Abstract

A part of speech (POS) tagging is one of the most well studied problem in the field of Natural Language Processing (NLP). A POS Tagger is the process of assigning correct tag like noun, adjective, verb, adverb etc to each word of the input sentence. Disambiguation rules and Tagset is vital parts of POS tagger. POS tagging is difficult for Marathi language due to unavailability of corpus for computational processing. In this paper, a POS Tagger for Marathi language using Rule based technique is presented. Our proposed system find root word using morphological analyzer and compare the root word with corpus to assign appropriate tag. If word has assigned more than one tags then by using grammar rules ambiguity is removed. Meaningful rules are provided to improve the performance of the system.

References

- Jyoti Singh Nisheeth Joshi Iti Mathur &quot;Development of Marathi Part of Speech Tagger Using Statistical Approach&quot;, Advances in Computing, Communications and
Informatics (ICACCI), 2013.
- Namrata Tapaswi Suresh Jain, &quot;Treebank Based Deep Grammar Acquisition and Part-Of-Speech Tagging for Sanskrit Sentences&quot; Software Engineering (CONSEG), 2012 CSI Sixth International Conference on.
- Sankaran Baskaran, Kalika Bali1, Tanmoy Bhattacharya, Pushpak Bhattacharyya, Monojit Choudhury, Girish Nath Jha, Rajendran S. 5, Saravanan K. 1, Sobha L. 6, and KVS Subbarao &quot;A Common Parts-of-Speech Tagset Framework for Indian Languages&quot;.

Index Terms
Computer Science Information Sciences

Keywords
Part of Speech (POS) Tagset Tokenizer Stemmer Morphological analyzer Disambiguation