

Chapter 7

7 CONCLUSION

Overall, conclusions drawn from the above experiments with regard to the research questions are as follows:

We studied different approaches of Named entity recognition and their problems. We focused on collection huge gazetteer classes. Our results show more accurate and perfect recognition of Named Entities. Depending upon the collection of data our system will trace any named entities of almost all inflectional languages. We got success in tracking multiple proper and common noun words in a sequence up to five words in a row (using five gram technique).

Prefixes , root and suffixes (The Morphological features) are very good pointers of named entity detection type, especially in tasks where, the category of a named entity (PERSON, LOC, ORG, and so on) has to be anticipated, but also when the limitations of a named entity need to be predicted. Orthographic features that represent the capitalization characteristics of a word are especially useful in predicting these limitations. Seed-list features are also very helpful in Named Entity Recognition (NER) detection, but the actual improvement is highly dependent on the classification algorithm used. Like morphological features, seed-list information has the greatest effect on tasks where the type of a NE has to be determined. Boundary identification does not benefit from the use of gazetteers. Using the stacked predictions of another level classifier as features also improves generalization performance. The feature set that gave the best performance after the experiment rounds contained all of the 100 initial features. The suboptimal feature set of the two-step ceiling performance contained the following 22 features: five windowed words and their POS tags, suffix FOCUS, contains Punctuation FOCUS, prefix FOCUS, prefix R1, four seed-list features, and four features representing the windowed predictions of the first level classifier.