Abstract

The paper describes about the development of a Named Entity Recognition (NER) system for Geological text using Conditional Random Fields (CRFs). The system makes use of the different contextual information of the words along with the variety of features that are helpful in predicting the various named entity (NE) classes. The NE tagged geological corpus was developed from the collection of scientific reports and articles on the geology of the Indian subcontinent has been used to build up the system. The training set consists of more than 2 lakh words and has been manually annotated with a NE tag set of seventeen tags. The system is able to recognize 17 classes of NEs with 75.8% F-measure.

Reference


Index Terms
Computer Science Information Systems

Key words
Geological Corpus Named Entity Recognition
Precision Recall F-measure
Geographic references