

A RULE BASED PUNJABI DIALECT CONVERSION SYSTEM

Name of Student : Anterpreet Kaur (1411640)

Deptt. : CSE

Guide : Parminder Singh & Kamaldeep Kaur

Mode of Study : F. Time

ABSTRACT

At a recent time, informal form of language is surged dramatically on internet resources as web becomes a media for people to share their ideas and thoughts via personal blogs and social articles. In case of Punjabi language, there is deficiency of processing tools for Dialectal Punjabi as compared to Standard Punjabi language processing tools. Our work is the pioneer work which depicts the conversion of three dialects of Punjabi and illustrates the need and complications for Punjabi dialect language. In this thesis work, dialect conversion system has been developed for processing Punjabi dialects namely Majhi, Malwai and Doabi. A rule-based approach is used by proposed system to convert the Standard Punjabi input text into its equivalent Malwai dialect, Majhi dialect and Doabi dialects. To identify Standard Punjabi words in given input text, various methods are used and then selected identified words are converted using rule-based component which has two parts: bilingual dictionary and morphological transfer rules. The proposed conversion system has nine bilingual dictionaries that contains 5,558 words related to Standard Punjabi, Majhi dialect, Malwai dialect and Doabi dialect. These dictionaries are used only for word-to-word mapping but only direct mapping of source text into its equivalent target text is not adequate, so morphological conversion rules for conversion to Majhi, Malwai and Doabi dialects have been developed in thesis that replace specific portion of input word. The transfer rules are implemented on source sentences and then words are replaced as stated by rules. The results of proposed dialect conversion system depict the remarkable accuracy. The accuracy of Standard Punjabi to Majhi dialect, Standard Punjabi to Malwai dialect and Doabi dialect conversion is found to be 96.58%, 96.48% and 97.54% respectively.