Implementation of DJ Rule Based Algorithm for Dhuni-Vishleshan of Compound Punjabi Words

Deepjot Kaur*, Navjot Kaur
Department of Computer Science & Engineering
Sri Guru Granth Sahib World University
Fatehgarh Sahib, Punjab, India.

Abstract: Dhuni-vishleshan describes the process by which one word is broken into many words. It is the new software for Punjabi language. Various rules can be made and after that we can implement the words according to their rules. In Punjabi, words are a sequence of characters. There is a little amount of work is completed in this area. A word can be of two types-simple and compound. A simple word consists of roots. A compound word is also called as co-joined word can be broken up into two or more words. The problem to which this paper is concerned, is breaking up of Punjabi compound words into constituent words. In this paper, the rules for breaking the compound words into simple words have been applied. The problem of this paper is to break the compound word into constituent words with the help of rules of ‘dhuni-vishleshan’ in Punjabi.

Keywords: compound words, DJ rule based algorithm, dhuni-vishleshan, phonetics, gurmukhi script.

I. Introduction

A) Phonetics: Phonetics is the study of speech sounds of humans that appear in all human languages to represent the meanings. In phonetics we deal with different sounds neither letters. The task of phonetics is to provide brief of speech. Phonetics plays an important role in improving our communication. Alphabets and words are spelled correctly that is must [1]. Eg. A child cries and informs it mother that it is hungry. In this condition no language is used. For communication language can spoken or written. In this case sound matters. Mostly sounds are produced by air stream from lungs through any other speech organs. It is the root of the speech sounds.

B) History of Punjabi Language: Punjabi sometimes spelled Panjabi, belong to the Indic group of the Indo-European family of Languages. Punjabi is the tonal language. Tonal being that it differentiate the words by tones [2]. Punjabi language is used in both parts of Punjab in India and Pakistan. In India and Pakistan the written standard for Punjabi is known as Majhi that is called after the Majha region of Punjab. This script was created by Guru Angad Dev Ji. This language is the mother language of more than 100 million people of Pakistan, India, Canada and America. In India it is the official Language of Punjab state, and is additionally spoken within the neighboring states of Haryana and Himachal Pradesh. The Punjabi language is closely connected with the Sikh religion. Its alphabet, recognized as Gurmukhi, was the vehicle for recording the teachings of the Sikh gurus. It was invented by the second of the gurus within the 16th century. The word “Gurmukhi” means “Guru’s mouth.” Gurmukhi script is used for Punjabi language and it is the 11th widely spoken language in the world. Almost 100 million people speak different accent of this language as their first language.

1) Gurmukhi Consonants: The Gurmukhi script has thirty five akhar or consonants, a twin of the Punjabi alphabet as well as 3 vowel and thirty two consonants. Each character represents a phonetic sound. The alphabetical order of the Gurmukhi script area unit classified to make a grid of 5 horizontal and 7 vertical rows. Some characters have a nasal sound [3].

2) Gurmukhi Vowels: In Punjabi language letters are joined by a line at the top. In this there is no concept of upper and lower case letters. The gurmukhi script can be separate into three zones i.e. upper, middle and lower. There are ten vowels, three semi-vowels and three half-characters are used in Punjabi language [4]. In spoken language a vowel could be a sound that is pronounced with associate vocal tract such as teeth, lips, tongue. Vowels are the affecting class of sound in any language. They play a significant role within the pronunciation of any words.

II. Proposed Algorithm

Phonetics is the study of speech sounds of humans that appear in all human languages to represent the meanings. The work has been done in the area of English and similar languages. Punjabi is the 11th widely spoken language. There is the very little amount of work is completed in this field. Developing programs that understand a natural language is a difficult task. They contain an infinity of various sentences. The problem to which this paper is concerned, is breaking up of Punjabi compound...
words into constituent words. Eg. ਦਿਮਾਂ → ਦਿ + ਮਾ + ਅਨ. Sometimes a person cannot pronounce difficult word so it’s a easy way to pronounce by separating the words by applying several rules. Several rules can be made according to Punjabi laga, ਦਿ, ਦਾ, ਦਾਣ, ਦਾਸ, ਦਾਣੀ, ਦਾਸ਼, ਦੋਤ, ਦੋ ਅਨਦੇ ਅਤੇ Punjabi vowels are also used like ਅ, ਆ, ਇ, ਈ, ਉ, ਊ, ਏ, ਐ, ਓ, ਔ and Punjabi vowels are also used like ਅ, ਆ, ਇ, ਈ, ਉ, ਊ, ਏ, ਐ, ਓ, ਔ. On this various rules will be made and after that we can implement the words according to rules. Some examples of ‘dhunivisheshan’ of Punjabi Compound Words are:

Table I: Compound Words with their Outputs (Dhuni-vishleshan)

<table>
<thead>
<tr>
<th>Compound word</th>
<th>Dhuni-vishleshan</th>
</tr>
</thead>
<tbody>
<tr>
<td>ਮਹਿਤਵਪੂਰਕਾਂਤ</td>
<td>ਮਹਿਤਵਪੂਰਕਾਂਤ</td>
</tr>
<tr>
<td>ਪੇਤੇਕਸ਼ਣੀਵਾਂ</td>
<td>ਪੇਤੇਕਸ਼ਣੀਵਾਂ</td>
</tr>
<tr>
<td>ਲੁਕਾਕਾਰ</td>
<td>ਲੁਕਾਕਾਰ</td>
</tr>
</tbody>
</table>

Algorithm:- Dhuni-vishleshan is a recently developed software for Punjabi language. It is a application that is developed in .net. The algorithm used for the implementation of this module is the DJ Rule Based Algorithm.

**Step 1**: Load data from database.
**Step 2**: Select the word from the database or whether enter manually.
**Step 3**: Splitting the string into character by character.
**Step 4**: Now, comparing the characters:-

a) If character = ਦਾ
   Replace it with ਦਾ

b) Else If character = ਦਾਣ
   Replace it with ਦਾਣ

c) Else If character = ਦਾਸ
   Replace it with ਦਾਸ

d) Else If character = ਦਾਸ਼
   Replace it with ਦਾਸ਼

e) Else If character = ਦੋਤ
   Replace it with ਦੋਤ

f) Else If character = ਦੋਤ
   Replace it with ਦੋਤ

g) Else If character = ਦੋ ਅਤੇ
   Replace it with ਦੋ ਅਤੇ

h) Else If character = ਦੋ ਅਤੇ
   Replace it with ਦੋ ਅਤੇ

i) Else If character = ਦੋ ਅਤੇ
   Replace it with ਦੋ ਅਤੇ
j) Else If character = ൅
    Replace it with ਲ
k) Else If character = േ
    Replace it with ਲ
l) Else If character = െ
    Replace it with ਲ
m) Else If character = ൅ਐ
    Replace it with ਲ
go) Else If character = ൅ਉ
    Replace it with ਲ
h) Else character = ൅਼
    Replace it with ਲ

Step 5: Concatinate the final character for final output

III Experiment & Result

Dhuni-vishleshan is a recently developed software for Punjabi language. It is an application that is developed in .net. With in which the work is done on MS Excess at the back-end tool and front-end tool is .net. The algorithm used for the implementation of these module is the Rule Based Algorithm. Accuracy is the significant issue to be examined. So, to measure the accuracy of our algorithm we implement experiments on number of different words. Whenever the application is started, the window shown in figure 1 will appear which contains the text area, where the user can enter the text. In this, we can choose a word from the database or enter a word manually.

Fig. 1: Snapshot of main screen.

In the following snapshot Fig.2 shows the working of Dhuni-vishleshan. First the user will choose the word from database or whether enter manually. After that they will get the output on clicking button ਆਉਟ ਉਟ. Fig. 3 shows its word to sound rule.
In Fig. 3, the word is entered by the user manually.

If the user types the word "ਪੰਜਾਬੀ ਵੇਲਸ" manually, then it will show the output as: ਇਹ ਹਦ ਆ ਸਰਮਾਂ ਅਧਰਾਂ and it gives the correct output. In Fig. 4, the word is taken from the database by user.
If the user choose the word ਲਿਖਣ-ਸੂਚਿਕ from the database then it will show the output as: ਵਿਕਟ ਬਾਲ - ਮੂੰ ਟ ਦੇ ਆਲ and it gives correct output.

We perform testing on different words corresponding to our algorithm. To measure the accuracy of our algorithm. We perform testing on 24231 words. After testing we obtain 99.9% accuracy.

Results:- We have tested the system by first giving the input from database that contain approximately twenty four thousand words where our system has given no error. After that we have entered the words manually that also gives no error. So we can say that our system has good accuracy. Following is the part of the document.
In this work, we have develop the DJ Rule-Based algorithm on words according to their rules. With the help of this algorithm we have noted an accuracy of 99.9% depending upon the number of rules that are implemented. As future work, we can use the sound button for pronouncing the word and further implementation can be done on the line or paragraph also. This software can be beneficial for those people who are learning punjabi. With this software one can learn about the very important aspect of Punjabi Grammar i.e. ‘Dhunivishleshan’ is in an straightforward and interesting way that can give entirely new dimension that add new way to traditional approach to Punjabi Teaching. This can also be used to solve and test the problems related to Punjabi Grammar.

References

IV Conclusion

<table>
<thead>
<tr>
<th>Word</th>
<th>Output</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ਅਠਾਦਧਵ</td>
<td>ਅਠਾਦਧਵ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਠਾਦਧਵ</td>
<td>ਅਠਾਦਧਵ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਮਤੰਸਾਰ</td>
<td>ਮਤੰਸਾਰ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
<tr>
<td>ਅਈ</td>
<td>ਅਈ</td>
<td>Correct</td>
</tr>
</tbody>
</table>

© 2013, IJARCSSE All Rights Reserved


