Abstract

Part-Of-Speech (POS) tagging is defined as the Natural Language Processing (NLP) task in which each word in a sentence is labeled with a tag indicating its appropriate part of speech. Of the entire supervised machine learning classification algorithms, second order Hidden Markov Model (HMM) and Conditional Random Fields (CRF) is chosen in this work for POS tagging of Kannada language. Training data includes 51,269 words and test data consists of around 2932 tokens. Both set being disjoint and taken from EMILLE corpus. Experiments show that the accuracy of the tools based on HMM and CRF is 79.9% and 84.58% respectively.

References

- Gimenez, J. and L. Marquez, 2003. Fast and Accurate Part-of-Speech Tagging: The...
SVM Approach Revisited. In Proceedings of the Fourth RANLP.
- Proceedings of IJCAI- 2007, Workshop on Shallow Parsing for South Asian Languages (SPSAL-2007), Hyderabad, India
- F. Sha and F. Pereira. 2003 Shallow parsing with conditional random fields. In
Kannada Part-Of-Speech Tagging with Probabilistic Classifiers

Proceedings of HLT-NAACL.

Index Terms

Computer Science
Artificial Intelligence

Keywords
Natural Language Processing Part Of Speech Tagging Hidden Markov Model Conditional Random Fields