Abstract

Named Entity Recognition (NER) is used to locate and classify atomic elements in text into predetermined classes such as the names of persons, organizations, locations, concepts etc. NER is used in many applications like text summarization, text classification, question answering and machine translation systems etc. For English a lot of work has already done in
field of NER, where capitalization is a major clue for rules, whereas Indian Languages do not have such feature. This makes the task difficult for Indian languages. This paper explains the Named Entity Recognition System for Punjabi language text summarization. A Condition based approach has been used for developing NER system for Punjabi language. Various rules have been developed like prefix rule, suffix rule, propername rule, middlename rule and lastname rule. For implementing NER, various resources in Punjabi, have been developed like a list of prefix names, a list of suffix names, a list of proper names, middle names and last names. The Precision, Recall and F-Score for condition based NER approach are 89.32%, 83.4% and 86.25% respectively.

Reference

- Takeuchi K. and Collier N. 2002 Use of Support Vector Machines in extended named entity Recognition. In the proceedings of the sixth Conference on Natural Language Learning (CoNLL-02), Taipei, Taiwan, China.
- Srihari R., Niu C. and Li W. 2000 A Hybrid Approach for Named Entity and Sub-Type Tagging. In the proceedings of the sixth Conference on Applied Natural Language Processing.
- Amandeep Kaur, Gurpreet Singh Josan and Jagroop Kaur. 2009 Named
- Praneeth M Shishtla, Karthik, Prasad Pingali and Vasudeva Verma
- Karthik Gali, Harshit Surana, Ashwini Vaidya, Praneeth Shishtla and Dipti Misra Sharma.2008 Aggregating Macine Learning and Rule Based Heuristics for NER. In the Proceedings of the IJCNLP-08 workshop on NER for South and South East Asian Languages (pp 25-32). Hyderabad, India.

Index Terms
Computer Science Natural Language Processing

Key words
Punjabi language Named Entity Recognition (NER) Condition based approach
Punjabi proper nouns identification
Named Entities (NE) etc.