ijRASET

International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Hinglish 2 Hindi Converter

Rubana Khan[1], Rajatkumar Singh[2], Tathagat Khanorkar[3], Vinit Girde[4], Divyesh Ramteke[5], Payal Atey[6]

1Assistant Professor, 2, 3, 4, 5, 6, StudentDepartment of Computer Technology, Priyadarshini College of Engineering, Nagpur, India

Abstract: Hinglish is a language which is a combination of English and Hindi, especially a type of English that involve many Hindi words. The language, which is nearly English with a Hindi-ascent, serves as common language spoken by common man of India. From ancient time, the mixing of Hindi and English words has been occurring. The mother tongue of the middle-class and the so-called up-to-date people through which millions of Indians connect. It mixes Hindi and English words together, it does not follow English sentence format, it creates its very own variant of English called as Hinglish. It was a matter of time before in India Hinglish was born, much like Chinglish in China and Singlish in Singapore.

I. INTRODUCTION

Hinglish is a mixture of the English and Hindi languages that you confront very usually in India. In other sentence, the language resulting from the habit of frequently mixing English and Hinglish vocabulary has become widely known as Hinglish. The people who do not feel confident while talking in English use Hinglish as a second language. English has been very widespread in India ever since the British Raj, of course, and remains a very common language in that area of the world.

Most educated Indians speak English very well, and it has remained the language of the business, education, and technology sectors in that country. Hindi is one of the official languages of the Republic of India, spoken by an approximate 260 million people. It's not shocking that in a place such as India – where English is a crucial skill for success and Hindi is a common local language – that people would speak both languages, but the fascinating thing is that it's now almost universal in the area for people to speak a mixture of both languages relate to as Hinglish.

The language is mostly a subconscious code-switching endeavour, with speakers openly mixing English words and sentences with Hindi (and even some other local languages). Most Hinglish speakers don't even realize that they are swapping back and forth – they are naturally using the best words and phrases from each language to communicate their thoughts. In the modern communication-based society, no natural language suggests to have been left untouched by the craze of code-mixing.

For different communicative objective, a language uses linguistic codes from another language. This allow rise to a mixed language which is neither totally the host language nor the adventitious language. The mixed language poses a contemporary threat to the complication of machine translation.

It is unavoidable to identify the "adventitious" elements in the source language and process them appropriately. The adventitious elements may not appear in their original form and may get morphologically mutated as per the host language. Further, in a convoluted sentence, a section/assertion may be in the host language while another section/assertion may be in the adventitious language. Code-mixing of Hindi and English where Hindi is the host language, is a common phenomenon in day-to- day language usage in Indian metropolis.

The scenario is so regular that people have started seeing this a different variation thoroughly and calling it by the name Hinglish. In this paper, we exhibit a mechanism for machine translation of Hinglish to pure (standard) Hindi and pure English forms. Hinglish does not have any official script like rule of grammar and literature for references. Some people thinks mixing of languages is a bad idea because then the respective languages loss their purity and identity. The mix of Hindi and English is the language of the street and college campus as like "Hungry kya?".

Hinglish language mostly spoken between young generations and the rich elite who want to come across as more "western". Hinglish language is like adding spice and variety toward truly global language.

A. Area Affected By Hinglish

Hindi and English languages both are individually monolinguals but the use of both this language makes them a Hinglish speaker. The user feels it is continent to use Hinglish, a mixture of both English and Hindi language and it also increases their fluency in a language some time. In an English Job market many people have lost their job due to problems in English language fluency but this does not mean by any means the people are entirely shifting towards Hindi language, now a day many individuals are now shifting

their focus towards Hinglish which has gained high value in recent times. It is very useful when the user has to use limited English and have to make a bold impression to get the opportunity. The use of Hinglish can prove very beneficial, especially in a local market for the advertisement purpose.

Over time, our model suggests that all three speech communities will survive, and Hinglish is not likely to replace English or Hindi, as some reports are suggested.

B. *Why hinglish?*

In the present scenario, there is more than 100 crore people are use Hinglish language. The companies are also going towards Hinglish language so they can reach as maximum people as possible. In today's scenario everyone is looking to target each and every person to sell their product and enhance their business. Hinglish is a best way to reach them. Hinglish is a language of common people. We can easily get connected to this people by using Hinglish language.

C. *Objective Of The Project*

Hinglish 2 Hindi converter allows the user to represent the Hindi text using Hinglish words. Hinglish language has really started to influence into various aspects of today's Indian youth in both home and work environments. Nowadays people are not trying to just mingle both English and Hindi words, but they are using it in such a innovatory and dynamic way which can be easily seen in today's advertising era. In this competitive advertising world English alone can't prove beneficial to sell a product in the market.

Hinglish is the language of advertising and it is going to get more and more popular and will determinate the barricade of this industry in our country. Corporate, finance, human resource, operational issues, business cultural widely use this Hinglish language. This language is very useful in helping teachers in corporate language objective into their lesson plan. In India we do have our mother tongue usually regional language official language Hindi and international language English. Communication with Indians have become much easier with the language called Hinglish.

D. *Literature Survey*

Paul Taylor, Alan W Black, Richard Caley they have

described a new way of storing data related to languages in text to speech format. The language entities such as words, phones and diaphones can be stored as featured structures in a general object defined as linguistic item.

These objects can contain any random information in

the feature structure and can be configured at run time. The words with the same pronunciation or the words which have same vocal sound are stored in the database. These kind of language structures are commonly defined with trees or lists. Utterance of any word contain all types of structure of language contained in a single statement [1].

Deepa S.R., Kalika Bali, A.G. Ramakrishnan, Partha Pratim Talukdar has proposed about the problem of breaking Hindi complex words and it affects the quality of phonetize for Hindi Speech Synthesis. As the words in Hindi language are mainly the compound words and they do not have space or hyphen as a connector between them, therefore there

arise a problem of synthesizing the Hindi compound

words with those existing methods of synthesizing. They proposed a new method of taking out a Hindi compound words from the compilation of whole Hindi language [2].

N. Sridhar Krishna, Partha Pratim Talukdar, Kalika Bali, A.G. Rama Krishnan has written about one of the smallest unit of speech which can make one word different from other word (phoneme) for Hindi. there is the Classification and Regression Tree (CART) based database for the identification of words [3]. Lakshmi Sahu and Avinash Dhole described text-to- speech.

(TTS) system for Indian languages (Viz., Hindi,

Telugu, Kannada etc.) to generate human voice or

speech through a speech synthesizer. In this system,

synthesized speech can be generated by joining the

prices of recorded speech or from text generated on

home computer. The quality of speech synthesizer is

approved if the synthesized speech is same as the human voice and is easy and clear to understand [4]. M. Habibullah Pagarkar, Lakshmi Gopala krishnan, Nimish Sheth developed a new method using Devanagari script.

E. Translation Logic

Here we are going to input from the user. User will type the text (Hinglish text) which is written in English. We will provide a text box for writing the hinglish text. To convert the Hinglish text into Hindi character there is a converter is required. For converting the given English character into a Hindi character there is requirement of a converter which convert the given English character into corresponding Hindi character. So, if the user type input as "Hum" then the output should be "हम".

For this we make two separate files, one file will store Hindi character and their equivalent English character. Second file will store the other properties of file for removing the complexity of the program. We have recorded the consonants and vowels of Hindi alphabets probably known as Barahkhadi into the equivalent English character. When user give the input into the textbox our converter directly converts this into corresponding Hindi character without checking any semantic error.

Our logic maps that character into Hindi character. For mapping of those text or character we used type casting of character by using the phenomena of Unicode. We convert this Unicode into a string by using type casting. We make this software into java hence for storing the Hindi character and their equivalent English Character we created different classes. By using this classes, we convert their Unicode into equivalent string value.

So in this way we directly map the hinglish text into Hindi- text or character.

F. Procedure

Here we are going to use a textbox where the user has to type the hinglish text written in English language which will be getting convert into the respective Hindi character. For instance, if the user types 'hum' in our textbox then the output in the other textbox will be showing result "हम". In order to achieve that there is requirement of a converter which will convert the given English character into its equivalent Hindi characters. We are going to use two different separate files. One of them containing Hindi characters along with their equivalent English values. And the Second file will store other properties of a file for removing the complexity of the program.

We have also stored the barahkhadi, the vowels and the consonants of the Hindi language. When the user types any Hinglish word the converters directly convert this into respective Hindi characters without checking any semantic errors. Our files directly map that character into the Hindi character. For mapping we have used a phenomenon called Unicode, through which we can type cast a character.

This Unicode is again converted into a string by using type casting. We have made this Software using java so for storing different English and Hindi characters we have created different classes. Using this classes, we convert the given Unicode, into the string value.

G. Proposed System

To give input through textbox, user will enter Hindi text but written using English letters so Transliteration engine is required so that it can convert English characters to corresponding Hindi characters (not translation). So, if the input is "mera" output should show "मेरा" i.e. it will not check any semantic meaning of "mera". We have used two String tables, one table contains Hindi letters, and other table contains equivalent English text. We have recorded Hindi barahkhadi, and most probable words in Hindi and saved every letter, word separately. English letters are used for naming stored files, so if the user types any Hindi letter or word using English text we can simply concatenate the speech files with the name to get a final output. For example, if user types 'स', its English equivalent is 'S'. So we have a speech file by name 'db.properties' which will be played when 'स' is to be pronounced.

H. Data Flow Diagram
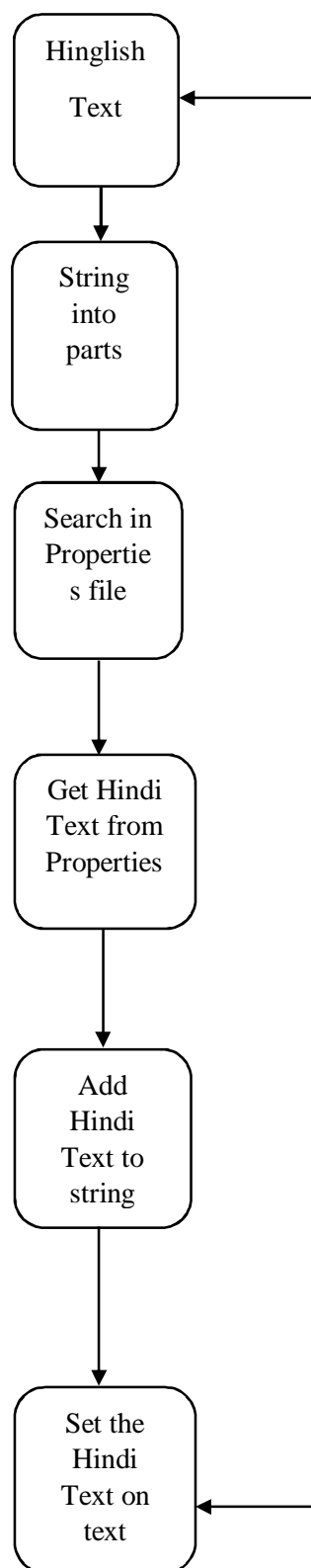
II. CONCLUSION

As it is observed that considerate amount of people is a bilingual speaker which means people who can able to maintain Hindi language and tends to speak mix Hindi and English known as Hinglish. This language has now often regarded as a modern culture language as it connects the English culture with lower urban class culture.

The main aim of this project to explain transliteration (i.e. Hinglish to Hindi character mapping). There can be some future improvements to this project like to implement it to read a Hindi pdf file written in Devanagari script.

REFERENCES

[1] Paultaylor, Alan W Black, Richard Caley, "THE architecture of the festival speech synthesis".

[2] Deepa S.R., KalikaBali , A.G. Ramakrishnan, Partha Pratim Talukdar, A.G. Ramakrishnan "Tools for the   development of a Hindi speech synthesis system",5th ISCA Speech Synthesis Workshop – Pittsburg.

[3] N. Sridhar Krishna, Partha Pratim Talukdar, Kalika Bali, A.G. Ramakrishnan , "Duration Modeling for Hindi Text-to-Speech Synthesis System".

[4] Lakshmi Sahu and Avinash Dhole, "Hindi & Telugu Text-to-Speech Synthesis (TTS) and inter-language text Conversion",International Journal of Scientific and Research Publications, Volume 2, Issue 4, April 2012, ISSN 2250-3153

[5] M. Habibullah Pagarkar, Lakshmi Gopalakrishnan, Nimish Sheth, Rizwana Shaikh, Virag Shah "Language Independent Speech Compression using Devanagari Phonetics" , cs.jhu.edu. 200http://developer.android.com/distribute/googleplay/publish/ preparing.html. (2013) .

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)