text is sent for translation to chosen Machine Translation System. These translated outputs are reinserted into original format documents.



```
ans-unit id='17ab11f6-029f-429a-9db6-ea162bdce191' lt:paraID='17ab11f6-029f-429a
source xml:lang='English'><mrk mtype='x-coretext'>Search</mrk></source>
 <target xml:lang='Hindi'><mrk mtype='x-coretext'>"खोज"</mrk></target>
trans-unit id='95bd89ab-06ee-4f17-904a-0dc17b428c87' lt:paraID='95bd89ab-06ee-4f
 <source xml:lang='English'><mrk mtype='x-coretext'>Welcome to Tourism in India<
   <target xml:lang='Hindi'><mrk mtype='x-coretext'>"भारत में पर्यटन के लिए आपका स्वागत है"</
 <trans-unit id='14a261b1-2d29-41b8-a234-1197a525b9c7' lt:paraID='14a261b1-2d29-
   <source xml:lang='English'><mrk mtype='x-coretext'>02.07.14</mrk></source>
```

Figure 6. Information stored in XLIFF_Doc

In the merging process, the Output Format Rebuilding module retains the file format after getting the final translated output. It uses machine-translated output text and format information stored during input format extraction. The output of the rebuilding module would be a translated file with the original input format or format as per the user's requirement.

The proposed system is independent of the formalism used in the MT system. Here in this experiment, we have used a rule based machine translation system. Being belong to a divergent language family, the word order among source and target languages differ widely. The rule based MT System is capable to maintain the linkage of the target corresponding to source tokens. To uphold this, linguistic morphological knowledge, syntactic rules and semantic analysis corresponding to the source and target languages take place. Analyzer for structure analysis of source sentence has been identified using a natural language parser. Target language gets generated with a Natural language generator using transfer grammar linkage. This mapping has been maintained at the sentence level and token level of source and target language during the process.

5. Machine Translation System (MTS)

Every MTS requires software programs to perform translations and grammars and automatic dictionaries that will support translations. The ideal aim of every MTS is to generate the best and appropriate possible translation of a given source sentence without any human assistance. This is achieved by some variations that take place at various stages. The quality of generated output can be improved by pre-editing the source text or input text. Pre-editing is nothing but compensating the input text by marking some clause boundaries, prefixes, suffixes and inserting some pre-processing rules which will simplify the input for the system to understand. Controlling the vocabulary is also one of the ways which may give a fine quality to the translated output. Post-editing is done to make the output perfect which also means the synthesis of the generated output. This provides semantic to the generated target output sentence.

This is a web-based MT application that is developed for a specific domain. The various phases include interdependent modules such as pre-processing, pre-parser, TAG based parser and generator, morphological analyzer and synthesizer. In the pre-processing module, we analyze the input source text lexicons using lexical analysis. This analysis isolates the text into different paragraphs, paragraphs into different sentences and sentences into words(lexicons). Named Entity Recognition (NER) rules[25] are used for performing basic chunking of these lexicons along with phrase marking which involve linguistic heuristics. In the pre-parser module, the lexicons are classified known as Part of Speech (POS). This is known as a POS tagger. Here tagging of lexicons is done which are linguistic categories of words or lexical items. This is given based on the semantic and morphological behavior of the word. This POS tagging is parsed and generated using TAG. The TAG parser syntactically and semantically assembles this category-tagged source sentence sequence and converts it to derivation tree format which has all the elementary tree sets. The category sequence generated is independent of words in the source sentence and thus can be reused for input sentences that have similar category sequences. The category sequence is adjusted according to the input derivation structure and output derived tree [24].

In terms of the selection of trees, TAG parsing and generation involves consequential complexity of calculation of substitution and adjunction of trees on one another. The complexity of the parsing algorithm in TAG is O(n6) based on time and it is O(n4) based on space concerning the length of the source text [19]. A compiled sequence thus generated is given to the TAG generator to interpret the lexicons of the target language for a generation. Subsequently, further smoothening of output for fluency is carried out by a language synthesizer module. Here the derived structure is given for lexicalization with the target language lexicons. A bunch of linguistic rules is applied to this lexicalized derived structure to give grammatical meaning which will generate a refined target output. Let us see an example of the above translation steps in MTS when an English sentence is given as a source sentence for translation.

A. **Input Sentence:** The best time to visit Jaipur city is between October and March.

B. **Pre-Processed Output:** 0=The 1=best-time 2=to 3=visit 4=Jaipur-city 5=is 6=between 7=October 8=and 9=March

C. **Pre-Parsed Source Sentence Structure:** 0=The-best-time=NN 1=to-visit=VB 2=Jaipur-city=NN 3=is=VBZ 4=between=RP 5=October=NN 6=and=CC 7=March==NN

D. **Generated Lexicalized Target Sequence:** 2=NN=□□□□□-□□□ 1=VB=□□□□□-□□-□□□ 0=NN=□□□□□□□□□ □□□ 5=NN=□□□□□□□ 6=CC=□□ 7=NN=□□□□□ 4=RP=□□-□□□ 3=VBZ=□□

Here in the above example, as said earlier during pre-processing the long input sentence is segmented into short sentences and punctuation marks if any are fixed with the use of delimiters also the material not required for translation is blocked. Blocking the inappropriate material proves very fruitful for getting a successful translation. Part Of Speech (POS) category of each token (word) is given to get category sequence in the pre-parser module which is the pre-parsed structure of the sentence. This structure is the input of the generator module where the target category sequence is generated. This structure is generated using TAG parsing formalism. This category sequence analyzes each token along with their respective categories and outputs a lexicalized target structure sequence. Post-editing which is also known as synthesis is done to assure the translation quality along with grammatical improvements. This process induces grammatically meaningful sentences. This online translation engine generated output is passed on for rebuilding in the user specified format which is processed via XLIFF format.

6. Translation memory for machine translation (TM) database

XLIFF is a format for transacting between content and respective files used for localization if we stand for XMLized Localization Interchange File Format. The translatable text is separated into the localizable unit which is the text which undergoes translation. Translation memory enables the translation of segments which include sentences, phrases, or paragraphs present in documents to be searched for the segments in the database to replace directly the translated target text despite undergoing the stages of translation. This is beneficial for large data translations.

Translation memory is built by storing source text with its corresponding target text while first time translation. The source text is given a sentence ID which is numeric calculus of ASCII character of every character in the source string which also succeeds with language ID which is again mathematical calculation of ASCII character of every character in language string.

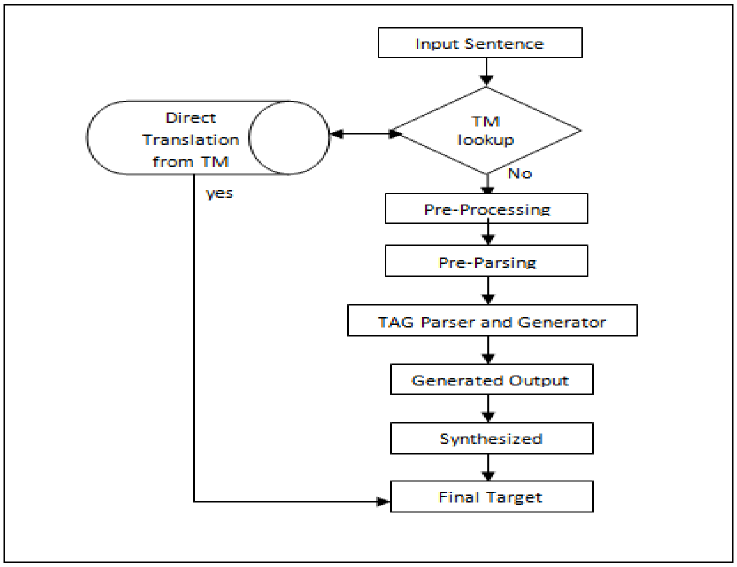


Figure 7. Flowchart for MT system using TM

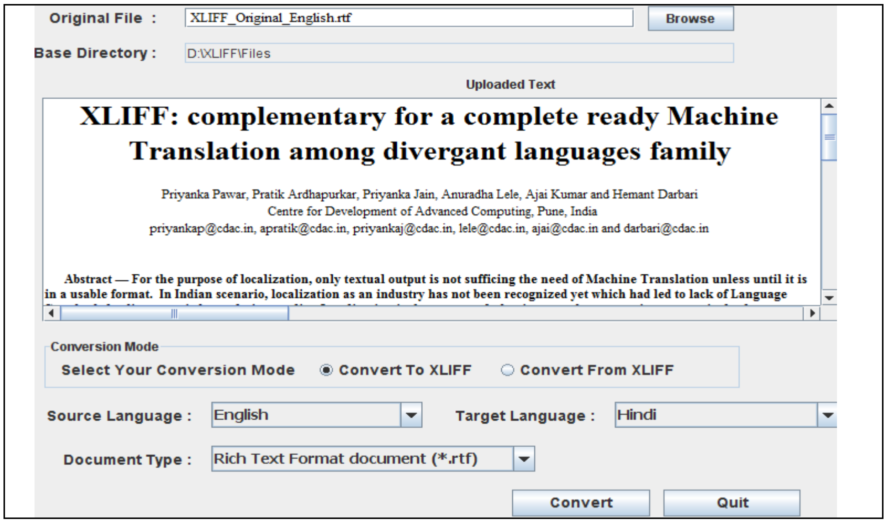
This sentence ID along with source text and its corresponding target text is stored in the database. Whenever the segments get repeated this sentence ID is matched for check and its existence provides target text without going for all phases of the translation system.

Above [Figure 7] shows the flowchart of the MT System using Translation Memory. The workflow changes drastically as the source sentence gets direct target text i.e translated output of required target language after performing the lookup if the source-target sentence pair is stored in DB while first execution. The lookup is performed using sentence ID which is calculated for source sentence using ASCII code of each character and language ID which is again calculated using ASCII code of target language selected. This sentence ID is the summation of this code which stays unique for similar sentences. The parsing and generation steps are followed in the only condition if the input source sentence or segments of the same are not present in the TM of the database. This has accelerated the system which is the primary requirement of an online machine translation system.

7. Experiments and result

We have designed and implemented a Graphical User Interface of a conversion tool as shown in Fig. 8 developed in Java using NetBeans IDE [10]. It automates and thus accelerates the process of file format extraction to re-format translated data using XLIFF. The user uploads the input file in the desired data format. This file is converted to XLIFF by selecting the conversion mode after translation into the target language. This tool provides the facility of conversion of file format to XLIFF and XLIFF to file format. The user selects the desired conversion mode and file format to send the file for machine translation. Machine Translation System goes through various pre-post and parsing-generation modules where rules play a key role which is based on the grammatical linguistic study.

The translation takes place through Machine Translation System (MTS) using Tree Adjoining Grammar (TAG) which is a Rule Based Machine Translation technique. We have also included the translation memory mechanism in this tool at the back end. The translation process provides output in the selected target language by the user. This MTS system can translate text to 8 Indian Languages including Hindi, Marathi, Oriya, Urdu, Tamil, Bengali, Gujarati and Bodo. As a grammatical study in Indian languages vary the rules depend and differ for every individual language. The patterns implemented based on these rules are matched and a sequence of the category is generated. This generated sequence is lexicalized which



Original File : XLIFF_Original_English.rtf

Base Directory : D:\XLIFF\Files

Uploaded Text

XLIFF: complementary for a complete ready Machine Translation among divergant languages family

Priyanka Pawar, Pratik Ardhapurkar, Priyanka Jain, Anuradha Lele, Ajai Kumar and Hemant Darbari
Centre for Development of Advanced Computing, Pune, India
priyankap@cdac.in, apratik@cdac.in, priyankaj@cdac.in, lele@cdac.in, ajai@cdac.in and darbari@cdac.in

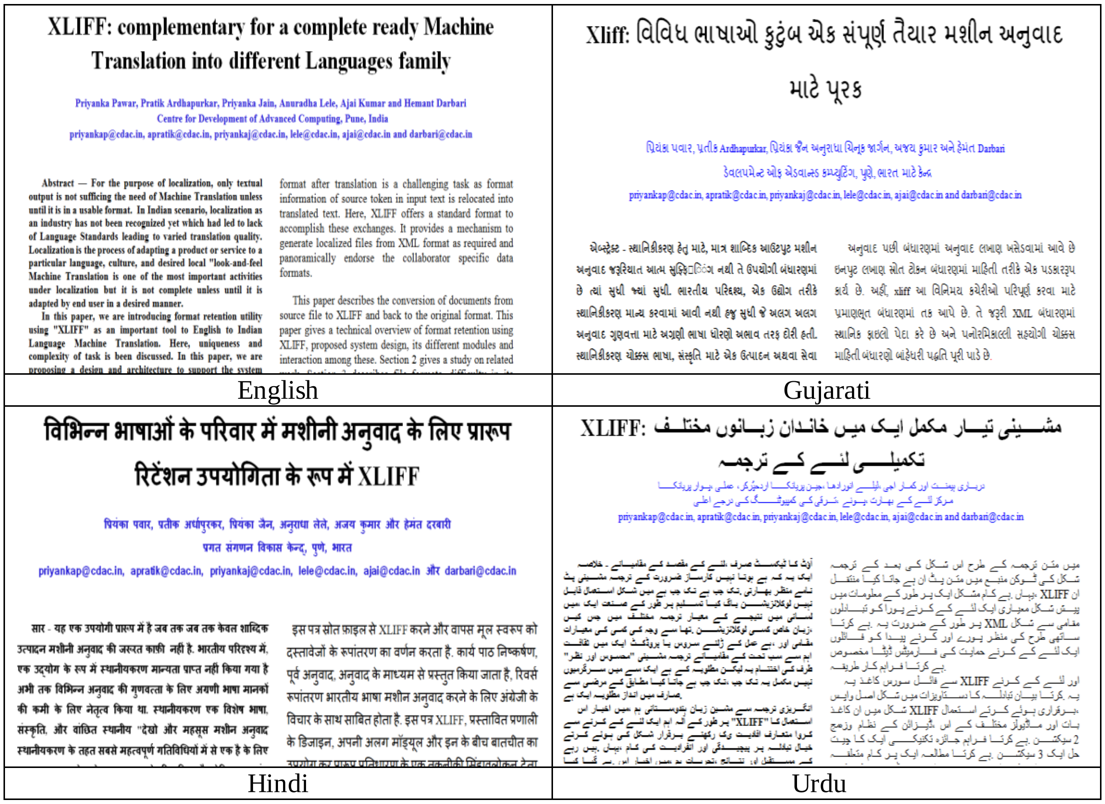
Abstract — For the purpose of localization, only textual output is not sufficing the need of Machine Translation unless until it is in a usable format. In Indian scenario, localization as an industry has not been recognized yet which had led to lack of Language

Conversion Mode
Select Your Conversion Mode ● Convert To XLIFF ○ Convert From XLIFF

Source Language : English
Target Language : Hindi
Document Type : Rich Text Format document (*.rtf)

Figure 8. GUI for format conversion tool

produces the target output. This targeted output is transformed into an XLIFF file. The translated file acquires the original format of the source file using extracted tag of the source file. While retaining user has to again select the conversion mode to convert the file from XLIFF to the required format. The translated XLIFF file will get converted to the original format or specified format in document type with formatting and styling present in the original input file. The text area provided in the tool visualizes the input original file data to the user before translation. This file can also be edited in this text area before proceeding it for the translation process. XLIFF format also gives the flexibility to retain a file in a different format than the original format. The experiment has been done on RTF, HTML, and XLS file formats. The conversion tool is capable to retain font color, font size, font features like bold, underline, italic. It can handle tables, images and header/footer format of the document. Indentation, justification, bulletins, indexing, paragraph and page formatting is also being retained after translation. The Format Conversion Tool retains various formats using XLIFF and has made rebuilding the translation of Indian languages a smooth process.



XLIFF: complementary for a complete ready Machine Translation into different Languages family

Priyanka Pawar, Pratik Ardhapurkar, Priyanka Jain, Anuradha Lele, Ajai Kumar and Hemant Darbari
Centre for Development of Advanced Computing, Pune, India
priyankap@cdac.in, apratik@cdac.in, priyankaj@cdac.in, lele@cdac.in, ajai@cdac.in and darbari@cdac.in

Abstract — For the purpose of localization, only textual output is not sufficing the need of Machine Translation unless until it is in a usable format. In Indian scenario, localization as an industry has not been recognized yet which had led to lack of Language Standards leading to varied translation quality. Localization is the process of adapting a product or service to a particular language, culture, and desired local "look-and-feel Machine Translation is one of the most important activities under localization but it is not complete unless until it is adapted by end user in a desired manner.

In this paper, we are introducing format retention utility using "XLIFF" as an important tool to English to Indian Language Machine Translation. Here, uniqueness and complexity of task is been discussed. In this paper, we are proposing a design and architecture to support the system format after translation is a challenging task as format information of source token in input text is relocated into translated text. Here, XLIFF offers a standard format to accomplish these exchanges. It provides a mechanism to generate localized files from XML format as required and panoramically endorse the collaborator specific data formats.

This paper describes the conversion of documents from source file to XLIFF and back to the original format. This paper gives a technical overview of format retention using XLIFF, proposed system design, its different modules and interaction among these. Section 2 gives a study on related

English

Xliff: વિવિધ ભાષાઓ કુટુંબ એક સંપૂર્ણ તૈયાર મશીન અનુવાદ માટે પૂરક

પ્રિયંકા પવાર, પ્રતીક Ardhapurkar, પ્રિયંકા જૈન અનુરાધા ચિનૂક જાર્ગન, અજય કુમાર અને હેમંત Darbari
ડેવલપમેન્ટ ઓફ એડવાન્સ્ડ કમ્પ્યુટિંગ, પુણે, ભારત માટે કેન્દ્ર
priyankap@cdac.in, apratik@cdac.in, priyankaj@cdac.in, lele@cdac.in, ajai@cdac.in and darbari@cdac.in

એબ્સ્ટ્રેક્ટ - સ્થાનિકીકરણ હેતુ માટે, માત્ર શાબ્દિક આઉટપુટ મશીન અનુવાદ જરૂરિયાત આત્મ સુફિ□િંગ નથી તે ઉપયોગી બંધારણમાં છે ત્યાં સુધી જ્યાં સુધી. ભારતીય પરિદ્દશ્ય, એક ઉદ્યોગ તરીકે સ્થાનિકીકરણ માન્ય કરવામાં આવી નથી હજુ સુધી જે અલગ અલગ અનુવાદ ગુણવત્તા માટે અગ્રણી ભાષા ધોરણો અભાવ તરફ દોરી હતી. સ્થાનિકીકરણ ચોક્કસ ભાષા, સંસ્કૃતિ માટે એક ઉત્પાદન અથવા સેવા

અનુવાદ પછી બંધારણમાં અનુવાદ લખાણ ખસેડવામાં આવે છે ઇનપુટ લખાણ સ્રોત ટોકન બંધારણમાં માહિતી તરીકે એક પડકારરૂપ કાર્ય છે. અહીં, xliff આ વિનિમય કચેરીઓ પરિપૂર્ણ કરવા માટે પ્રમાણભૂત બંધારણમાં તક આપે છે. તે જરૂરી XML બંધારણમાં સ્થાનિક ફાઇલો પેદા કરે છે અને પનોરમિકાલ્લી સહયોગી ચોક્કસ માહિતી બંધારણો બાંહેધરી પદ્ધતિ પૂરી પાડે છે.

Gujarati

विभिन्न भाषाओं के परिवार में मशीनी अनुवाद के लिए प्रारूप रिटेंशन उपयोगिता के रूप में XLIFF

प्रियंका पवार, प्रतीक अर्धापुरकर, प्रियंका जैन, अनुराधा लेले, अजय कुमार और हेमंत दरबारी
प्रगत संगणन विकास केन्द्, पुणे, भारत
priyankap@cdac.in, apratik@cdac.in, priyankaj@cdac.in, lele@cdac.in, ajai@cdac.in और darbari@cdac.in

सार - यह एक उपयोगी प्रारूप में है जब तक जब तक केवल शाब्दिक उत्पादन मशीनी अनुवाद की जरूरत काफी नहीं है. भारतीय परिदृश्य में, एक उद्योग के रूप में स्थानीयकरण मान्यता प्राप्त नहीं किया गया है अभी तक विभिन्न अनुवाद की गुणवत्ता के लिए अग्रणी भाषा मानकों की कमी के लिए नेतृत्व किया था. स्थानीयकरण एक विशेष भाषा, संस्कृति, और वांछित स्थानीय "देखो और महसूस मशीन अनुवाद स्थानीयकरण के तहत सबसे महत्वपूर्ण गतिविधियों में से एक है के लिए

इस पत्र स्रोत फ़ाइल से XLIFF करने और वापस मूल स्वरूप को दस्तावेजों के रूपांतरण का वर्णन करता है. कार्य पाठ निष्कर्षण, पूर्व अनुवाद, अनुवाद के माध्यम से प्रस्तुत किया जाता है, रिवर्स रूपांतरण भारतीय भाषा मशीन अनुवाद करने के लिए अंग्रेजी के विचार के साथ साबित होता है. इस पत्र XLIFF, प्रस्तावित प्रणाली के डिजाइन, अपनी अलग मॉड्यूल और इन के बीच बातचीत का उपयोग कर प्रारूप प्रतिधारण के एक तकनीकी सिंहावलोकन देता

Hindi

XLIFF: مشـــینی تیـــار مکمل ایـک میں خانـدان زبـانوں مختلـف تکمیلـــی لئـے کـے ترجمـہ

دربـاری ہیمنت اور کمار اجی ،لیلے انورادھا ،جین پریانکـــا اردھپُرکر، عملی ،پـوار پریانکـــا
مرکز لئـے کـے بھارت ،پونـے ،ترقی کی کمپیوٹنـــگ کی ترجمے اعلی
priyankap@cdac.in, apratik@cdac.in, priyankaj@cdac.in, lele@cdac.in, ajai@cdac.in and darbari@cdac.in

آؤٹ کا ٹیکسٹ صرف ،لئے کے مقصد کے مقامیانے ۔ خلاصہ ایک یہ کہ ہے ہوتا نہیں کارساز ضرورت کے ترجمہ مشینی پٹ تامے منظر بھارتی ،تک جب ہے تک جب ہے میں شکل استعمال قابل نہیں لوکلائزیشن یاگ کیا تسلیم پر طور کے صنعت ایک ،میں لسانی میں نتیجے کے معیار ترجمہ مختلف میں جس کیں ،زبان خاص کسی لوکلائزیشن ،تھا سے وجہ کی کمی کی معیارات مقامی اور ،ہے عمل کے ڈھلنے سروس یا پروڈکٹ ایک میں ثقافت اہم سے سب تحت کے مقامیانے ترجمہ مشینی "محسوس اور نظر" طرف کی اختتام یہ لیکن مطلوبہ کے ہے ایک سے میں سرگرمیوں نہیں مکمل یہ تک جب ،تک جب ہے جاتا کیا مطابق کے مرضی سے ،صارف میں انداز مطلوبہ ایک ہے

انگریزی ترجمہ سے مشین زبان ہندوستانی ہم ،میں اخبار اس استعمال کا "XLIFF" پر طور کے آلہ اہم ایک لئے کے کرنے سے کروا متعارف افادیت ی ک رکھنے برقرار شکل کی ہونے کرتے خیال تبادلہ پر پیچیدگی اور انفرادیت کی کام ،یہاں ،ہیں رہے

میں متن ترجمہ کے طرح اس شکل کی بعد کے ترجمہ شکل کی ٹوکن منبع میں متن پٹ ان ہے جاتا کیا منتقل ان XLIFF ،یہاں ہے کام مشکل ایک پر طور کے معلومات میں پیش شکل معیاری ایک لئے کے کرنے پورا کو تبادلوں مقامی سے شکل XML پر طور کے ضرورت یہ ۔ہے کرتا ساتھی طرح کی منظر پورے اور کرنے پیدا کو فائلوں ایک لئے کے کرنے حمایت کی فارمیٹس ڈیٹا مخصوص ۔ہے کرتا فراہم کار طریقہ

اور لئے کے کرنے XLIFF سے فائل سورس کاغذ یہ یہ ۔کرتا بیان تبادلہ کا دستاویزات میں شکل اصل واپس ،برقراری ہونے کرتے استعمال XLIFF شکل میں ان کاغذ بات اور ماڈیولز مختلف کے اس ،ڈیزائن کے نظام وزمج 2 سیکشن ۔ہے کرتا فراہم جائزہ تکنیکی ایک کا چیت حل ایک 3 سیکشن ۔ہے کرتا مطالعہ ایک پر کام متعلقہ

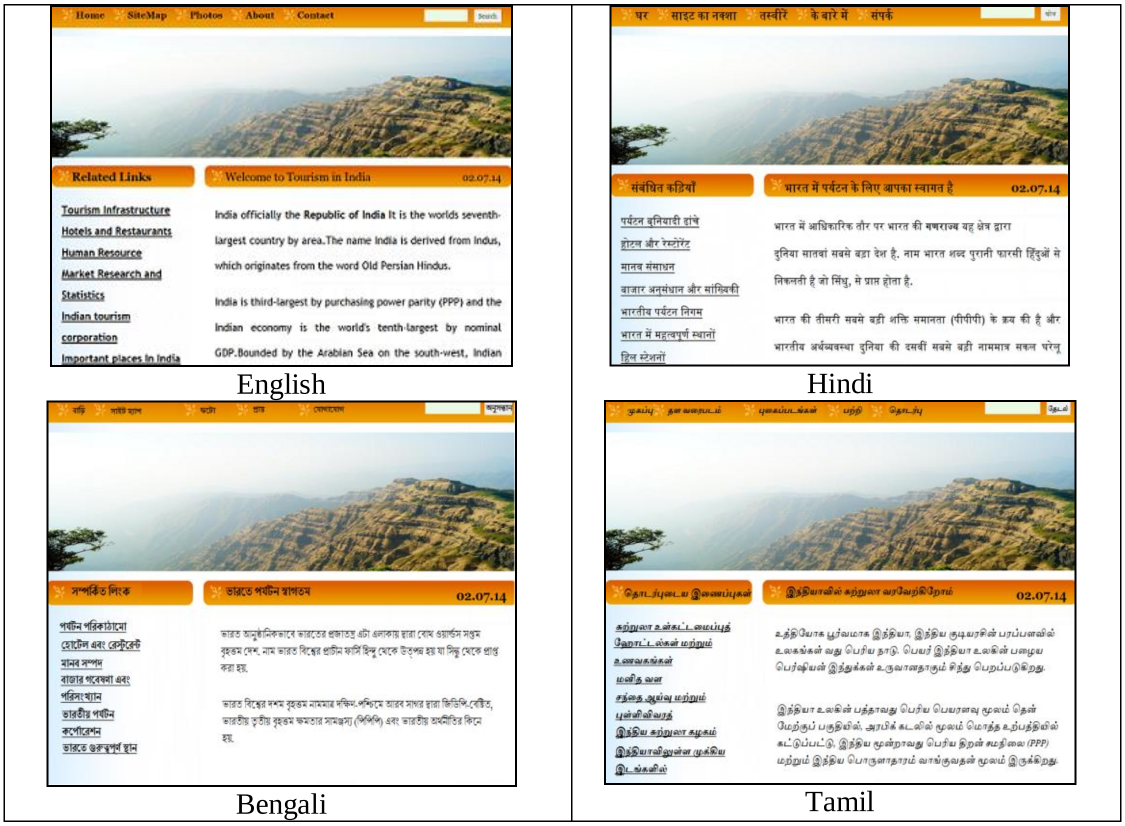
Urdu

Figure 9. Example of RTF tags extraction and rebuilding in Hindi and Urdu language respectively

As mentioned above, the formats like page, column, paragraph, font color, size and style have been retained as shown in [Figure 9] for RTF. Here, the input file is in RTF format as shown in ‘English’ [Figure 9]. This file is extracted and translatable data or text from the original document was taken for translation. Special programs called filters separate text and design. The non-translatable data which comes in the formatting part or designing part was stored in a skeleton file. Each translatable text fragment is reserved in a translation unit element called <trans-unit> in an XLIFF file. Skeleton had the mark of the id attribute of translation unit which simplified the mapping between skeleton and XLIFF file.

The filter intricacy is on which format is been parsed. As in the above example, RTF files have many tags to handle and retain as compared to an HTML file. The XML is a well documented format so XLIFF being an XML application can be said to be very clear, concise and practical informal specification. The translated XLIFF file after translation is ready to merge with the skeleton file based on special marks. If the translation does not take place for the segment, or if the translation which was included is not approved, the data in <source> tag or element is used as it is. After the insertion of every mark present in the XLIFF file, the skeleton file now becomes the translated document. Most documents require post fixation in the original document, whereas HTML, RTF and XML are formats that comparatively require very few post translation arrangements. Moreover, the Indian language itself varies in their structure so post fixation in such cases becomes mandatory. This also increases the complexity of filters. Thus translating a document using XLIFF as an intermediate file format.

XLIFF benefits with a set of mechanisms for pre-translations, modifications and diversity recording. It also supports widely binary objects like bitmap, icon, etc to be reserved along with the text. These advantages make RTF rebuilding an easy process in any target language. As all word processors and many other programs can read and excess RTF texts and documents easily this file format is majorly used. We have extracted data from the RTF file and rebuilt it after translation in different Indian languages. We had translated the English file in Gujarati, Hindi and Urdu languages and retained the formatting and styling of the original source file as shown in [Figure 9] above.



Related Links

Welcome to Tourism in India 02.07.14

Tourism Infrastructure
Hotels and Restaurants
Human Resource
Market Research and Statistics
Indian tourism corporation
Important places In India

India officially the Republic of India It is the worlds seventh-largest country by area.The name India is derived from Indus, which originates from the word Old Persian Hindus.

India is third-largest by purchasing power parity (PPP) and the Indian economy is the world's tenth-largest by nominal GDP.Bounded by the Arabian Sea on the south-west, Indian

English

संबंधित कड़ियाँ

भारत में पर्यटन के लिए आपका स्वागत है 02.07.14

पर्यटन बुनियादी ढांचे
होटल और रेस्टोरेंट
मानव संसाधन
बाजार अनुसंधान और सांख्यिकी
भारतीय पर्यटन निगम
भारत में महत्वपूर्ण स्थानों
हिल स्टेशनों

भारत में आधिकारिक तौर पर भारत की गणराज्य यह क्षेत्र द्वारा दुनिया सातवां सबसे बड़ा देश है. नाम भारत शब्द पुरानी फारसी हिंदुओं से निकलती है जो सिंधु, से प्राप्त होता है.

भारत की तीसरी सबसे बड़ी शक्ति समानता (पीपीपी) के क्रय की है और भारतीय अर्थव्यवस्था दुनिया की दसवीं सबसे बड़ी नाममात्र सकल घरेलू

Hindi

সম্পর্কিত লিংক

ভারতে পর্যটন স্বাগতম 02.07.14

পর্যটন পরিকাঠামো
হোটেল এবং রেস্টুরেন্ট
মানব সম্পদ
বাজার গবেষণা এবং পরিসংখ্যান
ভারতীয় পর্যটন কর্পোরেশন
ভারতে গুরুত্বপূর্ণ স্থান

ভারত আনুষ্ঠানিকভাবে ভারতের প্রজাতন্ত্র এটা এলাকায় দ্বারা বোধ ওয়ার্ল্ডস সপ্তম বৃহত্তম দেশ. নাম ভারত বিশ্বের প্রাচীন ফার্সি হিন্দু থেকে উত্পন্ন হয় যা সিন্ধু থেকে প্রাপ্ত করা হয়.

ভারত বিশ্বের দশম বৃহত্তম নামমাত্র দক্ষিণ-পশ্চিমে আরব সাগর দ্বারা জিডিপি-বেষ্টিত, ভারতীয় তৃতীয় বৃহত্তম ক্ষমতার সামঞ্জস্য (পিপিপি) এবং ভারতীয় অর্থনীতির কিনে হয়.

Bengali

தொடர்புடைய இணைப்புகள்

இந்தியாவில் சுற்றுலா வரவேற்கிறோம் 02.07.14

சுற்றுலா உள்கட்டமைப்புத்
ஹோட்டல்கள் மற்றும் உணவகங்கள்
மனித வள
சந்தை ஆய்வு மற்றும் புள்ளிவிவரத்
இந்திய சுற்றுலா கழகம்
இந்தியாவிலுள்ள முக்கிய இடங்களில்

உத்தியோக பூர்வமாக இந்தியா, இந்திய குடியரசின் பரப்பளவில் உலகங்கள் வது பெரிய நாடு. பெயர் இந்தியா உலகின் பழைய பெர்ஷியன் இந்துக்கள் உருவானதாகும் சிந்து பெறப்படுகிறது.

இந்தியா உலகின் பத்தாவது பெரிய பெயரளவு மூலம் தென் மேற்குப் பகுதியில், அரபிக் கடலில் மூலம் மொத்த உற்பத்தியில் கட்டுப்பட்டு, இந்திய மூன்றாவது பெரிய திறன் சமநிலை (PPP) மற்றும் இந்திய பொருளாதாரம் வாங்குவதன் மூலம் இருக்கிறது.

Tamil

Figure 10. Example of HTML tags extraction and rebuilding of Bengali, Hindi and Tamil languages respectively

Format retention is more beneficial when it is used for website conversion as it provides the original look and feel of the web portal to the general mass in their desired language. Fig. 10 shows Format Retention of English webpage to Bengali, Hindi and Tamil Languages for HTML format. Lack of standards and Unicode operability and support is still a crucial aspect of Indian languages. XLIFF is a simple XML format and is used more or less solely for translation. Even though it is a recommended multi-platform open-source editor from the translation toolkit, it certainly has some pros and cons. XLIFF's structure needs to have increased simplicity, modularization, clarification of necessary metadata, and better workflow control. However, offline translators need to get familiar with the new tools.

8. Conclusion

XLIFF tenders a wide range of features that ease the storage of metadata consociated with the localization process. Particularly, it allows translation customers to provide their source data text with an optional set of contextual information, Translation Memory, reusable translations, font information, maximum text length, identifier and type of resource, notes and many other types of data. XLIFF is much more versatile, flexible and influential than any current format used by commercial translation or localization tools. In addition, the upcoming version will offer many ways to extend and customize it.

This paper has presented a technology to retain the format of the document after translating from the English Language to Indian Languages using XLIFF as an intermediate file format. We have also introduced machine translation techniques for Indian Languages using TAG formalism and intermediate modules of the same and also translation memory to speed up the process of parsing and generation which proves the benefits of non-execution of the tedious processes for every repeating sentence. It has also explained all the series of the process with simple examples. XLIFF is a solution that can be extended to support other formats to provide the translation completely.

References

[1] D. Anastasiou, "Survey on the Use of XLIFF in Localisation Industry and Academia", In: Proceedings of Language Resource and Language Technology Standards – State of the Art, Emerging Needs, and Future Developments Workshop, 7th International Conference on Language Resources and Evaluation (LREC), Malta, pp. 50–53, **(2010)**.

[2] HTML site [online], available: http://html.net [accessed 14jan15]

[3] Microsoft [online], available: http://msdn.microsoft.com/en-us/library/office/jj945830(v=office.15).aspx [accessed 22jan15]

[4] Microsoft [online], available: http://www.microsoft.com/en-in/default.aspx[accessed 22jan15]

[5] Xliff [online], available: http://www.opentag.com/xliff.htm [accessed 22jan15]

[6] Oracle [online], available: http://www.oracle.com/index.html [accessed 22jan15]

[7] Sap [online], available: http://www.sap.com/about.html [accessed 22jan15]

[8] Sdl [online], available: http://www.sdl.com [accessed 22jan15]

[9] XLIFF [online], available: http://www.xliff.org [accessed 22jan15]

[10] Netbeans [online], available: https://netbeans.org [accessed 22jan15]

[11] ITS 2.0 in XLIFF 2: Online available dated 30.10.2014 at - http://www.localizationworld.com/lwdub2014/feisgiltt/slides/ITS_2.0_in_XLIFF_2.pdf

[12] Raya, R. XML Localization Interchange File Format as an intermediate file format. IBM developer Works **(2004)**. Online available dated 30.10.2014 at - http://www.maxprograms.com/articles/xliff.html

[13]

[14] P. Reynolds, and T. Jewtushenko, “What Is XLIFF and Why Should I Use It?: A brief overview of the XML Localization Interchange File Format (XLIFF)”, **(2005)**, XML Journal, Online available dated 30.10.2014 at - http://xml.sys-con.com/node/121957?page=0,3

[15] Y. Savourel, “An Introduction to Using XLIFF”, Online available dated 30.10.2014 at - http://www.multilingual.com/articleDetail.php?id=1178

[16] W3C Internationalization Tag Set, Online available dated 30.10.2014 at - http://www.w3.org/International/multilingualweb/madrid/slides/lieske-its.pdf

[17] Word 2007: Rich Text Format (RTF) Specification, version 1.9.1. Online available dated 30.10.2014 at - http://www.microsoft.com/en-in/download/details.aspx?id=10725

[18] XLIFF Version 1.2 OASIS Standard. Online available dated 30.10.2014 at - http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html

[19] A.K. Joshi, L.S. Levy, and M. Takahashi, “Tree adjunct grammars”, Journal of Computer and System Sciences, Vol. 10, No. 1, pp. 136-163, February **(1975)**.

[20] M. Vasconcellos, and M. Leon, “SPANAM and ENGSPAN: Machine Translation at the Pan American Health Organization”, in J. Slocum (ed.) Machine Translation systems, Cambridge: Cambridge University Press, **(1988)**.

[21] B. Mellebeek, A.Khasin, J.Van Genabith and A.Way, “TransBooster: Boosting the Performance of Wide-Coverage Machine Translation Systems”, In Proceedings of the 10th Annual Conference of the European Association for Machine Translation (EAMT-05), Budapest, Hungary, pp. 189-197, **(2005)**.

[22] A.K. Joshi, L.S. Levy, and M. Takahashi, “Tree Adjunct Grammars”, Journal of Computer and Systems Sciences, vol.10, no.1, pp. 55-75, **(1975)**.

[23] A.K. Joshi, “How much context sensitivity is necessary for characterizing structural descriptions—tree adjoining grammars”, Natural Language Processing—Theoretical, Computational, and Psychological Perspectives, **(1985)**.

[24] J. Rogers, ‘A Unified Notion of Derived and Derivation Structures in TAG’, University of Central Florida, Gainesville, **(1997)**.

[25] D. Nadeau and S. Sekine, “A survey of named entity recognition and classification”, Lingvisticae Investigationes, vol.30, no.1, pp. 3-26, **(2007)**.

[26] L. Bowker, “Computer-Aided Translation Technology: A practical Introduction”, Ottawa: University of Ottawa Press, **(2002)**.

[27] M. Thawabteh, “The Intricacies of Translation Memory Tools: With Particular Reference to Arabic-English Translation”, Al-Quds University, Jerusalem, **(2013)**.

[28] w3schools XML Tutorial [online], available: http://www.w3schools/xml/ [accessed 22 jan 2015]

[29] L.M. Vázquez, and J.T. del Rey, “Teaching XLIFF to translators and localisers”, University of Salamanca, Spain, **(2015)**.

Authors

Ms. Priyanka Pawar is working as Project Engineer, AAI Group, C-DAC, Pune. She has completed her B.Tech and PG Diploma from ACTS-CDAC. Her area of interests are in Machine Translation, web technologies and Brain Computer Interface. She has published various papers in the field. priyankap@cdac.in

Ms. Anuradha Anand Lele is working as Joint Director in the Applied Artificial Intelligence (AAI) Group, C-DAC, Pune. She is handling various projects in the area of Natural Language Processing such as Machine Translation systems and Information extraction & Retrieval systems apart from applications development in Public health. She also has past experience in developing the Management Information systems. lele@cdac.in

Mr. Pratik Ardhapurkar is working as Project Engineer, AAI Group, C- DAC, Pune. He has completed his B.Tech and PG Diploma from CDAC. He is associated in Machine Translation technology and advanced web secutity systems. apratik@cdac.in

Mr. Ajai Kumar is working as Associate Director and Head, AAI Group, C-DAC, Pune. He is handling various projects in the area of Natural Language Processing, Information Extraction and Retrieval, Intelligent Language Teaching/Tutoring, Speech Technology [Synthesis & Recognition ASR], Mobile Computing, Decision Support Systems & Simulations and has published various national & international papers. ajai@cdac.in

Ms. Priyanka Jain is working as Principal Technical Officer, AAI Group, C-DAC, Pune. She has 14 years of experience in technologies related to Natural Language Processing (NLP), Machine assisted Translation (MT), Computer Based Training (CBT), Speech technologies, distributed architectures, Mobile Computing and Visualization. She has published various national & international papers. priyankaj@cdac.in

Dr. Hemant Darbari is working as Executive Director in C-DAC, Pune. He is one of the founding members of C-DAC, an R&D Institute set up by the Department of Electronics and Information Technology; Govt. of India for carrying out advanced research in new and emerging technological domains. He has to his credit, 85 Technical Papers that have been published in national & international Journals & Conference Proceedings. Email: Darbari@cdac.in