Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 01, January -2018

Hybrid machine translation from Marathi to English: A research evolution in machine translation

Prof. AbhaPathak[1], Anchal Kumari[2], Akanksha Prasad[3], Ashwini Topre[4], RutujaLondhe[5]

Department of Computer Technology, Pune.

Abstract - *India is multilingual country, in India people speak different dialect. Hence data present in different dialect and structure gives rise to dialect as a barrier in information retrieval. In state like Maharashtra where Marathi is the local dialect commonly used by the people, it gives rise to a non-adjustable environment for non-marathi users to live in. Vital information on current happenings on village and taluka level published in newspaper, destination and their distances indicated by the sign boards, advertisements and shops/hotels names/other vital information presented with local language causes a major setback among Masses. Information can be present in form of text, speech and image translating this information helps for sharing of information and ultimately information gain.The proposed system uses Image Processing and Machine Aided Translation for overcoming the above problem. Image processing is used because we are taking images as input and extracting useful information from it which has to be translated. Machine aided Translation helps to translate Information displayed in one dialect to other language(in our case from marathi to english). We are presenting the output not only in text format but also in speech. Currently System has been proposed for images which can be extended to speech and voice in future. Also many other dialects can be incorpated for the betterment of the user.*

Keywords- *Image Processing, Unicode, Machine aided Translation, Content based image retrieval, duplicate image detection, SQL, TTS Algorithm.*

I. INTRODUCTION

Machine enabled transformation is core research in Natural Language (NL) for removing language as obstacle in communication and information access with help of bi-lingual machine translation. Research work in Machine translation has been done from English to Hindi, English to Urdu to another language like telugu many native languages and foreign languages like Arabic, Chinese and Spanish . The research problem to address is to community of Marathi language. Lets take an instance of travelling. Suppose the user has to travel from Delhi(source) to Nashik(destination). While travelling there are several sign boards, name plates which are written in marathi language and the user being a non-marathi person doesn't understand anything. So our focus is to develop asystem in which when he/she focuses the camera of his/her mobile phone to the text it gets processedand the output is given in english language in two forms- speech and text.

II. PROBLEM STATEMENT

Traditional way of a Machine Translation is translation obtained by machine on large scale from source to target language. Marathi to English language Translator is of computer science and linguistics concerned with the interactions between computers and human (natural) language.

III. LITERATURE REVIEW

1. Marathi Text to Speech Synthesis Using Matlab: The system for text to speech mixture for Marathi language gives whole idea that how to convert Marathi text into language right from text dispensation to audio processing. The proposed system requires audio data base and very limited text data base. The TTS system has been developed on is Unicode software therefore UTF-8 encoding has used to read the Marathi text.

2. Issues in Parsing for Machine Aided Translation from English to Hindi:Resolving cases of uncertainty due to differences in the syntactic form of grammatical rules in basis and objective language is a main confront in the expansion of a Machine Aided Translation (MAT) system.

3. An Evaluation Methodology for English to Sinhala Machine Translation: This document present estimate method for English to Sinhala machine Translation system.

4. A Parser for Sinhala Language First Step Towards English to Sinhala MachineConversion: Sinhala language parser has been industrial as the first step towards English to Sinhala natural language translation.

5. A Robust Segmentation Technique for Line, Word and Character Extraction from Kannada Text in Low Resolution Display Board Images: consistent taking out of text appearance, words and font is one of the very important steps for growth of automated systems for considerate the text in low resolution display board images.

6. Text Region Extraction from Low Resolution Natural Scene Images using Texture Features: Automated systems for understanding display boards are finding many applications useful in guiding tourists, assisting visually challenged and also in providing location aware.

IV. MATHEMATICAL MODEL:

S = {I,P,O,S,F}
Where,

1. Input:

- IP = {I}.
- Where,
- I is set of images, provided as an input.

2. Procedure:

- Step 1: Initially capture the image using your mobile phone's camera.
- Step 3: Extraction and pre-processing of useful content of the image.
- Step 4: Perform comparison using unicode.
- Step 5: Using TTS Algorithm map the result with associated speech.
- Step 6 : Display the translated image and the speech.

3.Output: O: is the output presented in text and speech.
4.Success: Translation done successfully. Appropriate information represented.
5.Failure: Image not properly captured, Camera not responding, text not in preferred language.

V. PROPOSED SYSTEM

The proposed system uses Image Processing and Machine Aided Translation for overcoming the above problem. Image processing is used because we are taking images as input and extracting useful information from it which has to be translated. Machine aided Translation helps to translate Information displayed in one dialect to other language (in our case from Marathi to English). We are presenting the output not only in text format but also in speech.

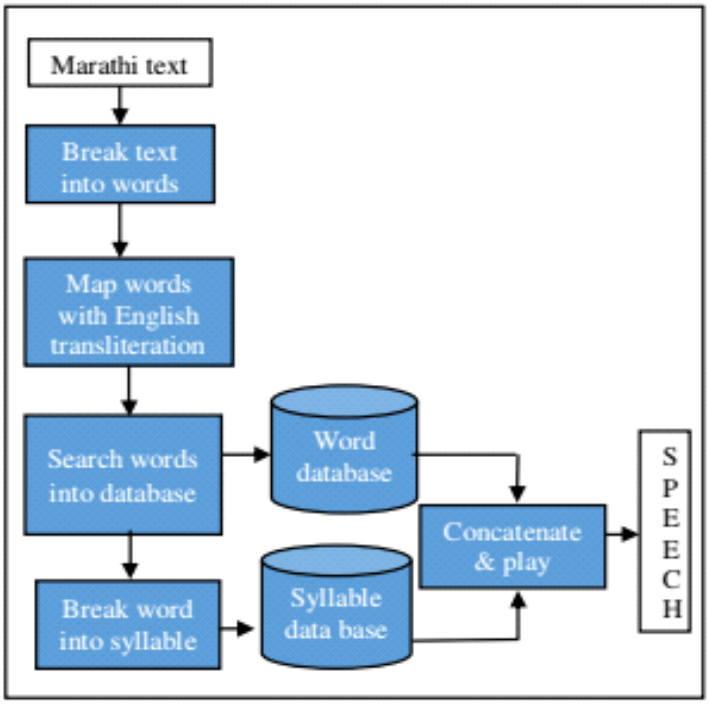

Fig 1: Block Diagram for proposed system

VI. KEYWORDS

Image Processing: Image processing is a method to execute some operation on an image, in order to get an improved picture or to extract some useful in sequence from it. It is a type of sign giving out in which contribution is an image and output may be image or associated with that image. Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too.Image processing basically includes the following three steps:

- Importing the image via image acquisition tools;
- Analysing and manipulating the image;
- Output in which result can be changed image or report that is base on image analysis.

Unicode: **Unicode** is a computing manufacturing standard for the consistent encoding, representation, and handling of text spoken in most of the world's writing systems.
"Content-based" means that the explore analyze the inside of the picture slightly than the metadata such as keywords, tags, or similes associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is attractive because searches that rely purely on metadata are dependent on annotation quality and completeness. Having human physically interpret images by entering keywords or metadata in a large database can be time overriding and may not capture the keywords preferred to describe the image. The estimate of the efficiency of keyword image search is subjective and has not been well-defined. In the similar view, CBIR systems have similar challenges in defining success.

SQL: Structured Query Language (SQL) is a typical processor words for relational file organization and data manipulation. SQL is used to query, insert, update and modify data. Most relational database support SQL, which is an additional benefit for database administrator (DBAs), as they are frequently necessary to support database across several dissimilar platforms.

TTS: Text-to-speech (TTS) is a type of speech synthesis request that is used to make a verbal sound description of the text in a processor document, such as a help file or a Web page. TTS can enable the reading of computer display information for the visually challenged person, or may basically be used to supplement the interpretation of a text message. Current TTS applications include voice-enabled e-mail and spoken prompt in voice reply systems. TTS is frequently used with voice recognition programs. There are many TTS crop available, including Read Please 2000, Proverbe Speech Unit, and Next up Technology's TextAloud. Lucent, Elan, and AT&T each have products called "Text-to-Speech."

VII. METHODOLOGY

The Marathi script uses Devanagari script. This script contains a set of 12 vowels and 36 consonant, which are known as **swar**and **vyanjan** respectively in the language. It also contains dependent vowels, which are known asMatra. All the vowels, consonants and dependent vowels have been stored in the database in the form of ASCII values with their English transliteration code. This is because Matlab being Unicode software, it first converts Marathi text to its equivalent English translation.The Marathi script uses Devanagari script. This script contains a set of 12 vowels and 36 consonant, which are known as **swar**and **vyanjan** respectively in the language. It also contains dependent vowels, which are known asMatra. All the vowels, consonants and dependent vowels have been stored in the database in the form of ASCII values with their English transliteration code. This is because Matlab being Unicode software, it first converts Marathi text to its equivalent English translation.

VIII. CONCLUSION

This paper focuses on More than two sets are of data Translations segments taken from site and submitted to our system. The results show that at times the output produced by other Translator is fluent but contains meaningless sentences and miss out the core sense of passage. The proposed system produces the rule based out with accurate sense and meaning conservation Statistical output is generated from corpus (collection of written text) incorporated in project the statistical output is also better and comparable. The Hybrid output is been found better in two inputs given to system. Ultimately Hybrid Approach is better to single Approach taken.

REFERENCES

[1] DarshnaBadhe, P.M.Ghate, Marathi Text to Speech synthesis: using Matlab:International Journal of Computer Science and Network, Volume 4, Issue 4, August 2015.

[2] PramodSalunkhe ,MrunalBewoor , SuhasPati , Shashank Joshi ,Aniketkadam,Summarization and Hybrid Machine Translation System for English to Marathi: AResearch Effort in InformationRetriveal System (H-Machine Translation ,DiscoveryThe International journal (ISI thomasretuers indexed).2015.

[3] PramodSalunkhe ,MrunalBewoor , Dr.SuhasPatil A ResearchWork on Englishto Marathi Hybrid Translation System, (IJCSIT) International Journal of ComputerScience and Information Technologies, Vol. 6 (3) 2015, 2557-2560.

[4] S. B. Kulkarni , P. D. Deshmukh Linguistic Divergence Patterns in English toMarathi Translation International Journal of Computer Applications (09 75 8887)Volume 87 No.4, February 2014. J. Ramanand, AkshayUkey, BrahmKiran Singh.

[5] Pushapak Bhattacharyya Mapping and Structural Analysis of MultilingualWordnets Bulletin of the IEEE Computer Society Technical Committee on Data Engineering.

[6] Sreelekha.S, Raj Dabre, Pushpak Bhattacharyya Comparison of SMT and RBMT,The Requirement of Hybridization for Marathi Hindi MT [online pdf].

[7]Tapas Kumar Patra, "Text to Speech Conversionwith Phonetic Concatenation", International Journal of Electronics Communication and Computer Technology (IJECCT) Volume 2 Issue 5 (September 2012).

[8]Mrs.Madhavi R. Repe," Prosody Model for MarathiLanguage TTS Synthesis with Unit Search and Selection Speech Database", IEEE International Conference on Recent Trends in information, Telecommunication and Computing 2010.

[9]Shruti Gupta, "Hindi Text To Speech System",ComputerScience And Engineering Department Thapar University Patiala June 2012.

[10]Snehali. K. Nandurkar, ZakirM.Shaikh, "SpeechGeneration Of Transliterated Hindi Text",International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 10, October 2014 ISSN 2319 - 4847

[11] Shreekanth.T, " An Unit Selection based Hindi TextTo Speech Synthesis System Using Syllable as a Basic Unit", IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 4, Issue 4, Ver. II(Jul-Aug. 2014).

[12] MrsMinaksheepatil, " Syllable" Concatenation forText to Speech Synthesis for Devnagari Script", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 9, September 2012.

