Survey of Named Entity Recognition Techniques for Various Indian Regional Languages

Abstract

Named entity recognition is a process and study of identification of entities that are proper nouns and classifying them to their appropriate pre-defined class, also called as tag. Named entity recognition is also called as entity chunking, entity identification and entity extraction. It is a sub task of information extraction, where structured text is extracted from unstructured text. Popular applications of NER are machine translation, text mining, data classification, question answering system. This paper presents a survey of different NERC techniques, approach, observations and features for Indian regional languages such as Hindi, Marathi, Bengali, Punjabi, Malayalam, Bengali, Kannada, Telugu, Tamil, Urdu, Oriya.

References

Survey of Named Entity Recognition Techniques for Various Indian Regional Languages

28. B. Sasidhar, P. M. Yohan, Dr. A. Vinaya Babu and Dr. A. Govardhan, “Named Entity Recognition in Telugu Language using Language Dependent Features and Rule based Approach “, 2008

**Index Terms**

Computer Science  
Pattern Recognition

**Keywords**

Named entity recognition, Named Entity, Machine learning techniques, Statistical Model, NERC