Abstract

In this paper, we present evaluation of URDU.KON-TB in the dependency parsing domain. The URDU.KON-TB treebank is developed on the bases of the phrase structure and hyper dependency structure which are only functional constituent's label. Treebank was annotated with three levels of annotation tagset, the semi-semantic POS (SSP), semi-semantic Syntactic (SSS) and Functional (F) tagset and was checked for the Phrase Structure Parsing domain. To evaluate this treebank in the Dependency Parsing domain we have selected MaltParser. To use data in the parser, we have converted the URDU.KON-TB treebank annotated data according to the CONLL format. The compatibility of data to CoNLL is also measured along with usability of data in the dependency parsing domain. To make the data compatible, few assumptions are taken. The converted data is used to evaluate the system by dividing 80% data as training data and 20% data as testing data. We have performed eight experiments. Four experiments are conducted with six different feature models with converted data. The experiments results show URDU.KON-TB treebank is not suitable for the dependency parsing as dependency relation because Head information was missing in the treebank. We then performed four experiments.
with an assumption based enhancement by adding Head information. The algorithm used to
train and test data is Nivre arc-agear algorithm. The new experiments show this treebank data
can be used to develop new dependency treebank for Urdu.

References

1. Abbas, Q. (2014). Building Computational Resources: The URDU. KON-TB Treebank and
   the Urdu Parser (Doctoral dissertation).
   Proceedings of Conference on Language and Technology (CLT10), SNLP, Lahore, Pakistan.
   Sciences thesis, Department of Computer Sciences, National University of Computer and
   Emerging (NUCES), Lahore, Pakistan.
   evaluation. In Proceedings of the Sixth Linguistic Annotation Workshop (pp. 157-165).
   Association for Computational Linguistics.
   semantic features make all the difference in parsing accuracy. Proc. of ICON, 8.
   Optimization. In LREC (pp. 2757-2763).
    MaltParser: A language-independent system for data-driven dependency parsing. Natural
    Language Engineering, 13(02), 95-135.
    Conference on Empirical Methods in Natural Language Processing and Computational Natural
    Language Learning (pp. 1455-1465). Association for Computational Linguistics.
12. Spreyer, K., & Kuhn, J. (2009, June). Data-driven dependency parsing of new languages
    using incomplete and noisy training data. In Proceedings of the Thirteenth Conference on
    Computational Natural Language Learning (pp. 12-20). Association for Computational
    Linguistics.
    morphosyntactic features in Hindi dependency parsing. In Proceedings of the NAAACL HLT 2010
    First Workshop on Statistical Parsing of Morphologically-Rich Languages (pp. 94-102).
    Association for Computational Linguistics.
    Parcing.
    Processing (pp. 16-27). Springer Berlin Heidelberg.
16. Abbas, Q. 2015, Morphologically rich Urdu grammar parsing using Earley algorithm,
    Natural Language Engineering (NLE), Vol.21(2), PP.1-36, ISSN: 1351-3249, DOI:
10.1017/S1351324915000133, Cambridge University Press, UK


**Index Terms**

Computer Science

Information Systems

**Keywords**

Phrase structure parsing, Data Driven Dependency Parsing, MaltParser