**A**

**SYNOPSIS REPORT**

**ON**

**Machine Transliteration: A Hybrid approach for natural language**

**Submitted by**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Name of student** | **Roll No.** |
| **1** | **Ms.Isha Godiyal** | **4222** |
| **2** | **Mr.Sumeet Lohare** | **4260** |
| **3** | **Mr.Pradip Khutwad** | **4265** |
| **4** | **Mr.Suprit Kulkarni** | **4266** |

**Under the Guidance of**

**Mr.S.B.Javheri**

****

**JSPM’s**

**Rajarshi Shahu College of Engineering**

**Department of Computer Engineering**

**[2012-2013]**

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Chapter** | **Topic** | **Page No.** |
|  | **Abstract** | **1** |
| **I** | **Introduction** | **2** |
| **II** | **Problem Statement** | **4** |
| **III** | **Objective ,Aim, Scope** | **5** |
| **IV** | **Feasibility study** | **7** |
| **V** | **Technical Details** | **8** |
| **VI** | **Technologies used** | **10** |
| **VII** | **Innovativeness and Usefulness** | **11** |
| **VIII** | **Project overview** | **12** |
| **IX** | **Papers Referred** | **14** |
| **X** | **References** | **15** |

**Abstract**

Machine translation is one of the most important applications of Natural Language Processing (NLP). Machine Translation is one of the most important branches of Artificial Intelligence. It is the translation of one natural language into another using automated and computerized means. For a multilingual country like India, with the huge amount of information exchanged between various regions and in different languages in digitized format, it has become necessary to find an automated process from one language to another. Sanskrit is the mother of all native language of India. A great storage of knowledge with various subjects is kept alive and fresh Sanskrit lore for thousands of years. The state of Uttarakhand in India has ruled Sanskrit as its second official language. NASA and others have been looking at Sanskrit as a possible computer language since its syntax is perfect and leaves little room for error. English to Sanskrit translator is very useful to people in India. Sentence in English is translated into Sanskrit using hybrid approach which includes rule based approach and Example Based Machine Translation (EBMT) & from Sanskrit it is easier to transform into native Languages. EBMT has emerged as one of the most versatile, computationally simple and accurate approaches for machine translation. The main advantage of rule based approach is the easy implementation and small memory requirement. Thus we are going to combine both approaches to obtain a ‘good enough’ translation as opposed to a perfect translation aimed by earlier machine translation efforts. The basic objective of this system is to convert the simple English SVO format statements into its equivalent Sanskrit form. We look forward to this concept as being useful for generation of complex statements which includes ‘*sandhi*’ rules.

*Keywords:* machine translation, natural language processing, artificial intelligence, hybrid approach, rule based dictionary approach, example based approach (EBMT), Sanskrit.

**1. Introduction**

**Introduction:** In the past twenty years, much time, effort, and money has been expended on designing an unambiguous representation of natural languages to make them accessible to computer processing. The concept of machine translation domain of natural Language processing (NLP) and area of Artificial Intelligence (AI) is very useful in providing interaction user with a machine, which understands different languages spoken by the common man. It presents the user with an interface, with which he feels more comfortable.

We always in need for translators because of many languages are spoken in the world .The use of the machine translator is to make people speaking and understanding different languages , share ideas and views and also to communicate with one another people. With the help of Machine Translation concept, we may progress towards achieving the goal using easily available computer systems.

A survey of the machine translation systems that have been developed in India for translation from English to Indian languages and among Indian languages reveals that the MT software are used in field testing or are available as web translation service. These systems are also used for teaching machine translation to the students and researchers. English is a widely spoken language hence we can call it a "global language". The system will benefit the users who want to learn or understand Sanskrit or want to translate the text available in English into Sanskrit. But, why to use Sanskrit? Since, Sanskrit is mother of 85 percent Indian Language therefore it is an obvious choice for the target language. Sanskrit has been a favorite language for linguistics, as it has well defined grammar and is well structured. Sanskrit being a synthetic language is quite scientific in its approach and is based on formula or sutra.

There are various approaches to translate languages from source language to target language. In our system we are going to design and implement a system which use a hybrid approach i.e. a combination of two important approaches that is example based approach and rule based approach.

The Example Based Machine Translation (EBMT) is one of the most popular machine translation mechanisms which retrieve similar examples with their translation from the example data base and adapting the examples to translate a new source text. This will be a great attempt towards bringing people together and sharing their views if this project is successfully implemented.

**2. Problem Statement**

**Problem Statement**

Design and implementation of machine translator for English to sanskrit with hybrid approach including rule based and example based approach to obtain a good enough translation for SVO formats of English statement.

**3. Objective, Aim, Scope**

**Objective**

In short the objective of our proposed system is:

* Parsing the given English sentence.
* Detecting the most probable errors in the sentence.
* Translate statement by using example based approach if it is already in corpus.
* Translate the source code to target code by rule based approach if it didn’t found in corpus.
* Improving translator efficiency by providing bi-lingual dictionary, statement corpus and on-line dictionary support.

**Aim**

The aim of this is to provide an overview of our software product which we are going to develop.

Our aim is

* To translate basic and simple SVO(Subject-Verb-Object) pattern English statements into equivalent Sanskrit statements.
* To do translation with the help of hybrid approach to increase throughput.
* To provide English to Sanskrit bilingual dictionary support.

**Scope**

Translation of any natural language into its equivalent target natural language is quite complex task. It requires much more linguistic knowledge. It is impossible to write rules that cover all a language. In case of modifying some rules, it does not only change the incorrect sentence in to correct sentence, it may be further affecting on the correct sentence also. Common error occur in machine translation is one is choosing incorrect meaning and another one is incorrect sequence.

The translation process involves the many different rules interacting in many different ways. Due to which it is hard to extend or modify.

In translation of English to Sanskrit language there are various aspects to take into consideration like simple statements, sandhi, visarg rules, samasa rules, dhatu path rules etc.

Hence at this starting phase we are going to limit our scope for simple English S-V-O patterns statements with sandhi vichcheda rules. This aim we are going to achieve with the help of combining two different machine translation approaches i.e. example based and rule based machine translator.

**4. Feasibility Study**

Feasibility study is made to see if the project on completion will serve the purpose of the organization for the amount of work, effort and the time that spend on it. Feasibility study lets the developer foresee the future of the project and the usefulness. A feasibility study of a system proposal is according to its workability, which is the impact on the organization, ability to meet their user needs and effective use of resources. Thus when a new application is proposed it normally goes through a feasibility study before it is approved for development.

The document provides the feasibility of the project that is being designed and lists various areas that were considered very carefully during the feasibility study of this project.

**5. Technical Details**

**Technical Details:-**

We have used a hybrid approach i.e. a combination of the RBMT (Rule based Machine Translation) and EBMT (Example Based Machine Translation). This hybrid approach in its purest sense makes use of source and target parallel sentences along with the benefits of Rule Based Machine Translation.

For translation generally rule based, Statistical based and Example Based approaches are used and Rule Based methods for machine translation is divided in to another three types:

1. **Transfer based machine translation**: - Type of machine translation based on idea of Interlingua and is currently one of the most widely used area. It is necessary to have an intermediate representation that captures the “meaning” of the original sentence in order to generate the correct translation.

2. **Dictionary based machine translation**: -Machine translation can use dictionary based approach, which means that the words will be translated as they are by a dictionary.

3. **Interlingua based machine translation**: Machine translation can use dictionary based approach mean the text to be translated, is transformed into an interlingua, i.e. source or target language-independent representation. The target text is then generated out of the interlingua.

The Example Based Machine Translation (EBMT) approach is often characterized by its use of a bilingual corpus as its main knowledge base, at run-time. It is essentially a translation by analogy and can be viewed as an implementation of case-based reasoning approach of machine learning.

The advantage of this approach is that it can be language independent and makes minimal use of linguistic resources as they are sparse in our case. We aim to build rapid English to Sanskrit language Hybrid Transliteration systems making the most of the available minimal resources - parallel texts and bi lingual dictionaries.

Design is a process through which requirements are translated into a representation of software. The design of Hybrid approach of Machine Transliteration can be shown as:-

Morphological Analysis

Lexical Analysis

Pattern Directing Parsing

Rule Based Pattern

Example Based MT approach

Sanskrit Text Generator

**6. Technologies Used**

**Technologies Used**

|  |  |
| --- | --- |
| Technology Used | Core JAVA |
| Tools | JDK 1.5 and Above, Net beans 7.3.1 and Above, |
| Database | MySQL.SE 6 |
| Operating System | Windows XP Or Above |

**7. Innovativeness and Usefulness**

**Innovativeness and Usefulness**

* Sanskrit and computers are a perfect fit. NASA the most advanced research centre in the world for cutting edge technology has discovered that Sanskrit, the world's oldest spiritual language is the only unambiguous spoken language on the planet. NASA and others have been looking at Sanskrit as a possible computer language since its syntax is perfect and leaves little room for error. Hence to help the scientist to define various kinds of things in Sanskrit language this system will be used as an initial step.
* It will help in mass media as there has been a short daily news broadcast on state-run All India Radio. Sanskrit news is broadcast on TV and on the internet as part of the DD National channel. Hence application will guide and help reporters in making news.
* It will ease the translation of English to Sanskrit in the state of Uttarakhand where Sanskrit as its second official language.
* For the foreigners as Sanskrit is the liturgical language of Hindus, it is used during worship in Hindu temples throughout the world.
* Most of the organization in various countries especially organizations of government uses Sanskrit language in their motto. This application will guide the organizations to decide the motto.

**8. Project Overview**

**Project Overview**

There are different approaches used for the machine translation form one language to another language. Machine translation can apply a technique based on linguistic set of laws, which means that words will be translated in a linguistic mode the most suitable words of the target language will replace the ones in the source language. Three different approaches to MT have been generally used-

1. **Rule Based machine translation (RBMT)**
2. **Example based machine translation (EBMT)**
3. **Statistical machine translation (SMT) approach** and recent approaches to machine translation are knowledge-based, corpus-based, hybrid methods etc.

The rule based machine conversion model includes transfer based machine translation, interlingua machine conversion and dictionary based machine conversion paradigms. It uses the rules for their mapping process. The concept of RBMT systems is based on between the structure of the given source sentence with the structure of the target output sentence, necessarily preserving their unique meaning. The example based machine translation (EBMT) approach to machine conversion is often characterized by its use of a bilingual amount or corpus with similar texts as its main source language knowledge base, at run-time. The Example Based Machine Translation (EBMT) is one of the most popular machine translation mechanisms which retrieve similar examples with their translation from the example data base and adapting the examples to translate a new source text.

In proposed system we are going to combine both approaches and will try to get best features from those. Hybrid machine translation (HMT) leverages the strengths of example-based and rule-based translation methodologies.

In proposed system we will provide simple SVO order English statements as an input and we will get corresponding Sanskrit statement.

The flow will be like this; first the statement is get send to the example based machine translator (EBMT) module. Here it will get checked with the present corpus, now if system found the statement already present in the corpus then it will give directly output. If we didn’t found any corpus in present database then it will get transfer to the rule based machine translator (RBMT) module. Here input will go through various steps.

Tokenization: This module splits the given sentence into chunks of strings delimited by spaces. These strings may be simple words or compound words coalesced by the rule of English Grammar. By applying the rules of English grammar assign appropriate category to words like (noun, verb, noun phrase etc.)

*Tagging*: Here each word is get tagged with appropriate means.

*Parse tree generator*: It generate a parse tree using grammar rules of source language.

*Bilingual Dictionary & Sanskrit sentence generator*: Here it’ll find translation of all English words into Sanskrit dictionary and rearrange the words in Sanskrit using its grammar rule to format and a meaningful statement.

In this paper we consider dictionary rule based approach is for translation and synthesizer. In dictionary based approach words are stored in Database dictionary and when we got input then English sentence are separated from sentence by tokenization then morphological analysis is done, after getting the words its search into English dictionary and according to word its category e.g. (noun, verb) is assigned. If we compare the Grammar for both English and Sanskrit then English sentences always in order of subject-verb-object format while Sanskrit has free word order.

**9. Papers Referred**

**Papers Referred**

[1] Vimal Mishra and R. B. Mishra “Study of Example Based English to Sanskrit Machine Translation”

[2] Sinha RM.K. and Jain A., “AnglaHindi: An English to Hindi Machine-Aided Translation

System.”, “MT SUMMIT IX”, New Orleans, Louisiana, USA-2003

[3] English to Sanskrit Machine Translator Lexical Parser And Semantic Mapper Ms.Vaishali.M.Barkade", Prof. Prakash R. Devale ", Dr. Suhas H. Patil'.

[4] A Design Towards English to Sanskrit Machine Translation Synthesizer System Using Rule Based Approach D. T. MANE, P. R. DEVALE AND S.D. SURYAWANSm

[5] Jyoti Srivastava, Sudip Sanyal, “A Hybrid Approach for Word Alignment in English-Hindi Parallel Corpora with Scarce Resources” 2012 International Conference on Asian Language Processing.

**10. References**

**References**

[1] Vimal Mishra and R. B. Mishra “Study of Example Based English to Sanskrit Machine Translation”

[2] Sandeep R. Warhade, Prakash R. Devale, Dr. S.H.Patil , “Statistical Machine Translation Approach for English-to-Sanskrit Translation in Ubiquitous Environment”, International Journal of Engineering and Innovative Technology (IJEIT), June 2012

[3] Sitender, Seema Bawa,Dept. of CSE, Thapar University, Patiala, Punjab, India “Survey of Indian Machine Translation Systems” IJCST -Jan. - March 2012

[4] Ms.Vaishali.M.Barkade, Prof. Prakash R. Devale , Dr. Suhas H. Patil “English to Sanskrit Machine Translator Lexical Parser And Semantic Mapper”, National Conference On "Information and Communication Technology[NCICT-10]

[5] D. T. MANE, P. R. DEVALE AND S.D. SURYAWANS , “A Design Towards English to Sanskrit Machine Translation Synthesizer System Using Rule Based Approach, IJMRAE, July2010,

[6] Latha R. Nair, David Peter S. “Machine Translation Systems for Indian Languages”, International Journal of Computer Applications (0975 – 8887), February 2012

[7] Sinha RM.K. and Jain A., “AnglaHindi: An English to Hindi Machine-Aided Translation System.”, “MT SUMMIT IX”, New Orleans, Louisiana, USA-2003

[8] Mrs. Namrata Tapaswi, Dr. Suresh Jain, Mrs. Vaishali Chourey, “PARSING SANSKRIT SENTENCES USING LEXICAL FUNCTIONAL GRAMMAR”, 2012 International Conference on Systems and Informatics.

[9] Jyoti Srivastava, Sudip Sanyal, “A Hybrid Approach for Word Alignment in English-Hindi Parallel Corpora with Scarce Resources” 2012 International Conference on Asian Language Processing.

**Internet Sources**:

[1] http://bhavinpathak.webs.com/sanskritinmodernera.htm

[2] http://en.wikipedia.org/wiki/Sanskrit

[3] http://www.vedicsciences.net/articles/sanskrit-nasa.html