Abstract

We report our dependency parsing experiments on two Indian Languages, Telugu and Hindi. We first explore two most popular dependency parsers namely, Malt parser and MST parser. Considering pros of both these parsers, we develop a hybrid approach combining the output of these two parsers in an intuitive manner. For Hindi, we report our results on test data provided in the for gold standard track of Hindi Shared Task on Parsing at workshop on Machine Translation and parsing in Indian Languages, Coling 2012. Our system secured unlabeled attachment score of 95.2% and labelled attachment score 90.7%. For Telugu, we report our results on test data provided in the ICON 2010 Tools Contest on Indian Languages Dependency Parsing. Our system secured unlabeled attachment score of 92.0% and labelled attachment score 69.5%.

References

- Sabine Buchholz and Erwin Marsi. 2006. CoNLL-X shared task on multilingual dependency parsing. In Tenth Conf. on Computational Natural Language Learning (CoNLL).
- Ryan McDonald, Kevin Lerman, and Fernando Pereira. 2006. Multilingual dependency analysis with a two-stage discriminative parser. In Tenth Conference on Computational Natural Language Learning (CoNLL-X), pp. 216–220.
Improving Indian Language Dependency Parsing by Combining Transition-based and Graph-based Parsing


Index Terms

Computer Science
Programming Language

Keywords

Dependency Parsing; Telugu; Hindi; Malt Parser; MST Parser