Evaluation of Punjabi Named Entity Recognition using Context Word Feature

Abstract

Named Entity Recognition is the task of identifying and classifying Named Entities in the given text. In this paper evaluation of Named Entity Recognition in Punjabi language has been performed using context word feature. Words preceding and succeeding the target word are very helpful in determining its category. In this work context word feature of word window 7, 5 and 3 have been used. Experiments have been performed using different training and test sets. In this evaluation a Named Entity Tagset of 14 tags namely PERSON, ORGANIZATION, LOCATION, FACILITY, EVENT, RELATIONSHIP, TIME, DATE, DESIGNATION, TITLE-PERSON, NUMBER, MEASURE, ABBREVIATION and ARTIFACT has been used. It has been observed that word window 7 and 5 have given better results as compared to word window 3. Although F-scores and Precision values of word window 7 are slightly higher than that of word window 5 but recall of word window 7 was found to be lower than that word window 5.

References


Index Terms

Computer Science

Text Processing

Keywords

Named Entities  Named Entity Recognition  Punjabi Language  Context word feature