NERHMM: A Tool For Named Entity Recognition Based On Hidden Markov Model

Sudha Morwal and Deepti Chopra

[1]Department of Computer Engineering, Banasthali Vidyapith, Jaipur (Raj.), INDIA

sudha_morwal@yahoo.co.in
deeptichopra11@yahoo.co.in

ABSTRACT

Named Entity Recognition (NER) is considered as one of the key task in the field of Information Retrieval. NER is the method of recognizing Named Entities (NEs) in a corpus and then organizing these NEs into diverse classes of NEs e.g. Name of Location, Person, Organization, Quantity, Time, Percentage etc. Today, there is a great need to develop a tool for NER, since the existing tools are of limited scope. In this paper, we would discuss the functionality and features of our tool of NER with some experimental results.

KEYWORDS

HMM, NER, F-Measure, Accuracy, NEs

1. INTRODUCTION

Named Entity Recognition is the process that involves finding the NEs in a corpus and then be able to distinguish them into various classes of NEs such as person, location, organization, time, river, sport, vehicle, country, state, quantity, number, time etc. The various applications of NER are: Question Answering, Information Extraction, Automatic Summarization, Machine Translation, Information Retrieval etc. [8][11]

There are many challenges that have to be dealt with while performing Named Entity Recognition in Indian languages. Indian languages lack in proper resources, so before performing Named Entity Recognition in Indian languages, we have to carry out the task of Corpus development which include doing annotation on the raw text, preparing Gazetteer etc. Indian languages are free word order, inflectional and morphologically rich in nature. In Indian languages, there are numerous named entities that also exist as common nouns in the dictionary.

2. RELATED WORK

Natural language toolkit (NLTK) is a free and open source computational linguistic tool. Apart from Named Entity Recognition, this tool can be used for performing tokenization, classification, stemming, parsing, tagging etc. NLTK provides support in carrying out research in areas like linguistic, artificial intelligence, machine learning, information retrieval etc.[21]

Scikit-learn also known as Scikits.learn is an open source machine learning tool. It efficiently implements the algorithms of Hidden Markov Model. [22]