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

- A. Bharati, V. Chaitanya and R. Sangal; Computational Linguistics in India: An Overview; Proceedings of the 38th Annual Meeting on Association for Computational Linguistics 2000, VOL 38; PART 1, PP 595-596
- B. N. Patnaik; Computational linguistics for Indian Languages; Symposium on Indian Morphology Phonology & Language Engineering IIT Kadaipur India March 19-21 2004 PP 55-57.
- C. Samuelsson and A. Voutilainen; Comparing a Linguistic and a Stochastic Tagger; Proceedings of the 35th Annual meeting of the ACL and 8th Conference of the European chapter of the ACL 1997 PP 246-253.
- Dinesh Kumar & Gurpreet Singh Josan; Part of Speech Tagger for Morphologically rich Indian languages: A Survey; International Journal of Computer Applications Vol. 6, No. 5, September 2010.
- 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
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


- S. Abney &quot;Part-of-Speech Tagging and Partial Parsing&quot;, Corpus-Based Methods in Language and Speech Processing 1996,

Index Terms

Computer Science  
Natural Language Processing

Keywords

Pos Tagger  Morphological Analysis  Rule-based.