**SYNOPSIS REPORT**

**Text Analysis Using Hierarchical Approach**

**Submitted by**

Yash Jangid (Enroll No. R970218068)

Rishabh Mishra (Enroll. No. R970218044)

Abhinav Kamboj(Enroll. No. R970218067)

Arjun Gupta(Enroll No. R970218010)

BRANCH :- Oil & Gas Informatics

SEMESTER :- V

**Under the guidance of**

Mr. Rahul Kumar Singh

Assistant Professor

Department of Informatics



**SCHOOL OF COMPUTER SCIENCE**

UNIVERSITY OF PETROLEUM & ENERGY STUDIES

Bidholi Campus, Energy Acres, Dehradun – 248007.

**June Dec-2020**

 **School of Computer Science**

**University of Petroleum & Energy Studies, Dehradun**

**Project Proposal Approval Form (2020)**

**Minor-1**

**PROJECT TITLE:**

Text Analysis Using Hierarchical Approach ( using parse tree)

**ABSTRACT-**

In this minor project we have to present a Machine Translation System for getting name entity of different language using rule – based machine translation technique & Hierarchical approach. This project discusses the language diversion with a recommended solutions to handle the diversion , And the system consists of different bilingual dictionaries. The system uses context free grammar in Chomsky normal for language processing and parsing techniques for processing the input sentence.

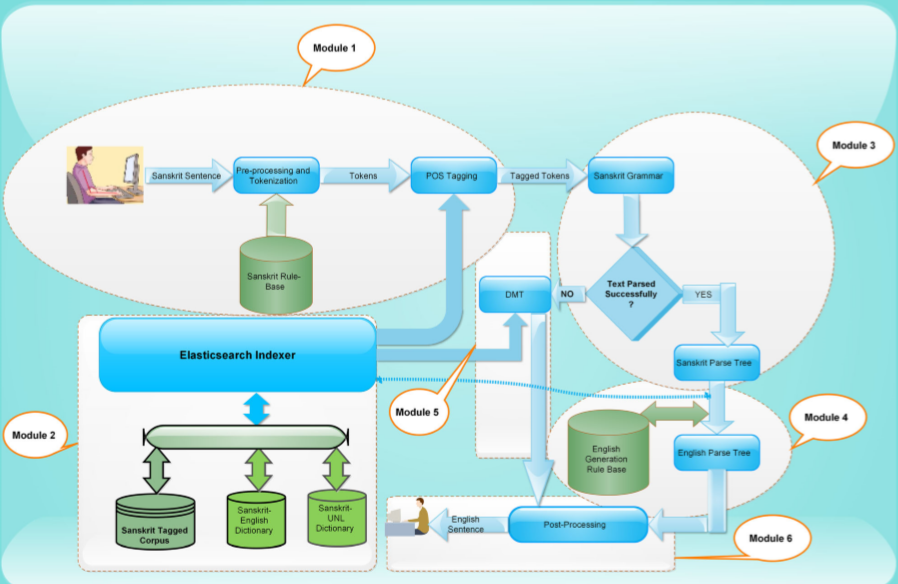
**KEYWORDS:**

Machine translation , Different language grammar , parser , Bilingual evaluation understudy, part of speech tagging.

**INTRODUCTION**

The most utilized language in the World is "English". So it's been a provoking undertaking to change over the any language into most helpful language like English. But Machine translation which is known for translating text from one language to another through computers , so, we can easily and efficiently convert the any text into English text. We can also do the same with Human translations but it is not a core solution to this problem. This system is very helpful in teaching and learning any language to understand its features like, unambiguous, well- structured grammar etc. And also helpful for the research purposes.

Whereas Machine translation system is the most beneficial way for this purpose because it includes high translation speed , cheap / less cost, easy to translate into multiple language at once in multilingual environment, store more data in comparison of human brain and most useful benefit is its availability , we can use this system at anytime anywhere.



**Fig 1.1 :** Architecture to purpose Sanskrit to English Machine translation system

**OBJECTIVES-**

* Preprocessing and tokenization
* Part of speech tagging for feature extraction
* Hierarchical representation of text

**METHODOLOGY**

In order to complete the project , we have to divide the project into different segments . This project is divided into six segments .

1. Gives the introduction about language and Machine translation system.
2. Defines the various divergence of languages occurs during translation of language with recommendation to handle them.
3. Then we have to define the various six modules used to proposed Machine translation system.
   * + Source language preprocessing
     + POS (part of speech) tagger
     + Language grammar
     + Parsing table
     + Language tree generation
     + Parse tree generation for parsing table
4. Fourth segment gives the details of data dictionary , various rules base, tagged corpus and technology used for implementing the purposed system.
   * + Experimental setup
5. Defines the evaluation methods used for the evaluation of the purposed system with the comparison of existing system.
6. Final conclusion followed by the References (Which is shown at the end of the Synopsis)

**SYSTEM REQUIREMENTS**

1. Hardware:

* RAM: 4GB

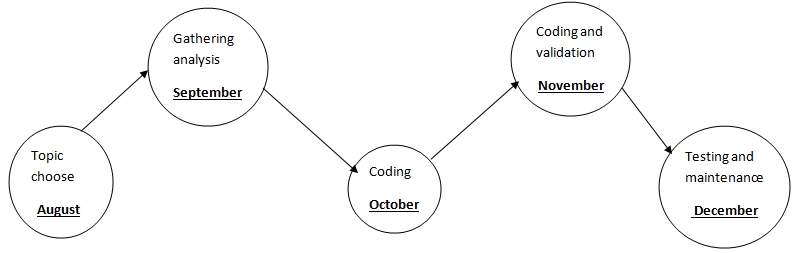
2.)Software:

* Any c-compiler

3.)Operating System:

* Windows.

**SCHEDULE**-

****

**Pert Chart**

**REFERENCES**

1. <https://link.springer.com/chapter/10.1007%2F978-3-540-30211-7_27>
2. <http://sanskrit.jnu.ac.in/corpora/JNU-Sanskrit-Tagset.htm>
3. <https://www.tandfonline.com/doi/full/10.1080/03772063.2018.1528187>
4. <http://sanskrit.jnu.ac.in/corpora/tagset.jsp>
5. https://link.springer.com/chapter/10.1007%2F978-81-322-1143-3\_4

**Synopsis Draft verified by**

**Project Guide HOD**

**(Name & Sign) (Dept. of Systemics)**

**INSTRUCTIONS**

1. Min/Max. no. pages: **7** pages excluding Front Page.
2. The contents in the document should be provided in such a way that the **Verification of the Synopsis Report** should be available in the last page along with the contents and NOT in a separate page.
3. Type your content in this document itself with all Pages except the front page numbered. Bottom, Center, Times New Roman, Bold, Font size 10.
4. Margins: Left(1”), Right (0.75), Top & Bottom(0.6”).
5. Chapter Heading: Center Aligned, Bold, All Caps, Time New Roman, Font size 16
6. Main Heading: Heading under a chapter should be Left Aligned, Bold, Not underlined & not italicized, Times New Roman, Font size 14
7. Sub-Heading: Left Aligned, Bold, Not underlined & not italicized, Times New Roman, Font size 12
8. Content Format: Justified, Not Bold & Underlined, Not Italicized, Time New Roman, Font size: 12, line spacing: 1.5