Abstract

Part-of-speech tagging in Marathi language is a very complex task as Marathi is highly inflectional in nature & free word order language. In this paper we have demonstrated a rule-based Part-of-Speech tagger for Marathi Language. The hand-constructed rules that are learned from corpus and some manual addition after studying the grammar of Marathi language are added and that are used for developing the tagger. Disambiguation is done by analyzing
the linguistic feature of the word, its preceding word, its following word, etc. After testing the system with three data sets we got encouraging results. The accuracy of our system is of an average 78.82% after testing it on three different data sets.

References

- F. M. Hasan, N. UzZaman and M. Khan "Comparison of Different POS Tagging Techniques (n-gram, HMM and Brill's Tagger) for Bangla"; Proceedings of the
Part-of-Speech Tagger for Marathi Language using Limited Training Corpora

International Conference on Systems, Computing Sciences and Software Engineering (SCS2 06) of International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CIS2E 06), December 4 - 14, 2006.

- F. M. Hasan, N. UzZaman, and M. Khan, "Comparison of Unigram, Bigram, HMM and Brill’s POS Tagging Approaches for some South Asian Languages.", Proceedings of the Conference on Language and Technology (CLT07), Pakistan, August 7 - 11, 2007

Index Terms

Computer Science
Natural Language Processing

Keywords

Pos Tagger   Morphological Analysis   Rule-based.