INTRODUCTION

The Natural Language Processing (NLP) is a challenging and interesting area of machine learning, where the task of training our system to learn the semantic analysis of individual words of any language. The language may be anything used to express the emotions or feeling of every person. But the emotions or feeling of a person may be same, but the type of language used to represent the feelings or emotions is different. The Language is different but the feelings are same. Thus, we need a system, which understands the emotions or feelings expressed in different languages by understanding the meanings of expression in a particular context.

INTRODUCTION

The Named Entities shall be recognized in various other languages like English, which is an international language and one regional language Telugu. English Named Entity Recognition

There are many research papers which has addressed the Named Entities in English like describes a Parts of Speech based method of identifying the Named Entities in English, Another et.al has addressed the problem of identifying the Named Entities in English. We understood the ideas of these methods and incorporated with a new ideology called Rule based approach. Et.al proposed a new way of identifying an entity in English Text documents like Names of a person, Place, Organization, but the feelings are same. Thus, we proposed a new method of identifying the Named Entities in Regional Language Kannada.
Telugu Named Entity Recognition
Another regional language has been used addressed with the named Entity Recognition called Telugu. The Telugu named entity recognition has been addressed by et.al. The et.al has given their significant contribution to Telugu language to address the problem of identifying the names in Telugu language. This et.al uses the method of Amarappa et.al to determine the Kannada Named Entities, which focuses on only names of a person. Thus, understanding the need of recognizing the named entities in regional language Kannada, we proposed new method called “Rule based- Kannada Named Entity Recognition”.

DATA OF KANNADA TEXT
The Kannada text documents collected from the government organization Central Institute of Indian Languages is a benchmark Kannada Text document. The Kannada Text document consists of data items in different files numbered 101.txt, 102.txt, 103.txt, 104.txt, 105.txt and 106.txt. All these text documents consist of Kannada textual information, whose information can be observed from the fig.1 and fig.2. The Kannada text document shown in fig.1 and fig.2 indicates that the document consisting of different named entities like Name of a person, Place, Designation of a person Recognitions awarded to a person and Organization, where the person is working. So we considered a data, which exhibits Kannada textual information based on these benchmark information contained in it. Further, the data file consists of information of various other entities, which shall be observed from it. But the objective of our proposed method is restricted to recognize only few entities like Name of a person, place, Organization, Designation and Recognition of awards. The other named entities can be recognized from any of the documents of Central Institute of Indian Languages (CIIL).

PROPOSED RULE BASED-KANNADA NAMED ENTITY RECOGNITION
The Rule based approach has been applied to recognize Kannada Named Entities like Name of a Person, Place, Designation, Organization and other related Kannada Named Entities of Natural Language represented in the form of Kannada Text.

The proposed method makes use of several rules formulated and designed to recognize Kannada Named Entities. The rules include many formulations, while extracting the Kannada named entities has been incorporated into the proposed method. The proposed method focuses on semantic analysis of the Kannada text while recognizing the Kannada Named Entities. The semantic analysis of the Kannada Text has been carried out by following the Rule Based Kannada Named Entity Recognition (R-KNER).

The sequence of steps followed to train the system, while understanding the meanings of individual words in such a manner that the semantics shall be learned by dividing the raw textual information into different sentences. Thereby reducing the complexity of understanding the semantic analysis of the Kannada textual information. Further, the sentences are divided into individual words, which eventually helps in applying the rules to annotate different words of a sentence and their
meanings associated with every sentence of raw Kannada Textual Information.

The proposed R-KNER has been designed to understand the individual words of a Kannada text by applying the rules, which were formulated. The features of individual annotated words have been extracted by training our system as such it understands the Kannada words and their meanings in a particular context. The R-KNER makes use of several rules, which is shown in the table.

**Algorithm R-KNER**

Input: The raw Kannada Text is given to a trained system

Description: The trained system understands and recognizes the Kannada Named Entities

Output: The prediction of Kannada Named Words into different classes of words.

Step 1: [Partitioning]
Divide the raw Kannada text into different sentences.

Step 2: [Partitioning]
Divide each sentence into different words

Step 3: [Annotation]
Annotate the individual text into different possible set of values.

Step 3.1: [Rule based Feature Extraction]
Annotate individual words

Step 3.2: Apply set of rules over an annotated words

Step 3.3: Apply feature extraction

Step 3.4: SVM Classification to determine names of different entities.

Step 3.5: Recognize Named Entities.

Step 4: [Output]
Display the recognized named entities.

End

The feature extraction is one of the significant steps of our proposed R-KNER. As we required to recognize the meaning of individual Kannada words, we used Rule Based feature extraction technique. So that the classification of words into different Kannada Named Entities are possible. The proposed makes use of multi-class support vector machine (SVM) classification algorithm to classify the individual Kannada words into different classes. The classes considered for classification of Kannada words into different classes include

- Name of a Person
- Place
- Designation
- Organization

The above four classes of SVM consists of many words, which fall into it based on the features extracted from our Rule Based feature extraction technique. The SVM Classifier has been used with these 4 classes of classifier. These 4 classes of words are used to calculate the True Positive Rate (TPR) against False Positive Rate (FPR).

The TPR is calculated by,

\[
TPR = \frac{\text{Predicted result}}{\text{Total number of words}}
\]

(1)

\[
FPR = \frac{\text{Incorrect Prediction}}{\text{Total number of words}}
\]

(2)

**CASE STUDIES**

The case studies have been performed to test the proposed R-KNER over a dataset consisting of different Kannada Named Entities. The case study 1 performs the task of finding the designation of a person by applying the proposed R-KNER method. Similarly, case study 2 performs the task of finding the Names of a person, which shall be observed from fig. 2 and fig.3.

**Case Study 1**

We utilized the features of font gubbi to extract the information of Kannada text and returned the details of it in a format, which is suitable for predicting the results of classification. The Kannada font Gubbi is used to determine the Kannada Named Entities. So that all information of a Kannada text consisting of Names of a person, Designation of a person, places shall be found by using the Rule Based feature extraction technique. The results of the proposed method over a dataset provided by Central Institute of Indian Languages (CIIL) shall be observed from the generated output file. The results of it are shown in fig.2.

**Case Study 2**

The proposed method has been used to determine the names of a person from the Kannada text documents provided by CIIL, the result of applying the proposed method over a data set consisting of names of a person shall be seen in fig.3.

**RESULTS AND ANALYSIS**

The proposed R-KNER has yielded good results over a dataset and the performance of producing the results of the proposed method shall be seen in the tabulated results. The accuracy of the proposed method has shown its impact and goodness of extracting the Kannada Named Entities.

**Results**

R-KNER has produced the output of recognizing the Kannada characters and the semantic analysis of the relation of each Kannada words with respect to other Kannada words in a sentence. The accuracy of identifying the Kannada Named Entities shown in tabulated results.

**Table 3: The result of comparison of proposed R-KNER over other contemporary methods**

<table>
<thead>
<tr>
<th>Method</th>
<th>TPR</th>
<th>FPR</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nymble</td>
<td>75.46</td>
<td>24.54</td>
<td>75.46</td>
</tr>
<tr>
<td>MENE</td>
<td>56.51</td>
<td>43.49</td>
<td>56.51</td>
</tr>
<tr>
<td>Association Rule Mining</td>
<td>66.36</td>
<td>33.64</td>
<td>66.36</td>
</tr>
</tbody>
</table>

The proposed method has yielded a good result over a dataset with 89.32% in finding the Kannada Named Entities, which shall be seen in Results and Analysis section.
### RULE BASED KANNADA NAMED ENTITY RECOGNITION

<table>
<thead>
<tr>
<th>Maximum Entropy</th>
<th>67.54</th>
<th>32.46</th>
<th>67.54</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed R-KNER</td>
<td>89.32</td>
<td>10.68</td>
<td>89.32</td>
</tr>
</tbody>
</table>

**Fig. 4:** The result of classification of Multi-class SVM in determining the Kannada Named Entities of dataset provided by CIIL.

**Analysis of Proposed method over other Existing methods of different language**

The accuracy of the proposed method over a dataset provided by CIIL for recognition of Kannada Named Entities shall be seen in table 3, below.

The proposed R-KNER has yielded good results of accuracy of 89.32% in finding the Kannada Named Entities correctly, while the rest 10.68% of false positive has been observed from the proposed method.

**CONCLUSION**

The proposed R-KNER has yielded good results over a dataset and the results of the proposed method have shown its significance in producing the results. Thus, we state that proposed R-KNER has contributed the task of recognizing the Named Entities in Kannada. The results of the proposed method shall be used as a reference point of research, while recognizing the meaning of the Kannada Named Entities.

**REFERENCES**