International Journal of Computer Applications Technology and Research
Volume 3– Issue 9, 547 - 549, 2014, ISSN: 2319–8656

A Review on a web based Punjabi to English Machine Transliteration System

Navpreet kaur
C.S.E Department
G.Z.S PTU Campus
Bathinda, India

Paramjeet Singh
C.S.E Department
G.Z.S PTU Campus
Bathinda, India

Shveta Rani
C.S.E Department
G.Z.S PTU Campus
Bathinda, India

Abstract: The paper presents the transliteration of noun phrases from Punjabi to English using statistical machine translation approach.Transliteration maps the letters of source scripts to letters of another language.Forward transliteration converts an original word or phrase in the source language into a word in the target language.Backward transliteration is the reverse process that converts the transliterated word or phrase back into its original word or phrase.Transliteration is an important part of research in NLP.Natural Language Processing (NLP) is the ability of a computer program to understand human speech as it is spoken.NLP is an important component of AI.Artificial Intelligence is a branch of science which deals with helping machines find solutions to complex programs in a human like fashion.The transliteration system is going to developed using SMT.Statistical Machine Translation (SMT) is a data oriented statistical framework for translating text from one natural language to another based on the knowledge.

Keyword: Transliteration,Mapping,Translation,Dictionary

1. INTRODUCTION

Transliteration is a process that maps the sounds of one language to scripts of another language.The system performs the process of transliteration of noun phrases of Punjabi to English using SMT approach.Punjabi Language is written from left to right using gurmukhi script and Punjabi language consist of consonents, vowels, halant, punctuation and numerals.The gurmukhi script was derived from sharda script.The Punjabi Language contains Thirty-five distinct letters.English language is written in roman scripts.There are 26 letters in English.Out of which 21 is consonants and 5 are vowels.Punjabi language is an official language of Punjab.It can be understand or read by the person who knows Punjabi.Opposite to it English is an international language.so the person who have no knowledge about Punjabi can convert the file Written in Punjabi into English using Punjabi to English transliteration system.SMT uses the concept of development of Machine learning system from the existing names stored in the database system.Development of database table for uni-gram,bi-gram,tri-gram,four-gram,five-gram,six-gram and upto ten-gram to store the results obtained from the learning phase of the system.Various algorithms for conversion of anmollipi into Unicode is used so that it can be used as input to the system This topic of machine transliteration has been used in different language to convert from one language to another language.Various techniques has been applied to this system Diect mapping like rule based approach etc.Transliteration is different from Translation system.Translation from Punjabi to English means to translate each word in Punjabi to its English equivalent whereas the transliteration means to write them sensing the characters in the word e.g. “nvdIp “in Punjabi is transliterated in English as “navdeep” where n for “n” v for “v” d for “d” p for “p” .This system can be developed using transliteration process using a database of transliterating characters.To develop this system first of all we have to collect names of proper nouns from various sources such as person names,cities rivers,countries,states etc.We have to store these names in Punjabi and its English equivalent in database.Then we have to develop an algorithm to convert the Punjabi font into Unicode so that it can be given as input to the system.Then to develop the algorithm for learning phase of the system.The system will learn from existing data entries.

Three Main Approaches are used for machine translation:

Direct Machine Translation (DMT) system is a simple form of machine translation system. In DMT, a word to word translation of the input text is performed and the result is obtained in the DMT, a language which is called a source language (Punjabi) is given as input and the output is received which is called a target form of output text.

Rule Based Machine Translation (RBMT) is also known as Knowledge Based Machine Translation system. It is a system which is based on linguistic infomation related to source and target languages and retrieves this information from dictionaries (bilingual) and grammars which includes semantic and syntactic information of each language. RBMT system generates output text from this information.

Statistical Machine Translation (SMT) is a new approach which is based on statistical models and in this approach; a word is translated to one of a number of possibilities based on the probability. The whole process is performed by dividing sentences into N-grams. N-gram is a contiguous sequence of n items from a given text. The items can be phonemes, letters, and words. An N-gram of size 1 is known as a unigram; size 2 is a bigram; size 3 is a trigram. Larger sizes are represented by the value of n i.e. four-gram, five-gram and so on. Statistical system will analyze the position of N-grams in relation to one another within sentences.

2. EXISTING WORK

Transliteration and translation has been studied in different languages.These systems has been developed in different languages pairs.We have studied different literature related to transliteration system.Gurpreet singh josan and gurpreet singh lehal has developed Punjabi to hindi machine transliteration system by combining character to character mapping using rule based approach.This paper shows that the system produced transliteration in hindi from Punjabi with an accuracy of 73% to 85%.Vishal goyal and Gurpreet singh has developed hindi to Punjabi machine translation system using the rule based techniques.The overall efficiency of this system hindi to Punjabi is 95%.Another system has been developed by Kamaldeep and Dr. Vishal goyal of using hybrid approach for Punjabi to English transliteration system.This paper presents the Punjabi to English machine transliteration using letter to letter mapping as baseline and try to find out the improvements by statistical methods.To improve the accuracy various rules has been developed.Author has developed hybrid (statistical + rules) approach based transliteration system.Independent vowel mapping,dependent vowel mapping,consonant mapping,mapping of special symbols table is defined.The Overall accuracy of the system comes out to be 95.23%.Kamaljeet kaur batra and G.S.Lehal has developed rule based machine translation of noun phrases from punjabi to English.The paper presents the automatic translation of noun phrases from Punjabi to English using transfer approach.The system has analysis,translation and synthesis components.The steps involved are preprocessing,tagging,ambiguity resolution,translation and synthesis of words in target language.The accuracy is calculated for each step and the overall accuracy of the system is calculated to be about 85% for a particular type of noun phrases

3. PROBLEM

The problem domain to which this project is concerned is machine transliteration.In foreign and in some areas of india other than Punjab,most of population is not so familiar with Punjabi.As we know that all the data of government sector of Punjab is in Punjabi language because Punjabi is an official language of Punjab,people who are unaware of Punjabi can't understand it.For e.g.punjab state government has to send the report of malnutrition children to UNO.As all the reports are generally created in Punjabi language but it is not useful in foreign so there is a need to present it in English language,here the transliteration system is useful.Existing systems has been developed with mostly rule based techniques and hybrid techniques.we can't make as many rules as possible.We can develop this system with the help of SMT technique which can increase the efficiency of the system.In existing system some errors are occur e.g. sometimes when a name is pronounced in Punjabi it correspond to many English words. e.g."rxjIq" is convert in english as ranjeet,ranjit.So that system fail to guess which one is the best.Sometime user does not enter correct data due to which output is also not correct.e.g."mRIq" it is wrongly enter data we cannot use "R" with "m" in Punjabi language.Another issue related to the difference in the number of characters in Punjabi and English languages.There is a difference in the number of vowels and consonents.Sometime single character to multiple mapping are occur e.g. "v" can be used as v,w.So there is a need to develop algorithm to select the appropriate character at different situations.Existing system is developed on the bases of direct and rule based approach.They are using direct approach due to which the accuracy of system is very low.

4. CONCLUSION

In this paper we have discussed about the transliteration system which has been developed in different languages.Different techniques has been used to develop this system.the accuracy of each system is studied.The paper has addressed the problem arising in transliteration of Punjabi to English.This system can be developed with additional efforts.There are many issues left for further improvement.the system could be improved by improving the techniques.The system can be effectively developed with the help of using SMT technique.SMT take the view that every sentence in the target language is a translation of the source language sentence with some probability.The best translation is the sentence that has highest probability.The system can be develop by using database table for uni-gram,bi-gram and upto ten-gram to store the results obtain from the learning phase of the system.In Punjab state most of the official work is done in Punjabi language,so this transliteration system will help them a lot to transliterate Punjabi to English.

5. REFERENCES

[1] Gurpeet Singh josan and Gurpreet Singh lehal,A Punjabi to Hindi machine transliteration system,Computational Linguistics and Chinese language processing vol.15.no.2.june 2010 ,pp.77-102

[2]Vishal Goyal and Gurpreet Singh Lehal,Evaluation of hindi to Punjabi machine translation system,IJCSI international Journal of computer science issues,vol.4.no.1,2009 ISSN(Online):1694-0784

[3]Kamaldeep ,Dr.vishal Goyal,hybrid approach for punjabi to English transliteration system International journal of computer applications (0975-8887) volume 28-no.1,August 2011.

[4]Sumita rani,Dr.Vijay Laxmi,A review on machine Transliteration of related languages:Punjabi to Hindi

international journal of science,Engineering and technology research (IJSETR) volume 2,issue 3,march 2013

[5]Gurpreet Singh Josan1& Jagroop Kaur, “Punjabi to Hindi statistical machine transliteration” International Journal of Information Technology and Knowledge Management July-December 2011, Volume 4, No. 2, pp. 459-463.