Chemical Name Entity Recogintion

Preprocessing :

1. Sentence and word Tokenization
2. Lemmatization
3. Stop Word Removal
4. Generation of N-grams

Features Used:

1.) Lemma and POS tag of the word.
2.) Digit Count
3.) Upper and lower case counts
4.) Punctuation Characters
5.) Case Pattern Features
6.) Character N-grams

Model Used:

CRF++ software was utilized for training CRF models.

We trained the model with 90% data and tested the model with the remaining data.

We got an accuracy of 66.67%.