

Multi Script Text Transliteration to Punjabi

Kamaljeet Kaur and Parminder Singh*

Department of Computer Science and Engineering, Guru Nanak Dev Engineering College,
Ludhiana - 141006, Punjab, India; parminder2u@gmail.com

Abstract

Machine Transliteration is becoming an important research area in Machine Translation, which ensures phonology of target language word. This paper describes the transliteration system that has been developed for multi script text. The proposed system is capable of transliterating English and Hindi text into equivalent Punjabi text using a rule based transliteration approach. Transliteration is more phonetic process because transliteration is more closely related with utterance. Thus there is need to perform phonetic analysis in addition to orthography for forming rules. The accuracy of English-Punjabi and Hindi-Punjabi transliteration is found to be 94% and 98% respectively.

Keywords: Grapheme, Orthography, Phonetic, Translation, Transliteration, Transliteration Rules

1. Introduction

Transliteration means transliterating a word written in one language to another language without compromising with its phonetic properties. Machine translation is a process that transforms/translate message written in one script into message written in target script. It keeps exact meaning of the word. The most important requirement of machine transliteration system is to preserve the phonetic properties of source language after the transliteration in target language. Thus during transliteration, we have to maintain syllables' sound in words. In this paper we have presented an approach toward Multi Script Text Transliteration. The system of multi script transliteration can transliterate English and Hindi text into target Punjabi text.

The remaining part of this paper is organized in 4 sections. In section 2 we have described the related work that is done in the field of machine transliteration. Then we have described basic approach and various rules for multi script transliteration in section 3. In section 4, performance evaluation is discoursed. Finally in section 5, we have given conclusion of it.

2. Related Work

In transliteration, significant research has been done for Indian scripts and for foreign scripts. Knight and Graehl have presented a statistical model. The model is based on phoneme and it uses finite state transducer. This model applies various rules for mapping to perform backward transliteration. To transliterate words from Japanese back to English, they have proposed a method for automatic backward transliteration¹⁰. Lee and Choi have presented a statistical transliteration model for transliteration from English language to Korean language. They have compared two SMT based methods i.e. direct method and pivot method. After comparing two methods, they have proposed a more effective method for transliteration that is hybrid one¹¹. Oh and Choi discussed phoneme based English language to Korean language transliteration system. Phonetic information, orthography information, and information regarding context is used in this system¹³. Punjabi Machine Transliteration System has developed by Malik that uses transliteration rules to map words written in Shahmukhi script to its corresponding words written in Gurmukhi script¹². Saini and Lehal

* Author for correspondence

have performed transliteration from Shahmukhi script to Punjabi language. They have proposed a corpus based system for transliteration. The proposed system has been successfully tested on a small data of story, article, and poetry. It has claimed 91.37% of average transliteration accuracy¹⁵. To transliterate Hindi words into Punjabi equivalent, a transliteration system has proposed by Goyal and Lehal. Complex set of rules have implemented for more accurate transliteration⁵. Vijaya et al. have developed a system called “WEKA” that can perform transliteration from English language to Tamil language. They have demonstrated a transliteration model based on multi class classification approach. It is a Rule based system that has resulted accuracy rate of 84.82%¹⁶. Josan and Lehal have implemented a system for Punjabi to Hindi words transliteration. They have combined simple approach of character to character mapping with various rules and Soundex based enhancements. There is some dependency or contextual rules that are manually crafted⁷. Kaur and Josan have proposed transliteration system from English to Punjabi. It is statistical machine transliteration system that is created by using MOSES. MOSES is a statistical machine transliteration tool. The system has attained average accuracy of 63.31%⁹. Deep and Goyal have proposed a by using various mapping rules to transliterate person names from Punjabi to English. It is grapheme based method. The proposed technique has demonstrated transliteration for common names of persons, cities, states, rivers etc. and achieved 93.22% accuracy². Josan and Kaur have proposed Punjabi to Hindi text transliteration model. It is a statistical model that is builds on statistical techniques. It has claimed accuracy rate of 87.72%⁶. Dhore et al. have focused on transliteration of Hindi and Marathi to English using a phonetic based direct approach³. Rathod and Dhore have presented machine transliteration for Hindi and Marathi to English using Support Vector Machine. The system is tested for person names and place names and achieved accuracy of 86.52% for 5-gram¹⁴. Bhalla et al. have proposed English to Punjabi transliteration scheme using various transliteration rules. For name entities such as proper names and location names, probabilities are calculated. The proposed scheme has claimed 88.19% accuracy¹. Joshi et al. have used statistical machine learning approach to do transliteration from Roman script to Devanagari script⁸.

3. Methodology

3.1 Approach Followed

Rule based approach is followed for multi script text transliteration system. To achieve high performance and more accurate result of transliteration we have considered grapheme information and phonetic information during rule formation.

3.2 Proposed Machine Transliteration Process

To produce final transliteration result in target language i.e. Punjabi, the multi script input text passes through various phases, as shown in following Figure 1,

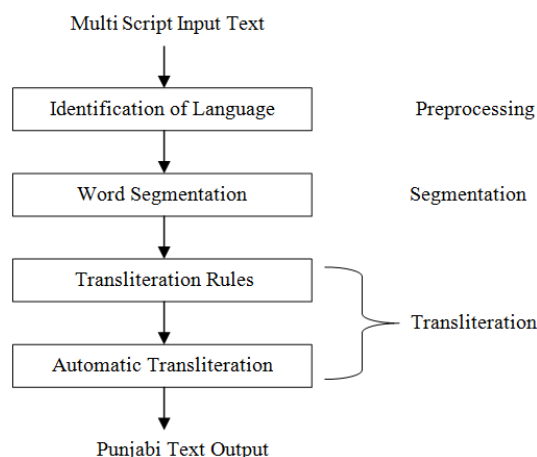


Figure 1. Basic Transliteration Approach.

3.2.1 Preprocessing

The main step of this module is to identify the language. The language is identified according to the Unicode of input words.

3.2.2 Segmentation

This module segments the source string. It generates transliteration units of Hindi/English language. It segments English and Hindi strings.

3.2.3 Transliteration

Based on framed rules, transliteration module maps English and Hindi text into resultant Punjabi text.

Preceding and following characters in a given word are also taken into consideration for more appropriate result.

3.3 English-Punjabi Transliteration Rules

In most of the cases, consonants of English language are directly mapped to the consonants of Punjabi language². But only direct character to character mapping is not enough for performing transliteration, because it may lead to poor result of proposed system. So there is need to develop different rules to improve accuracy of the system.

3.3.1 First Vowel Character and Last Vowel Character in Word

Table 1. Rules for First Vowel Character

Character Combinations at First Position	Source Word	Target Mapping	Target Word
'a' followed by consonant	aman	'ਅ'	ਅਮਨ
'an'	ankit	'ਅੰ'	ਅੰਕਿਤ
'ea'	eagle	'ਈ'	ਈਗਲ
'ei'	eight	'ਏ'	ਏਟ
'in'	India	'ਇ'	ਇੰਡੀਆ
'oi'	oil	'ਓ' + 'ਇ'	ਓਇਲ
'u' followed by 'double consonant'	utter	'ਉ' + 'ੳ'	ਉੱਤਰ

Table 2. Rules for Last Vowel Character

Character Combinations at Last Position	Source Word	Target Mapping	Target Word
'ia'	sonia	'ਆ'	ਸੋਨੀਆ
'a' preceded by consonant	samrala	'ਾ'	ਸਮਰਾਲਾ
'nu'	sonu	'ੰ' + 'ਊ'	ਸੋਨੂੰ
'oa'	goa	'ੋ' + 'ਆ'	ਗੋਆ
'e' + 2nd last character 'i'	like	Skip 'e'	ਲਾਈਕ

3.3.2 CVCC Pattern (Double Consonants)

If there are double consonants in English word then they are transliterated into Punjabi gemination symbol i.e. 'Addak'⁹. There are 2 geminates that represent nasalized sound i.e. 'mm' and 'nn'. So in most of the cases they are written with 'Tippi'.

Table 3. Rules for Double Consonants

Character Combinations	Source Word	Target Mapping	Target Word
'a' + double consonant	bhatt	'ੳ'	ਭੱਟ
'e' + double consonant	dress	'ੈ' + 'ੳ'	ਡਰੈੱਸ
'i' + double consonant	gill	'ਿ' + 'ੳ'	ਗਲਿ
'o' + double consonant	boss	'ੋ'	ਬੋਸ
'u' + double consonant	full	'ੂ' + 'ੳ'	ਫੁੱਲ

3.3.3 CVC Pattern (Short Vowels) and CVCe Pattern

Table 4. Rules for Short Vowels

Short Vowel	Source Word	Target Mapping	Target Word
'a'	fan, scan	'ੈ'	ਫੈਨ, ਸਕੈਨ
'e'	bed, pen	'ੈ' + 'ੳ'	ਬੈੱਡ, ਪੈੱਨ
'i'	big, swim	'ਿ' + 'ੳ'	ਬਿਗ, ਸਵਿਮਿ
'o'	dog, stop	'ੋ'	ਡੋਗ, ਸਟੋਪ
'u'	cut, bus	'ੳ'	ਕੱਟ, ਬੱਸ

Table 5. Rules for CVCe Pattern

Character Combinations	Source Word	Target Mapping	Target Word
current_char is 'a' second_next_char is 'e'	grace	'ੈ'	ਗਰੇਸ
current_char is 'i' second_next_char is 'e'	like	'ਿ' + 'ਈ'	ਲਾਈਕ
current_char is 'o' second_next_char is 'e'	vote	'ੋ'	ਵੋਟ
current_char is 'o' second_next_char is 'e'	tune	'ੀ' + 'ਊ'	ਟੀਊਨ

3.3.4 Resolving Ambiguities

In every language there are some words or characters that represent ambiguous pronunciation. For example, English

consonant 's' represent different pronunciations like 'ਸ' or 'ਜ' in Punjabi. For example pronunciation of word 'base' or 'page'. Other ambiguous character is 'g'. There is need to refer to surrounding characters in order to resolve this problem. Rules for handling such ambiguous characters are discussed in Table 6.

3.3.5 Silent Letters

Various rules are also framed for different silent letter combinations like ck, wr, kn, mn, mb, mp, dg, gh, wh etc. E.g.: trickà ਟਰਕਿ, wholeà ਹੋਲ, writeà ਰਾਈਟ, knowà ਨੋ, columnà ਕੋਲਮ, companyà ਕੰਪਨੀ, fridgeà ਫਰਿਜਿ, thoughtà ਥੋਟ, etc.

3.3.6 Unreduced Vowel (Schwa and Consonant)

Certain words in English language are often pronounced with full vowels with unstressed syllables. For example words careà ਕੇਅਰ, affairà ਅਫੇਅਰ, realà ਰੀਅਲ, potentialà

ਪੋਟੈਂਸ਼ੀਅਲ, crucialà ਕਰੁਅਲ, hereà ਹੇਅਰ, gorgeouslyà ਗੋਰਜੀਅਸ etc.

3.4 Hindi-Punjabi Transliteration Rules

Most of the letters in Hindi have one to one correspondence with Punjabi letters. Hindi and Punjabi are more closely related languages because there are various similarities in the syntax and vocabulary of both languages, the sound of most of the letters is same in both scripts. But only direct mapping does not provide accurate result. There is need to develop different rules for more accurate transliterations⁵.

3.4.1 Rules for Ambiguity Resolution

Hindi words may have more than one representation based on the perception of the transliterator. For example Hindi word 'प्यास' can be represented in Punjabi as 'ਪਿਆਸ', 'ਪੀਆਸ', 'ਪਯਾਸ', 'ਪਯਾਸ'. Most of the ambiguous words are

Table 6. Rules for Ambiguity Resolution

Character Combinations	Source Word	Mapping	Target Word
'g' followed by 'en'/'in'/'ic'/'im'	General, margin, logic	'ਜ'	ਜਨਰਲ, ਮਾਰਜਨ, ਲੋਜਿਕ
'ange'/'enge'	Orange, challenge	'ਜ'	ਓਰੇਂਜ, ਚੈਲੇਂਜ
'g' preceded by 'a'/'d' + followed by 'e'	page bridge	'ਜ'	ਪੇਜ, ਬਰਜਿ
's' preceded by 'ou'/'a' + followed by 'e'	mouse, base	'ਸ'	ਮਾਊਸ, ਬੇਸ
last 's' preceded by 'sse'/'che'/'ie'	classes, benches, parties	'ਜ'	ਕਲਾਸਜ਼, ਬੈਚਜ਼, ਪਾਰਟੀਜ਼
's' followed & preceded by vowel	Season	'ਜ਼'	ਸੀਜ਼ਨ

Table 7. Rules for Ambiguity Resolution in Hindi

Character Combinations	Hindi Word	Mapping	Punjabi Word
'य' is at first position	यादव	'ਯ'	ਯਾਦਵ
'यो'/'ये'/'या'/'यी' are at last position	कोरीयो, कारवायी	'ਓ'/'ਏ'/'ਯਾ'/'ਯੀ'	ਕੋਰੀਓ, ਕਾਰਵਾਈ
'या' is preceded by half character	प्यास	'ਿ'+'ਆ'	ਪਿਆਸ
'य' is followed by 'ी'/'इ'/'/ँ'	सोसायटी, सहायता, दायरे	'ਇ'	ਸੋਸਾਇਟੀ, ਸਹਾਇਤਾ, ਦਾਇਰੇ
'य' is followed by 'ि'	रासायनकि	'ਈ'	ਰਾਸਾਈਨਕਿ
'या' is preceded by 'ा'	लगाया	'ਇ'+'ਆ'	ਲਗਾਇਆ
'यू' is preceded by half character	न्यू	'ਿ'+'ਊ'	ਨਊ
'यु' is preceded by half character	व्यु	'ਿ'+'ਊ'	ਵਊ
'ण' is preceded by 'य'	नरायण	'ਇ'	ਨਰਾਇਣ
'यो' is preceded by half character	क्योर	'ਿ'+'ਓ'	ਕਓਰ
second and last character is 'या' and previous is consonant	गया	'ਿ'+'ਆ'	ਗਆ
'या' is preceded by 'ि'/'ी'	रीया	'ੀ'+'ਆ'	ਰੀਆ

Table 8. Rules for Middle Vowel

Character Combinations	Hindi Word	Punjabi Mapping	Punjabi Word
'ह' is preceded by 'े'	सेहत	'ि'	ਸਹਿਤ
'या' is preceded by 'े'	तैयार	'ि'	ਤਿਆਰ
'ह' is followed by '्र'	महफूज	'ਹਿ'	ਮਹਫੂਜ
'ि' is followed and preceded by 'ा'	जातविद	'ੀ'	ਜਾਤੀਵਾਦ
'ि' is preceded by half character	अग्निपिथ	'ੀ'	ਅਗਨੀਪਿਥ
'ि' is at second last position and 'या' is at last	एशिया	'ੀ'	ਏਸ਼ੀਆ
'ि' is at last	स्थिति	'ੀ'	ਸਥਿਤੀ

resulted due to the presence of Hindi consonant 'य'. The framed rules for resolving ambiguities are discussed in following Table 7.

3.4.2 Rules for Middle Vowel Character

Hindi vowels are mapped to different target characters based on their position in given word. For example Hindi vowel 'ि' is not always mapped to 'ि'. Sometimes it is mapped to 'ੀ' according to its position and surrounding characters in a given word. Table 8 shows various rules proposed for middle vowel character.

3.4.3 Rules for Half Characters

Half character form is used to represent sound duplication in Hindi language. But in Punjabi, gemination adhak 'ँ' is there to duplicate consonant sound⁶. The mapping of half character in target language is greatly depends on the next character combination. The mapping of such characters is shown in following Table 9.

Table 9. Rules for Half Characters

Character Combinations	Hindi Word	Punjabi Mapping	Punjabi Word
half 'क' is followed by 'ख'	मक्खी	'ੱ'	ਮੱਖੀ
half 'ग' is followed by 'घ'	मग्घ	'ੱ'	ਮੱਘ
half 'च' is followed by 'छ'	इच्छा	'ੱ'	ਇੱਛਾ
half 'ज' is followed by 'झ'	मज्झ	'ੱ'	ਮੱਝ
half 'ट' is followed by 'ठ'	मट्ठ	'ੱ'	ਮੱਠ
half 'त' is followed by 'थ'	पत्थर	'ੱ'	ਪੱਥਰ
half 'द' is followed by 'ध'	बुद्ध	'ੱ'	ਬੁੱਧ
half 'प' is followed by 'फ'	रप्फ	'ੱ'	ਰੱਫ
half 'ब' is followed by 'भ'	लब्भ	'ੱ'	ਲੱਭ

3.4.4 Rules for Nasalized Characters

Nasalization in Hindi is presented by anusvara, chandrabindu and some nasalized consonants. Nasal character 'Anusvara' and 'Chandrabindu' of Hindi are mapped to either 'Bindi' or 'Tippi' in Punjabi according to the matching rules. Commonly used nasalization consonants in Hindi are 'म', 'न' and 'ण'. When written with vowels or half character form then they are either mapped to 'Bindi' or 'Tippi' in Punjabi. Table 10 shows various rules proposed for handling Hindi nasalized characters.

Table 10. Rules for Nasalized Characters

Character Combinations	Hindi Word	Punjabi Mapping	Punjabi Word
'ँ' is not preceded by 'ँ' / 'ँ' / 'ँ' / 'ँ' / 'ँ'	मंत्री	'ੰ'	ਮੰਤਰੀ
'ँ' is at last position	दायराँ	'ँ'	ਦਾਇਰੇ
'ँ' is preceded by 'ँ' / 'ँ' / 'ँ' / 'ँ' / 'ँ'	कांग्रेस	'ँ'	ਕਾਂਗਰੇਸ
'ँ' is preceded by 'ँ' / 'ँ' / 'consonant'	मुकुंद	'ੰ'	ਮੁਕੁੰਦ
'ँ' is preceded by 'ँ' / 'ँ' / 'consonant'	हूँ	'ੰ'	ਹੂੰ
'ँ' is at last	कहाँ	'ँ'	ਕਹਾਂ
half 'म' / half 'न' / half 'ण'	पन्ना, पण्डति	'ੰ'	ਪੰਨਾ, ਪੰਡਤਿ
half 'म' / half 'न' are preceded by 'ँ' / 'ँ' / 'ँ' / 'ँ' / 'ँ' / 'ँ'	फ्रान्स	'ੰ'	ਫਰਾਂਸ

3.4.5 Rules for Common Conjunct Consonants

Conjunct consonant is a combination of two or more consonants which are pronounced together without the pronunciation of the inherent ‘a’ vowel between them. Hindi conjunct consonants have no direct mapping in Punjabi. These letters are mapped according to the rules as shown in Table 11.

Table 11. Hindi Conjunct Consonants Mapping

Hindi Characters to form Conjunct	Hindi Conjunct Consonants	Corresponding Punjabi Characters
क+क्ष	क्ष	ਕ+ਸ਼
ज+ज्	ज्	ਗਿ+ਅ
क+त्	क्त	ਕ+ਤ
त+र्	त्र	ਤ+ਰ
द+व	द्व	ਦ+ਵ
श+र्	श्च	ਸ਼+ਚ
द+द्	द्	ੱ+ਦ
द+ध	द्ध	ੱ+ਧ

4. Performance Evaluation

Word Accuracy Rate is used to evaluate machine transliteration system. It is defined in percentage i.e. number of accurate transliteration by total transliterations produced by proposed system. This system is successfully evaluated on data of news of more than 10,000 words and

other domains like names of persons, locations etc. The word accuracy rate of proposed transliteration system is shown in following Table 12. Figure 2 shows the snapshot of multi script text transliteration system.

Table 12. Accuracy of Transliteration System

Language	Accuracy (%)
English	94%
Hindi	98%

5. Conclusion

Here we have presented machine transliteration system that is developed for multi script text using a rule based approach. Transliteration does not rely only on source grapheme, it is a complex process, which. Because the transliteration is more phonetic, it is difficult to attain accurate results without considering phonetic information. Phonetic transliteration is followed to generate Punjabi equivalent of English and Hindi that works according to the transliteration rules. Besides direct character-to-character mapping, various rules have also been devised for transliteration in order to resolve various ambiguities and thus to obtain more accurate transliteration. Both orthographic and phonetic information is considered.

It is not easy task of developing transliteration system for different languages like English and Punjabi due to lot of differences in syntax and vocabulary. Transliteration of different languages pair is little bit difficult because there are many factors that affect the result such as difference in

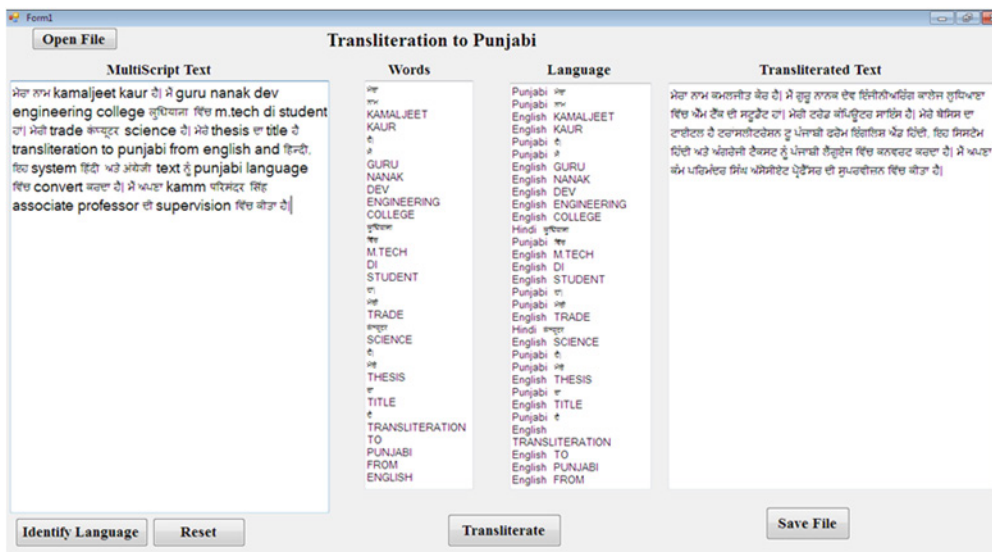


Figure 2. Snapshot of Multi Script Text Transliteration System.

Table 13. Result of Multi Script Text

Input Text	Output Text
<p>the आम आदमी पार्टी today said that the बीजेपी government economic policies were the same as pursued by the earlier मन्मोहन सिन्हा government, in a statement आप leader professor मन्मोहन सिन्हा said the people had expected a cut in the मा and डीजल prices by the new मोदी government, मोदी had promised the people to bring down the price of various commodities he pointed out, नरेंद्र मोदी said that the hike in the डीजल rates would hit the farmers it would also lead to inflation which बीजेपी had promised to curtail, the आम आदमी पार्टी leader said that the सरोमही अकाली दल a partner in the बीजेपी government must ask the नरेंद्र मोदी government to bring down the डीजल and पेट्रोल prices</p>	<p>ਦ ਆਮ ਆਦਮੀ ਪਾਰਟੀ ਟੂਡੇ ਸੈਂਡ ਦੈਟ ਦ ਬੀਜੇਪੀ ਗਵਰਨਮੈਂਟ ਏਕੋਨੋਮੀਕ ਪੋਲੀਸੀਜ਼ ਵਰ ਦ ਸੇਮ ਐਜ ਪੁਰਸੀਊਡ ਬਾਈ ਦ ਅਰਲੀਅਰ ਮਨਮੋਹਨ ਸਿੰਘ ਗਵਰਨਮੈਂਟ, ਇਨ ਅ ਸਟੇਟਮੈਂਟ ਆਪ ਲੀਡਰ ਪ੍ਰੋਫੈਸਰ ਮਨਮੋਹਨ ਸਿੰਘ ਸੈਂਡ ਦ ਪੀਪਲ ਹੈਡ ਐਕਸਪੈਕਟਡ ਅ ਕੱਟ ਇਨ ਦ ਪੇਟਰੋਲ ਐਂਡ ਡੀਜਲ ਪ੍ਰਾਈਸਜ਼ ਬਾਈ ਦ ਨੀਊ ਮੇਦੀ ਗਵਰਨਮੈਂਟ, ਮੇਦੀ ਹੈਡ ਪ੍ਰੋਮੀਸਡ ਦ ਪੀਪਲ ਟੂ ਬਰਗਿ ਡਾਊਨ ਦ ਪ੍ਰਾਈਸ ਆਫ ਵੇਰੀਅਸ ਕੋਮੋਡਿਟੀਜ਼ ਹੀ ਪੇਆਇੰਟਡ ਆਊਟ, ਨਰੇਂਦਰ ਮੋਦੀ ਸੈਂਡ ਦੈਟ ਦ ਹਾਈਕ ਇਨ ਦ ਡੀਜਲ ਰੇਟਸ ਵੁੱਡ ਹਟਿ ਦ ਫਾਰਮਰਸ ਇਟ ਵੁੱਡ ਆਲਸੇ ਲੀਡ ਟੂ ਇੰਫਲੇਸ਼ਨ ਵਚਿ ਬੀਜੇਪੀ ਹੈਡ ਪ੍ਰੋਮੀਸਡ ਟੂ ਕੁਰਟੇਲ, ਦ ਆਮ ਆਦਮੀ ਪਾਰਟੀ ਲੀਡਰ ਸੈਂਡ ਦੈਟ ਦ ਸਰੋਮਈ ਅਕਾਲੀ ਦਲ ਅ ਪਾਰਟਨਰ ਇਨ ਦ ਬੀਜੇਪੀ ਗਵਰਨਮੈਂਟ ਮੱਸਟ ਆਸਕ ਦ ਨਰੇਂਦਰ ਮੋਦੀ ਗਵਰਨਮੈਂਟ ਟੂ ਬਰਗਿ ਡਾਊਨ ਦ ਡੀਜਲ ਐਂਡ ਪੇਟਰੋਲ ਪ੍ਰਾਈਸਜ਼</p>

source and target script in writing, difference in number of alphabets of source and target language, phonetic properties, and length of word. The accuracy of Hindi to Punjabi transliteration is very good as both are closely related languages. English to Punjabi transliteration system also provides promising result.

This system is evaluated on data that is taken from different news and overall accuracy of the system is very good. It also does accurate transliteration of some names of person and location. Proposed system can be further improved by adding more rules for unique and unknown words.

6. References

1. Bhalla D, Joshi N. Rule based transliteration scheme for English to Punjabi. *International Journal on Natural Language Computing*. 2013; 2(2):67–73.
2. Deep K, Goyal V. Development of a Punjabi to English Transliteration System. *International Journal of Computer Science and Communication*. 2011; 2(2):521–6.
3. Dhore M, Dixit S, Dhore R. Hindi and Marathi to English NE Transliteration Tool using Phonology and Stress Analysis. *Proceedings of the 24th International Conference on Computational Linguistic*; Mumbai, India. 2012. p. 111–8.
4. Gill H, Gleason H. *A Reference Grammar of Punjabi*. Patiala, India: Publication Bureau, Punjabi University; 1986.
5. Goyal V, Lehal G. Hindi-Punjabi Machine Transliteration System (For Machine Translation System). *George Ronchi Foundation Journal, Italy*. 2009; 64(1):1–7.
6. Josan G, Kaur J. Punjabi to Hindi Statistical Machine Transliteration. *International Journal of Information Technology and Knowledge Management*. 2011; 4(2):459–63.
7. Josan G, Lehal G. A Punjabi to Hindi Machine Transliteration System. *Computational Linguistics and Chinese Language Processing*. 2010; 15(2):77–102.
8. Joshi H, Bhatt A, Patel H. Transliterated Search using Syllabification Approach. *Forum for Information Retrieval Evaluation*; Delhi, India. 2013.
9. Kaur J, Josan G. Statistical Approach to Transliteration from English to Punjabi. *International Journal on Computer Science and Engineering*. 2011; 3(4):1518–27.
10. Knight K, Graehl J. Machine Transliteration. *The 35th annual meetings of the Association for Computational Linguistics*; Madrin, Spain. 1998; 128–35.
11. Lee J, Choi K. English to Korean Statistical Transliteration for Information Retrieval. *Computer Processing of Oriental Languages*. 1998; 12:17–37.
12. Malik M. Punjabi Machine Transliteration. *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*. Sydney. 2006. p. 1137–44.
13. Oh J, Choi K. An English-Korean Transliteration Model Using Pronunciation and Contextual Rules. *Proceedings of the 19th International Conference on Computational Linguistics*. Taipei, Taiwan. 2002. p. 758–64.
14. Rathod P, Dhore M. Hindi and Marathi to English Machine Transliteration using SVM. *International Journal on Natural Language Computing*. 2013; 2(4):55–71.
15. Saini T, Lehal G. Shahmukhi to Gurmukhi Transliteration System: A Corpus based Approach. *Research in Computing Science (Mexico)*. 2008; 33:151–62.
16. Vijaya MS, Ajith VP, Shivapratap G, Soman KP. English to Tamil Transliteration using WEKA. *International Journal of Recent Trends in Engineering*. 2009; 1(1):498–500.