iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** XII **Month of publication:** December 2019

DOI: http://doi.org/10.22214/ijraset.2019.12037

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

Hindi to English Translation using NLTK

Samruddhi Waghmare[1], Ankita Sonanwane[2], Mamta Wagh[3]
[1, 2, 3]*Computer Engineering, JSMP'S Jayawantrao Sawant College of Engineering, Pune, Maharashtra.*

Abstract: ***Multilingual Machine Translation is a computerized system that is designed to translate source text from various natural languages into target text of another natural languages. Due to increased need of global communication, multilingual machine translation is the propel for researchers. With the advancement in technology, now we have computerized systems that can replace the human experts in particular domains. One such popularly adopted domain by researchers is artificial intelligence (AI) and NLP .Natural language processing(NLP) is a field of computer science concerned with the interaction between computer and human (natural) languages. It is becoming one of the most active area in the interaction between human and computer. These includes text system that integrate text and natural language. Machine translation (MT) system is available mostly for a specific language pair. Recently many MT systems are being designed for bilingual pairs. And the research is heading towards developing MT systems that are capable of handling more than two language pairs i.e. multilingual. MT systems still has so many issues to deal with. Recently developed multilingual systems are laden with the enormous size corpus .In this proposed system we are analyzing machine translation system using NLTK.***

Keywords: Meachine translation, NLP(Natural Language Processing) ,NLTK(Natural Language Tool Kit), AI(Aritifical Intelligence),Text to Speech.

I. INTRODUCTION

The main purpose of Natural Language Processing is for an English sentence to be interpreted by the computer and appropriate action taken. Despite the challenges, natural language processing is widely regarded as a promising and critically important set out in the field of computer research. The applications that will be possible when NLP capabilities are fully realized are impressive computers would be able to process natural language, translating languages accurately and in real time, or extracting and summarizing information from a variety of data sources, depending on the users' requests. This system focuses on the solution of the problems arising in the analysis or generation of Natural language text, such as syntactic and semantic analysis or compilation of dictionaries and grammars necessary for such analysis. It proposes the architecture for translating Natural language Hindi to English.

II. LITERATURE SURVEY

According to the traditional Systems and prior research, the methodologies and technologies used for translation are giving below in the table.

Sr.no	Paper name	published year	Technique used
1.	A Hybrid Approach For Hindi-English Machine Translation	2018	It has combined approach of rule based and Statistical approach.
2.	Hindi to english translation using concept net	2017	In this it uses the Knowledge based approch.
3.	Hindi to english machine translation	2016	In this only Statistical approach.

III. TECHNOLOGY USED

1) *Python:* Python is an interpreted , high-level, general-purpose programming language.
2) *NLTK:* NLTK3 library for python will be used for input stemming. This library serves as a toolkit for computational linguistics. Following is a list of the modules we will be using.

Token module provides basic classes for processing individual elements of text, such as words, or sentences. NLTK tokenize is used to tokenize incoming sentences. The words net lemmatize and porter stemmer is used parse the input's base words for comparison.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

IV. OBJECTIVE

The objective of our project is to translate into accurate and english sentence after parsing natural language using open source tools i.e NLTK and libraries. Users will be able to obtain english statement when the given input is hindi.

A. Flowchart

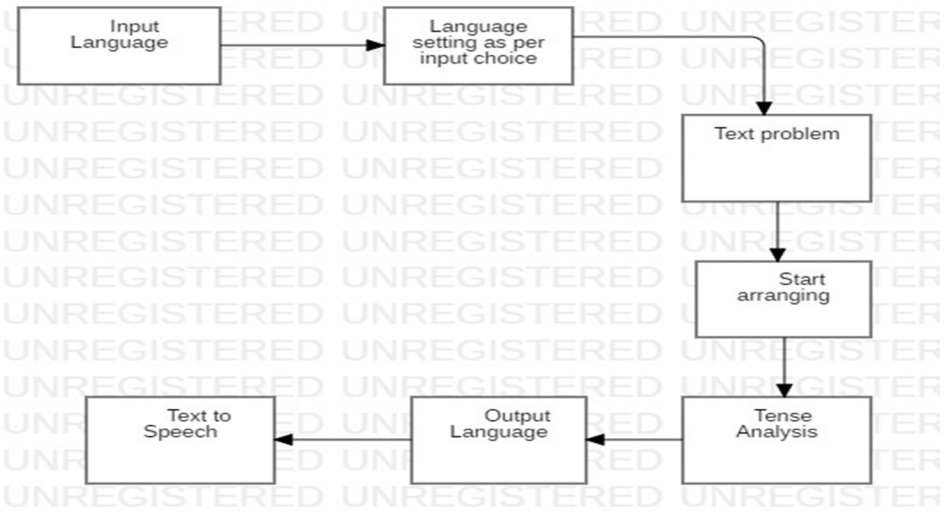


Figure 1: Flowchart of hindi to English translation

B. System Architecture

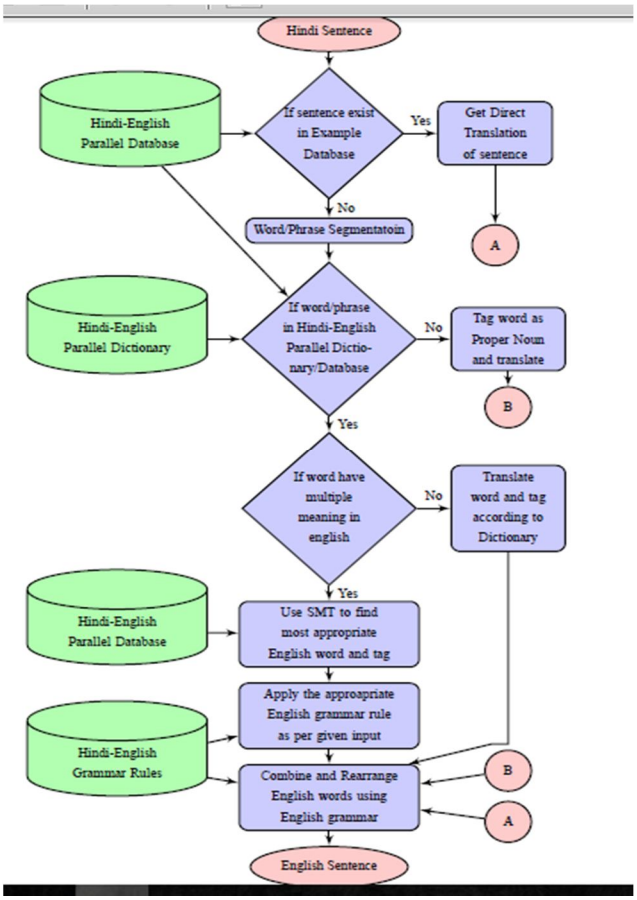


Figure 1: ARCHITECTURE

©IJRASET: All Rights are Reserved

C. Working Process

1) The first step in translation is to split the sentence into words or simple sentences.
2) If a part of the input is in the example database then we keep that part as it is and the remaining part is segmented into words.
3) When the segmentation is done, segments are actually translated and tagged independently.
4) If the segment is example based, it is directly converted to English.
5) Further, the remaining words are translated using the parallel Hindi-English Dictionary.

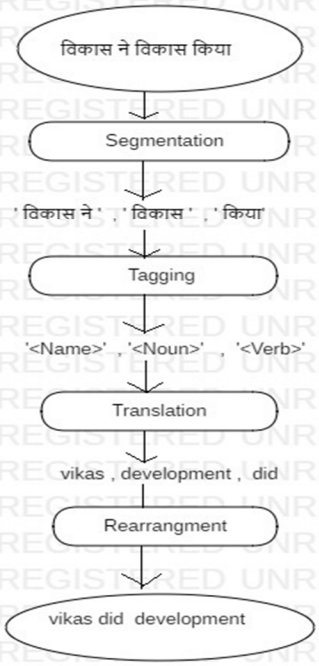


Fig1. Step-wise result without any drawback

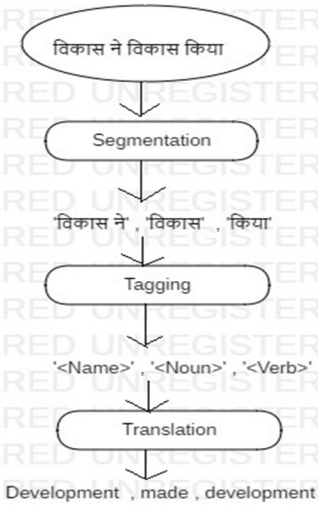


Fig2:Step-wise result with drawback

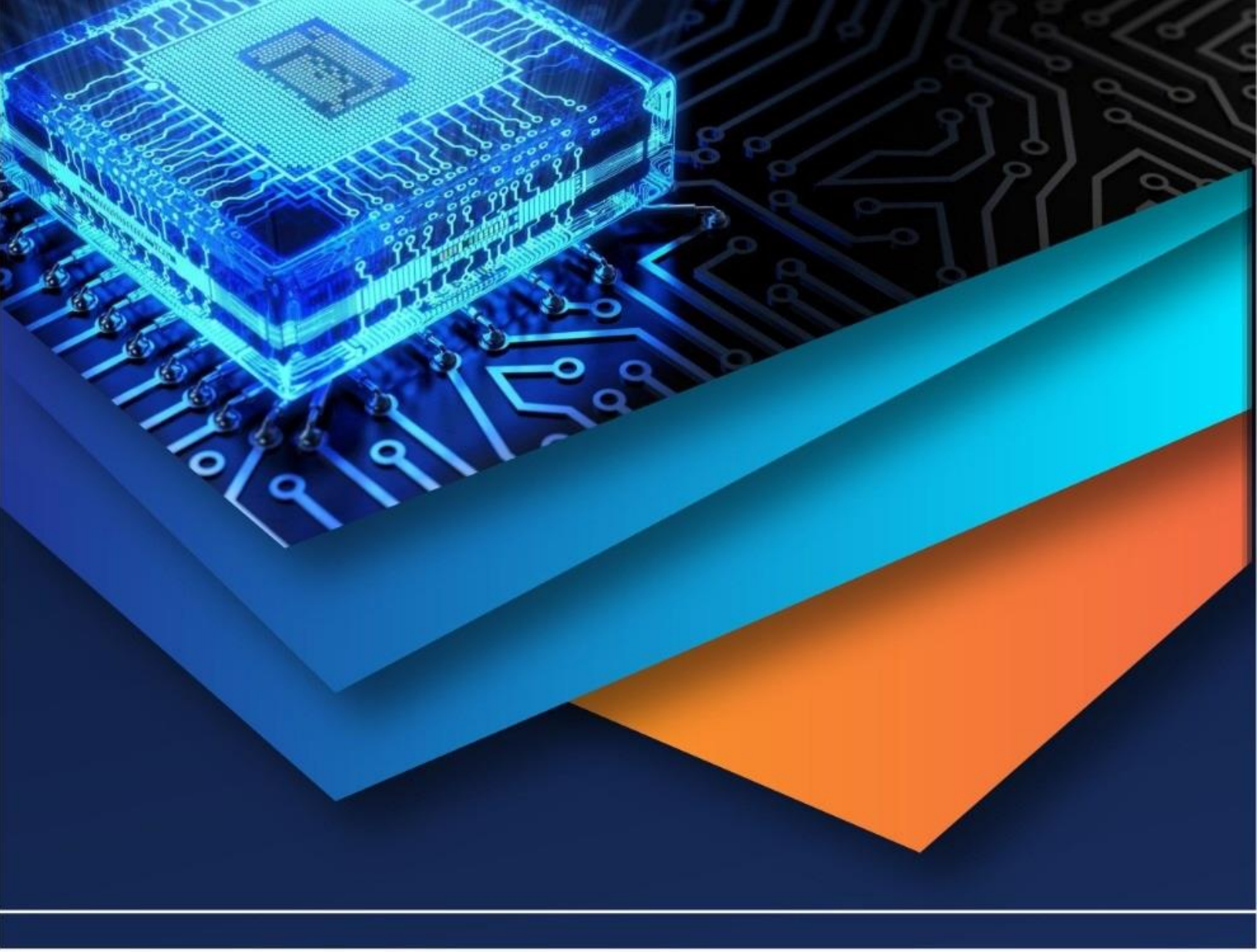
V. CONCLUSION

Hybrid model of translation (HMT) system is able to outperform many baseline translation systems based on direct, RBMT and SMT approaches individually.Currently it is not giving the good result on complex and multiple combine sentences.By using this method the user can get the output.

REFERENCES

[1] Alpaydin. Introduction to Machine Learning. Adaptive Computation and Machine Learning Series, MIT Press. 2014.
[2] Lotter, Hempel. Lernen, Lernschwierigkeiten – Diagnostik der Lernvoraussetzungen. Regierung Oberbayern. 2008.
[3] S. Chetlur. cuDNN: Efficient Primitives for Deep Learning. arXiv preprint arXiv:1410.0759c2. 2014
[4] T. Brants et. al.. Large Language Models in Machine Translation. EMNLP-CoNLL. 2007.
[5] W. Ding, et. al.. "Theano-based Large-Scale Visual Recognition with Multiple GPUs" ICLR. 2015.
[6] P. Flach. Machine Learning: The Art and Science of Algorithms that Make Senso of Data. Cambridge University Press. 2012.
[7] H. Fröning. GPU Computing Slides University of Heidelberg, ZITI. 2014.
[8] Y. Jia et. al.. Caffe: Convolutional Architecture for Fast Feature Embedding. arXiv preprint arXiv:1408.5093v1, 2014.
[9] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks. NIPS. 2012. [10] A. Krizhevsky. One weird trick for parallelizing convolutional neural networks. arXiv:1404.5997 [cs.NE]. 2014.
[10] Y. LeCun, et. al.. Backpropagation applied to handwritten zip code recognition AT&T Bell Laboratories. 1989.
[11] Y. LeCun, et. al.. Efficient BackProp Neural Networks: Tricks of the trade, Springer. 1998.
[12] K. P. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press. 2012.
[13] M. A. Nielsen, Neural Networks and Deep Learning. Determination Press. 2015.
[14] NVidia. User Guide – cuDNN Library NVidia. DU-06702- 001_v6.5. 2014.
[15] T. Paine et al.. Gpu asynchronous stochastic gradient descent to speed up neural network training. arXiv preprint arXiv:1312.6186. 2013.
[16] O. Russakovsky el. al.. ImageNet Large Scale Visual Recognition Challenge. arXiv preprint arXiv:1409.0575. 2014.
[17] S. Shalev-Shwartz, S. Ben-David. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press. 2014.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)