Named Entity Recognition (NER) seeks to locate and classify atomic elements in text into predefined categories such as names of person, organization, location, Quantities, Percentage etc. Named entities tell us the roles of each meaning bearing word in a sentence and hence identification of these entities certainly helps us to extract the essence of the text which is very
important in Question Answering (QA), Information Extraction (IE) and Summarization. The system presented here is a Named Entity (NE) Classifier created using Multiclass Support Vector Machines based on linguistic grammar principles. Malayalam NER is a difficult task as each word of named entity has no specific feature such as Capitalization feature in English. NERs in other languages are not suitable for Malayalam language since its morphology, syntax and lexical semantics is different from them. Also there is no tagged corpus available for training. For testing this system, documents from well known Malayalam news papers and magazines containing passages from five different fields such as sports, health, politics, science and agriculture are selected. Experimental results show that the average precision recall and F-measure values are 89.12%, 89.15% and 89.13% respectively.

Reference

- Burr Settles,” Biomedical Named Entity Recognition Using Conditional Random Fields

Index Terms

- Computer Science
- Natural Language Processing

Key words

- Named Entity Recognition
- Parts-of-Speech Tagger
- Phrase chunker
- Compound word splitter