

{tag} International Journal of Computer Applications
Foundation of Computer Science (FCS), NY, USA

[Volume 167](#)

-
[Number 12](#)

Year of Publication: 2017

Authors:

Saima Munir, Qaisar Abbas, Bushra Jamil

10.5120/ijca2017914492

{bibtex}2017914492.bib{/bibtex}

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).
2. Ali, W., & Hussain, S. (2010). Urdu dependency parser: a data-driven approach. In Proceedings of Conference on Language and Technology (CLT10), SNLP, Lahore, Pakistan.
3. Ali, W. (2010). Data-Driven Dependency Parsing for Urdu, MS (MPhil), Computer Sciences thesis, Department of Computer Sciences, National University of Computer and Emerging (NUCES), Lahore, Pakistan.
4. Bhat, R. A., Jain, S., & Sharma, D. M. (2012). Experiments on dependency parsing of Urdu. Proceedings of TLT11, 31-36.
5. Bhat, R. A., & Sharma, D. M. (2012, July). A dependency treebank of Urdu and its evaluation. In Proceedings of the Sixth Linguistic Annotation Workshop (pp. 157-165). Association for Computational Linguistics.
6. Abbas, Q. (2014). Semi-semantic part of speech annotation and evaluation. LAW VIII, 75.
7. Nivre, J., Hall, J., & Nilsson, J. (2006, May). Maltparser: A data-driven parser-generator for dependency parsing. In Proceedings of LREC (Vol. 6, pp. 2216-2219).
8. Bharati, A., Husain, S., Ambati, B., Jain, S., Sharma, D., & Sangal, R. (2008). Two semantic features make all the difference in parsing accuracy. Proc. of ICON, 8.
9. Ballesteros, M., & Nivre, J. (2012, May). MaltOptimizer: A System for MaltParser Optimization. In LREC (pp. 2757-2763).
10. Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., ... & Marsi, E. (2007). MaltParser: A language-independent system for data-driven dependency parsing. Natural Language Engineering, 13(02), 95-135.
11. Bohnet, B., & Nivre, J. (2012, July). A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In Proceedings of the 2012 Joint 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.
13. Ambati, B. R., Husain, S., Nivre, J., & Sangal, R. (2010, June). On the role of morphosyntactic features in Hindi dependency parsing. In Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages (pp. 94-102). Association for Computational Linguistics.
14. Nilsson, J. (2009). Transformation and Combination in Data-Driven Dependency Parsing.
15. Nivre, J. (2008). Sorting out dependency parsing. In Advances in Natural Language 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

17. N. Chomsky. Three Models For The Description Of Language. Information Theory, IRE Transactions on, 2(3):113–124, 1956.

18. PUNEETH, K. (2016). Dependency Parsing and Empty Category Detection in Hindi Language (Doctoral dissertation, International Institute of Information Technology Hyderabad).

19. GADE, R. P. (2014). Dependency parsing approaches for Indian Languages: Hindi and Sanskrit (Doctoral dissertation, International Institute of Information Technology Hyderabad).

20. J. Nivre, Inductive Dependency Parsing, Springer, 2006.

21. M. Marcus, B. Santorini, and M.A. Marcinkiewicz, "Building a large annotated corpus of English: The Penn Treebank", Computational Linguistics 1993

Index Terms

Computer Science

Information Systems

Keywords

Phrase structure parsing, Data Driven Dependency Parsing, MaltParser