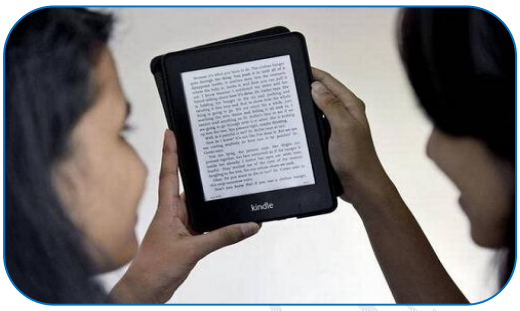
REVIEW OF RESEARCH
IMPACT FACTOR : 5.7631(UIF) UGC APPROVED JOURNAL NO. 48514 ISSN: 2249-894X
VOLUME - 8 | ISSUE - 7 | APRIL - 2019

A SYSTEMATIC LITERATURE REVIEW IN HINDI LANGUAGE



[1] Swapnil Justin and [2]Shabdika Pandey
[1] Asst. Prof. DCSA, St. AloysiusCollege, Jabalpur, MP.
[2] Student, DCSA, St. Aloysius College, Jabalpur, MP.

ABSTRACT :

In India we have total 22 officially declared languages in which Hindi is the language that is used by the most of the population. Under the language/text processing majority of work is done but we still lack in developing tools and application for our own language therefore there is a keen need to gather and integrate all the efforts putted and work done in Hindi text to provide the compact understanding of it for researchers.

This paper draws an attention of researcher to report systematic review focused to identify the challenges or troubles faced by universal teams during the translation or normalization of Hindi text .We have also recommended best tools that will be beneficial to understand the details of this language.

KEYWORDS : *"Hindi Text', 'Hindi Contents', 'Classification In Hindi', 'Hindi Text Language', 'Hindi Language Text Summarization', 'Hindi Text Normalization', Documents In Hindi', 'Hindi Text Description'.*

OBJECTIVES OF STUDY:

- To present the overview of works done in the field of Indian Language.
- To present major translation system present.
- To identify major issues regarding Hindi translation.

INTRODUCTION:

Systematic literature review is a kind of review where use some ways that should be organized in some manner, to gather data from various resources and produce new findings qualitatively or quantitatively.it gives the overview of the work done in that field through the resources we gone through.

According to Indian constitution Hindi id our official language still it is the language that is being used by the majority of population of India. Hindi is quite a rich language in every perspective and it been the Foundation stone for many other languages too.So for this much popular language need the technical transformation to cope up with the need and current trends rising ahead like translation, search, e-resources or processing for other purposes [1].

Language translation using machine is falls under the sub category of computational linguisticsin which we use software to translate text from one language to another language. As we already know that the modernization and digitalization of society screams for the various kind of needs to be fulfilledi.e. related to Hindi language to match with the technological advancements available now a days whether it is translation or processing for other fruitful purposes [2].

The study aims to present literature review on the development ofHindi language's computing in any form. Unlike English,her is not much to see for Hindi language so by this study we try to highlight the technologies available for Hindi language computing[2].

Here we took review over some papers i.e. "Anuvadak : Automatic Summarization of Text Documents Written in Hindi Language" and "Hindi language text search:- A literature review" ; where in Anuvadak , hybrid approach by combining Rule based and Example based translation methods been proposed

On the other hand "Hindi language text search: A literature review", It presents the various search engines available and also talk about issue faced when combinational word/sentence is searched."Automatic text summarization using supervised machine learning technique for Hindi language", Proposed an automatic text summarizer to summarize large text by extracting sentences using machine learning techniques. Now a days Neural Networks is being used for machine learning for Machine Translation even though many challenges are involved in Neural Machine Learning[1][5].

BACKGROUND AND RELATED WORK:

The very first work in automated text summarization was done by Luhn in 1958,Later Julian Kupiecin 1995 proposed some new features and in 2009 proposed a system based on fuzzy logic [1].There is a lot of work done in Hindi textcomputation and a lot more to be done ,According to the papers we are following, all the related work done in these three papers are as follows :

In "Anuvadak'', this paper is used to translate English language to Indian language by using natural language processing, here it propose a new approach called as hybrid approach by combining Rule based and Example based translation method .The implementation of the "Anuvadak" was done by using JAVA as front end and MySQL as a back end to store the examples and rules of parsing method. To overcome and solve some of the issues like gender and tense recognition in English to Hindi translation "Anuvadak" design is proposed as follows[2]:

Thetranslator "Anuvadak" accepts an input from user in English text form and thenperform following steps:

Step 1: Sentence identification

Step 2: Identify the Gender, Phrases, Prepositions and Auxiliaries

Step 3: Tens identification

Step 4: Translated text in Hindi

This paper presents an approach to the design anautomatic text summarizer for Hindi text that generates a summary by taking outsentences. The ain motive is to identify the most important sentences from the document and then compile it to form a summarized text output for which it uses the following methodology:

Input: A text file Output: A summarized textof original, as per compression ratio. 1. Read Input text File -Og 2. Pre-process the file Og .

//Preprocessing step 2.1 Segment text file into sentences. 2.2 Tokenize each sentence into words. 2.3 Remove stop-words [2].

Problem faced during development were:

- Gender identification
- Multi meaning words
- Subject implication

In paper "Hindi language text search : A literature review" represents the major problem of Hindi searching over the web, this review reveals the ability of a number of techniques and searched engines that have been developed to facilitate Hindi text. This paper also provides the list of various language translation systems.

Another task that is beyond translation is summarization of document from one language to another."Automatic text summarization using supervised machine learning technique for Hindi language". Some most common techniques of Machine Translation [5]:

1) Statistical Machine Translation (SMT)
2) Rule Based Machine Translation (RBMT)
3) Example Based Machine Translation (EBMT)
4) Hybrid Machine Translation

In the paper they proposed a technique and test for summarizing text document which is divided into 3 important parts : Pre-processing, Processing and Extraction.

1. Pre-Processing In pre-processing stage of this proposed technique, the text is first split into sentences, then sentences are further broken into words and then stop words are removed. Preprocessing stage involves 3 steps 1) Segmentation 2) Tokenization 3) Stop words removal.

a. Segmentation In this sentences are segmented based on sentence boundary i.e. specified by "|".
b. In Tokenization sentences are broken down into words by identifying the space and comma between the words.
c. Remove Stop-Words.

2. Processing stage here value of feature for every sentence is calculated. Our proposed technique makes use of six statistical features for calculating sentence score and each of them are explain below:

2.1 Sentence Paragraph Position (f1) : Higher values are assigned to the starting sentences and lower values are assigned to ending sentences of the paragraph.
2.2 Sentence Overall Position (f2) Sentence Overall Position values are calculated in context of the entire text in the document. Higher values are assigned to the starting sentences and lower values are assigned to ending sentences of the document
2.3 Numerical Data in Sentence (f3) : calculate numeric data value.
2.4 Presence of Inverted Commas (f4) :Presence of inverted commas is calculated.
2.5 Sentence Length (f5) : shorter and longer sentences are assigned lower values because it may carry the brief. Sentence length values are calculated using equations
2.6Keywords in Sentence (f6) Keywords are words that appear with high frequency in a text document because it carries essence of the sentence.

3. Extraction The sentences are now extracted and included in the final summary file based on the total lines possible in depending on the compression ratio intended [3].

CONCLUSION:

This paper discusses the development and issues in Hindi language computation and the techniques to solve the translation problem in Hindi text. It is observed that a number of problems still exist in the area of translation involving Hindi language. In future various aspects of language will be considered for translation like cue word, sense of humor and context information etc. and as it comes to summarization still it is not overall perfect because it only selects the word or sentence on the basis of specified weightage tagged to it instead of sensing the meaning of the sentence. We have to find a way to overcome this and make a smarter model that will be able to understand the sense behind the sentence to produce more sensible and meaningful summary.

REFERENCES:

1. Singh P, Tripathi A, "Hindi Language Text Search: a literature review", annals of library and information studies, vol. 64, 2017.
2. BagulSudhirD , Joshi U B , "anuvadak : english to hindi text translator", International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR), 2015.
3. Desai Nikita, Shah Prachi, "Automatic text summarization using supervised machine learning technique for Hindi language", IJRET: International Journal of Research in Engineering and Technology, 2016.

4. KhanShahnawaz and UsmanImran, "A Model for English to Urdu and Hindi Machine Translation System using Translation Rules and Artificial Neural Network", The International Arab Journal of Information Technology, Vol. 16, No. 1, January 2019.
5. Goswami Pankaj K., AsthanaShubham, "Unifying Indian Languages through Neural Machine Translation", Knowledge Digest for IT Community, CSI Communications, Volume 42, Issue No. 12, 2019.
6. KarthikeyanU., Dr. VanithaM., "A Study on Text Recognition using Image Processing with Datamining Techniques", International Journal of Computer Sciences and Engineering, E-ISSN: 2347-2693, Vol. 7, Issue-2, 2019 .