ABSTRACT

Transliteration is one of the key areas in natural language processing. Also, transliteration is an essential element of translation. This paper presents details about software interface that accepts a sentence in Marathi as an input, does its exact transliteration into English, and translates the same into English.

The main aim to develop this interface is to provide local language support in the web based commercial applications. There are two main modules of this interface viz. a) Transliteration and b) Translation. Transliteration is the conversion of text from one script to another [3].

Translation can form an essential part of transcription which converts text from one writing system into another [4]. Transliteration gives the word from a different language in letters that you can understand so as to be able to pronounce it [4]. The transliterated sentences are taken into consideration.

The flow of control interface is as follows

- The sentences consist of only 4 words.
- The sentence follows the grammar rules (Parsing tree).

The flow of control of the interface is as follows

1. INTRODUCTION

In the last few years remarkable advancements are seen in information retrieval method as information system faces challenges of managing information overhead. This information overhead is due to widespread of communication networks which in turn gives trigger to increasing usage of the World Wide Web i.e., WWW. This is mainly because of constantly increasing amount of formless data they make available to users. In English, a huge volume of textual data is available on internet and the same is for Marathi language. But, most of the available tools and techniques are English oriented. This shows that there is a lack of tool that does efficient text mining for both Marathi and English. To overcome this constraint, in this paper we have proposed a software interface that does automatic transliteration and translation of Marathi text into English.

The plan of study is divided into six stages of interface with graphical display. In this paper a general framework that can achieve the above mentioned goals and interface analogues to its processing is presented. This interface is useful in indentifying features of natural text like parts of speech through splitting of text into words and achieving the translated output through parsing.

Marathi is an Indo-Aryan language and it is the official language of Maharashtra state of India [2]. It is one of the 23 official languages of India [2]. It is the 19th most spoken language in the world [2]. Marathi has 4th largest number of native speakers in India [2]. The major varieties of Marathi are called

- Standard Marathi and
- Warhadi Marathi [2].

There are a few other varieties of it like Ahirani, Dangi, Vadvali, Samavedi, Khandeshi, and Malwani [2].

These statistics shows that Marathi is used by a large number of people and also communication relation between English and Marathi increases which in turn creates need for automatic, efficient and effective translation of these two languages. Translation activity has to consider various issues like vocabulary, grammar, context and semantics. The automatic translation is a bit tough process as languages considerably differ in their characteristics. The translation interface which is discussed in this paper is a miniature in this direction. Many existing automatic translation systems are available and they are working on rule-based approach.

2. IMPLEMENTATION DETAILS ABOUT TRANSLATION INTERFACE

The sentences are to be broke down into parts of speech through splitting of text into words and achieving the translated output through parsing.
boxes for input and outputs as well as to capture the intermediate stages of transliteration and translation.

Figure 2: Parse tree showing parts of speech
The following picture shows translation interface layout.

Image 1: Translation interface layout
By clicking buttons on the virtual keyboard, the input gets stored in the text box as shown in the image 3. On clicking “Split into words” button, all words of the input sentence splits and gets stored into different text boxes. This functionality is achieved by using following code snippet.

```csharp
protected void button2_Click(object sender, EventArgs e)
{
    string abc = msg.Text;
    int count = abc.Split(' ').Length;
    string[] words = abc.Split(' ');
    foreach (string word in words)
    {
    }
}
```

Image 2: Code snippet for splitting into words

Image 3: Inputting a Marathi sentence by using virtual keyboard
The connection between the code and the database is made. The English translation and part of speech for each word is retrieved from the database.

```csharp
    protected partial class Index : System.Web.UI.Page
    
    protected void Page_Load(object sender, EventArgs e)
    
    }
```

Image 4: Code snippet for database connection
The code to configure web.config file so that connection will be established with database is as given below.

Image 5: Code snippet for web.config file
The code for opening and closing connection with database is as given below.

```csharp
<connectionStrings>
    <add name="MyConnection" connectionString="Data Source=MyServer;Initial Catalog=MyDatabase;User ID=MyUser;Password=MyPassword" providerName="System.Data.SqlClient" />
</connectionStrings>
```

Image 6: Code snippet for opening and closing of database connection
The format in which Microsoft Access database is created for the interface is as given below.

Image 7: Database of Marathi words
The same database format for exact transliteration of Marathi to English is prepared as given below.

Image 8: Database for exact transliteration of Marathi words into English
The following image shows splitting of words for the entered Marathi sentence. Also, it shows word by word English translation and the type of part of speech for respective word.

Image 9: Splitting of words for entered Marathi sentence, its exact English translation and part of speech identification
On clicking the “Transliterate” button, the text in Marathi text gets converted into English.

The libraries used to develop this translation interface are as follows.

- **System**: It contains fundamental classes and base classes that define commonly-used value and reference data types, events and event handlers, interfaces, attributes, and processing exceptions [9].
- **System.data.OleDb**: It is the .NET Framework Data Provider for OLE DB (an API which allows accessing data from a variety of sources in a uniform manner) [9]. The .NET Framework Data Provider for OLE DB describes a collection of classes used to access an OLE DB data source in the managed space [9].
- **System.Web.UI.WebControls**: It contains classes that allow you to create Web server controls on a Web page [9]. Web server controls run on the server and include form controls such as buttons and text boxes [9].

### 3. CONCLUSION

- This interface does Marathi to English Transliteration as well as translation by considering that a sentence should contain four words.
- After splitting of words for the entered Marathi sentence, their exact English translation is carried out. In the next step, parts of speech are identified for respective words and then its exact transliteration and proper translation in English is done.

### 4. FUTURE SCOPE

- This interface can be extended for a sentence that consists of more number of words.
- Punctuation marks and numerical details have to be considered during translation of sentences.
- More number of grammar rules needs to be considered and according to that database design has to be done.
- The reverse feature that is transliteration and translation from English to Marathi can also be implemented.

### 5. ACKNOWLEDGEMENT

We would like to thank Prof. Dr. T. P. Singh, Director, Prof. Dr. Akshay Malhotra, Deputy Director and Prof. Shraddha Phansalkar, HOD Computer Science and Information Technology Department of Symbiosis Institute of Technology, Pune for giving us an opportunity to do research in this area and for their valuable support.

### REFERENCES

2. Marathi language details are obtained from [http://en.wikipedia.org/wiki/Marathi_language](http://en.wikipedia.org/wiki/Marathi_language)