

Review Article

RULE BASED KANNADA NAMED ENTITY RECOGNITION

M. Pushpalatha¹, Dr. Antony Selvadoss Thanamani²

¹Maharani's Science College for Women, Mysuru, India. pushpaharish78@gmail.com

²Professor and HOD, Department of Computer Science NGM College of Arts and Science, Pollachi, Bharathiar University, Coimbatore India. selvadoss@gmail.com

Received: 09.12.2019

Revised: 23.01.2020

Accepted: 10.02.2020

Abstract

The Objective of this research is to determine the names of individuals in Kannada text documents. The Kannada text document includes various Named Entities like Name of a person, Place, Organization, Designation and various other entities shall be determined from it. The method divides the collection of Kannada sentences into different sentences in the pre-processing stage, then dividing the sentence into different words in the second stage of pre-processing. Further, the words are annotated and then the rule based feature extraction is done to recognize the Kannada Named Entities. The multi-class SVM classification is employed to determine the Kannada Named Entities. The Rule based approach focuses on identifying the names of a person by applying the rules formulated, place, and designation of a person has been identified by recognizing the names then the word prior to the names has been considered as a designation of a person. The Rule based Kannada Named Entity Recognition recognizes the various words of the Kannada text document. It has yielded a good results of 89.32% from a proposed Rule based Kannada Named Entity Recognition.

Keywords: Pre-processing, Rule based Approach, Multi-Class SVM Classification, Named Entity Recognition.

© 2019 by Advance Scientific Research. This is an open-access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.31838/jcr.07.04.47>

INTRODUCTION

The Natural Language Processing (NLP) is a challenging and interesting areas of machine learning, where the task of training our system to learn the semantic analysis of individual words of any of the 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 indifferent languages by understanding the meanings of expression in a particular context.

The language used by every individual may be different, but the feelings or expressions are same. Based on this information, we trained a system to understand the meanings of different Natural Languages. One such language considered for understanding the context of it is Kannada Named Entity. The Kannada Named Entity Recognition is quite a challenging and tedious task, while training our system, as we face many challenges and difficulties in understanding the various vottaksharagalu of Kannada text and Dheerga of Kannada text documents.

The Kannada text documents consisting of all these entities have been observed in the dataset of CIIL. The dataset having the names like names of a person has been identified by using the specific rules mentioned in table.1. Where the name of a person is recognized by incorporating the rules like Dr. followed by any word is a name of a person, Prof. followed by any word is a name of a person, Further, the Titles like Dr. and Prof. are called as titles of a person, which shall also be called as a designation of a person. Similarly, in various other context of a Kannada textual representation, we recognized the organization, where the person is working or the names identified as an organization. Likewise, the place has been identified in a Kannada text document by applying various other rules formulated in Kannada. The rules formulated to identify these Kannada Named Entities shall be observed in Table.1. The Table.1 depicts the set of rules used in our proposed method. The Kannada Named Entities can be recognized by analyzing the text document into

various segments like sentences, words in preprocessing stage of our algorithm.

The Named Entities shall be recognized in various other methods, but the rule based method is another important method gives a very good result of understanding the semantic vision of any language in terms of training the systems with rules. So that the Kannada Named Entities shall be recognized and identified with it.

The structure of an entire research paper can be visualized in terms of various sections like Section 3. Focuses on proposed method, which has been incorporated with several Rules of recognizing the Kannada Named Entities. Section 4. Describes the case studies considered for recognizing the Kannada Named Entities like Name of a person and Location. Section 5 focuses on results of the proposed method and the analysis of the proposed method. Section 6 presents the conclusion of the overall objective of the research paper.

RELATED WORK

The proposed method has been built with an objective of recognizing the Kannada Named Entities. The section focuses on Named Entities of different 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 rule based method was not proposed by any of the research methodologies. Thus, we proposed a new method of identifying the Named Entities in Regional Language Kannada.

Telegu 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 (CILL).

<dances1><Aesthetics><Dance><1985><Book><ನೃತ್ಯ ಶಾಸ್ತ್ರ ಪ್ರಯೋಗ><ಡಿ.ಜನಾರ್ದನ ಶರ್ಮ (ಕಾಮತ್)><38>

Page 9

"ನೃತ್ಯ" ಎಂದಾಗ - ಅದರ ಭಾವವೆ "ಕುಣಿತ". "ಕಲೆ" ಎಂದಾಗ ಇದೊಂದು ವಿದ್ಯೆ, ನೃತ್ಯಕಲೆಯೆಂದು ಮೂರ್ಧುರ್ಯದ ಅಧ್ಯತೆಯಿಂದ ಇದನ್ನು "ಲಲಿತಕಲೆ" ಎಂದೂ ಕರೆಯುವುದುಂಟು. ಶಾಸ್ತ್ರಸಂಪನ್ನತೆಯ ದೃಷ್ಟಿ ವಿದ್ಯಾರ್ಥಿ ಮಾರ್ಗ. ಅಭ್ಯಾಸಾರ್ಥವಾದ ಹಾಗೂ ವಿಚಾರಾರ್ಥವಾದ ಕಲಾತ್ಮಕ ವಿದ್ಯೆ, ನಾಡ ಕಲಾಪರಂಪರೆಯ ಪುಣಿವಿಧಿ. ಭಾರತೀಯ ಸಂಸ್ಕೃತಿಯ ದೃಷ್ಟಿಯಿಂದ ನೋಡಿದರೆ ಇದೊಂದು ಪೌರಾಣಿಕ - ಜಾತಿತಿಕ - ಸಾಮ ಹೆಗ್ಗುರುತು. ಈ ಕಲೆಯ ಹಿನ್ನೆಲೆಯನ್ನು ಗಮನಿಸಿದರೆ; ಧರ್ಮ, ರೀತಿ - ನೀತಿಯನ್ನು ಉಪದೇಶಿಸುವ ನಾಡ - ನುಡಿಯನ್ನು, ಪ್ರತಿಬಿಂಬಿಸುವ ಪ್ರಾಚೀನ ಪರಂಪರೆಯ ಸ್ಮಾರಕ, ಜೀವನದ ದೃಷ್ಟಿಯಿಂದ ನೋಡಿದರೆ ಬದುಕಿಗೆ ತ ನೃತ್ಯ. ನೃತ್ಯ-ನಾಟಕ ಎಂದಾಗ ಬರುವ "ಕುಣಿತ" ಎಂಬ ಭಾವನೆಯು ನಿಜವಾಗಿಯೂ ನರ್ತನದ ಉಗ್ರತೆ, ಕುಣಿತವು - ಒಂದು ಆಕಾರ, ಕೊನ, ಬಾಗುವಿಕೆ, ನಿಲುವು, ನಿರ್ದಿಷ್ಟ - ನಿರ್ದಿಷ್ಟ ನಿಯಮಾದಿಗಳ ಸಂಭವನೆಯಿಂದ ಶ ಹಸ್ತಮುದ್ರಾಧಿ ಭಾವಭಂಗಗಳ, ಅಂಗವಿನ್ಯಾಸಾದಿಗಳ ಜಲುವಿನಿಂದ, ಯಂತ್ರರಂಗಾದಿ ದಾಗ ರಸಸಂಚಯದಿಂದ, ಹಾವ-ಭಾವ, ಲೀಲಾ, ವಿಲಾಸ, ವಿಜೃಂಭಿ ವಿದ್ಯಮಗಳ ಅಂತ್ಯವಾಗುವುದು, ಕಲ್ಪನಾ ಸಾಮಾನ್ಯವಾಗಿದ್ದು, ಹೊಸವ

Page 10

ಬಲೆಯಲ್ಲಿ, ಸೂಯ್ಯ ಸುಂದರ ಕನಸು ನೆನಸುಗಳಿಂದ, ಸಾಹಿತ್ಯ ಸುಧೆಯ ಹೊನಲಿನಲ್ಲಿ, ಹೊಮ್ಮುವ ವಿಚಾರದ ರೂಪದರ್ಶನತ್ರದ ಸತ್ಯತೆಯಿಂದ ಕಲಾರಂಗದಲ್ಲಿ, ಹಾಗೂ ಜನಜೀವನದಲ್ಲಿ, ತನ್ನದೇ ಆದ ಪ್ರತ್ಯೇಕ ಮಮತೆಯ ಪಡೆದಿದೆ - ನೃತ್ಯ, ಸಮುದ್ರದಂತಹ ವಿಶಾಲ ವ್ಯಾಪ್ತಿಯನ್ನು ಪಡೆದ ನೃತ್ಯ ಕಲೆಯನ್ನು ಅಕ್ಕಿಯ ಭಾವನೆಯಿಂದ ಕಂಡಲ್ಲಿ, ಇದೊಂದು ಅರಾಧನಾ ಮಾರ್ಗದ ಮೈಲಿಗಲ್ಲು, ನಿತ್ಯಾನಂದದಾಯಕವಾದ, ಜೈತನ್ಯಕಾರಕವಾದ, ಮಹಾನಿಧಿ, ಪಂಡಿತರು, ಪಾಪುರು, ಬುದ್ಧಿಜೀವಿಗಳು, ಉತ್ತಮರು, ಮೇಲ್ವಿಚಾರಣೆ, ಕಳೆದುಕೊಂಡವರು, ಶ್ರೀಮಂತರು ಎಂದೆಂಬ ಭೇದಭಾವವಿಲ್ಲದ ಪ್ರತಿಯೊಂದರೂ ಅವರವರ ಗುಣಗಳಿಗೆ ಅನುಗುಣವಾಗಿ ಅನಂದವನ್ನೀಯುವಂತಹ ಒಂದು ಮನರಂಜನೀಯ ಕಲೆ. "ಕಲೆ" ಎಂದರೆ - ಗುರುತು ಲಲಿತವಿದ್ಯೆ, ಕಾಂಠಿ, ಸೇರು ಮೊದಲಾಗಿ ನಿಫಂಟಿನ ಅರ್ಥ. "ಕಲ" ಎಂಬಲ್ಲಿ - ತಿಳಿ, ಕಲೆ, ಕಲಿತು ತೋರು ಎಂಬ ಭಾವಾರ್ಥವಿದೆ. ಈ ತೊ ಅಂಶಗಳೆ ಆಡಿದೆ. ಅಂತೆಯೇ ಬಗ್ಗಿನ ಎಲ್ಲಾ ವಿಚಾರಗಳನ್ನೂ ಭಾವನಗಳನ್ನೂ - ಪುಕಟಿಸುವ ಜೈತನ್ಯವನ್ನೂ ಪಡೆದಿದೆ. ಲೀಲಾಕಾನುಭವವನ್ನು ತಿಳಿಯುವುದು, ಅದರಲ್ಲಿ ಸೇರುವುದು, ಸೇರಿ ತಿಳಿದು ತಿಳಿಸುವುದು ಅರ್ಥಾತ್ ಸಾಧವನ್ನು ಕಲೆಯ ಮುಖಾಂತರ ಪ್ರಕಟಪಡಿಸುವುದೇ ನೃತ್ಯದ ರೀತಿ. ಮಾನಸಿಕ ಭಾವನೆಯನ್ನು, ಪ್ರತಿಯೇ ಲೀಲಾಭಾವಾದಿಗಳ ನಿಖರ ನಿಷ್ಕ್ರಿಯಲ್ಲಿ ರೂಪಗೊಳಿಸಿ, ಪ್ರಯೋಗ ವಿದ್ಯೆಯ ನಿಯಮಾದಿಗಳನ್ನು ನಿರ್ದಿಷ್ಟಪಡಿಸಿ ಅಲಿತ ಎಂದರೆ - ಸುಂದರವಾದುದು, ಸರಳವಾದುದು, ಮನೋಹರವಾದುದು, ಅನಂದದಾಯಕವಾದುದು - ಎಂಬ ಅರ್ಥವಿದೆ. "ವಿದ್ಯೆ" ಎಂದರೆ - "ಜ್ಞಾನ", "ತಿಳುವಳಿಕೆ", ಓದು - ಮೊದಲಾದ ಅರ್ಥವಿದೆ.

Fig. 1: Kannada Language dataset consisting of different entities like Name of a person, Place, Organization

<darwin1><Aesthetics><Biography><1981><Book><ಚಾರ್ಲ್ಸ್ ಡಾರ್ವಿನ್><ಡಾ. ಎನ್.ಎಸ್. ವೀರಪ್ಪ><35>

Page 5

ಮೂಲದ ಮಾತು (ದ್ವಿತೀಯ ಮುದ್ರಣ) ಮೌನದಿಂದ ಬಾಳಿ, ಜನಜೀವನದಲ್ಲಿ, ವಿಚಾರಕ್ರಾಂತಿಯನ್ನೆತ್ತಿದ ಮಹಾಪುರುಷರಲ್ಲಿ, ಚಾರ್ಲ್ಸ್ ಡಾರ್ವಿನ್ ಅಗ್ರಗಣ್ಯ. ಅವನು ತನ್ನ ವಿಶಾಸತತ್ವವನ್ನು ಪ್ರಕಟಿಸಿ, ಮಾನಸಿಕ ಸ್ವಾತಂತ್ರ್ಯವನ್ನು ದೊರಕಿಸಿಕೊಟ್ಟ ನಂತರ ಮಾ ನಿರಾತಂಕವಾಗಿ ಬೆಳೆಯುವಕಾಲವಾಯಿತು. ಆ ತತ್ವದ ಪ್ರಭಾವ ಇಂದಿನ ಎಲ್ಲ ವಿದ್ಯಾರಂಗಗಳ ಮೇಲೂ ಬಿದ್ದು, ಹೊಸದೊಂದು ಅಭ್ಯಾಸಕ್ರಮವೇ ಏರ್ಪಟ್ಟಿದೆ. ಯಾವ ವಿಷಯವನ್ನೇ ಆಗಲಿ, ಅದರ ಹುಟ್ಟು, ಬೆಳವಣಿಗೆ ಇ ವೇದದಾದುದು ಡಾರ್ವಿನ್‌ನ ಕಾಂಡದಿಂದಲೇ. ಮೇಲಾಗಿ ಡಾರ್ವಿನ್ ಅರ್ಥ ವಿಜ್ಞಾನಿ. ಅವನ ಸಂಶೋಧನೆಯ ದೃಷ್ಟಿಯಲ್ಲಿಯೂ ವಿಜ್ಞಾನಪ್ರಕೃತಿಯಲ್ಲಿ ಮೇಲ್ದಿಕ್ಕಿನಲ್ಲಿ ಪರಿಣಾಮಿಸಿ, ನಿರ್ಗತಾಸ್ತದ ಬೆಳವಣಿಗೆಯ ಡಾರ್ವಿನ್‌ನ ವ್ಯಕ್ತಿತ್ವ ಅಚ್ಚಳಿಯದಂಥ ಮುದ್ರೆಯೊತ್ತಿದೆ. ಸಮಾಜಸುಧಾರಕರಾಗಲಿ, ರಾಜಕಾರಣಜಟುಗಳಾಗಲಿ, ತಮ್ಮ ಜೀವನದಲ್ಲಿ, ಹಲವು ಘಟನೆಗಳನ್ನನುಭವಿಸಿ ಮಹಾಪುರುಷರಾಗುವರು. ಅವರ ಅನುಭವ ಮತ್ತು ದೃಷ್ಟಿಯಲ್ಲಿಯೂ ಪರಿಚಯವೇ ಒಂದು ದೂಡು, ಜ್ಞಾನ. ವಿಜ್ಞಾನಿಗಳ ಜೀವನದಲ್ಲಿ ಅಂಥ ಘಟನೆಗಳು ಸಂಭವಿಸಿದರೂ ಅವೇ ಅಷ್ಟೊಂದು ಪ್ರಭಾವವಲ್ಲ; ಜನತೆಯ ಕಲ್ಯಾಣದೃಷ್ಟಿಯಿಂದ ಅವರು ನಡೆಸುವ ಸಂಶೋಧನ ದೃಷ್ಟಿಯಲ್ಲಿಯೂ ಹೆಚ್ಚು ಅಕರ್ಷಣೀಯವಾದವು. ಆದುದರಿಂದ ಈ ಗ್ರಂಥದಲ್ಲಿ, ಜೀವನ ಚರಿತ್ರೆಗೆ ಪ್ರಾಧಾನ್ಯ ಕೊಟ್ಟಿರುವಂತೆ ಸಂಶೋಧನಾಂಶಗಳಿಗೂ ವಿಜ್ಞಾನತತ್ವ ದೃಷ್ಟಿಯಲ್ಲಿ ಪ್ರಾಧಾನ್ಯ ಕೊಟ್ಟು, ಕಥನಕಾರ್ಯವನ್ನು ವಿಜ್ಞಾನಿಗಳ ಜೀವನಕಥನಗಳ ನಮ್ಮ ಸಾಹಿತ್ಯದೇವಿಗೆ ಹೊಸ ಅಧ್ಯಾಯ. ಈಚೆಗೆ ನಮ್ಮ ನುಡಿಯಲ್ಲಿ ವಿಜ್ಞಾನಶಾಸ್ತ್ರದ ವಿವಿಧ ಶಾಖೆಯ ಗ್ರಂಥಗಳು ಹೊರಬೀಳುತ್ತಿವೆ. ಹಾಗೆಯೇ ವಿಜ್ಞಾನಿಗಳ ಜೀವನವನ್ನೂ ನಾವು ಅರಿತುಕು ಕಲವಾಯ ಗ್ರಂಥಗಳು ನಾವು ನನ್ನೊಡೆಯವನು ಸಾರಿದಂತೆ ವೈಜ್ಞಾನಿಕ ಗ್ರಂಥಗಳೂ ಬೆಳಕಿಗೆ ಬಂದು ನಾಡ ಮಕ್ಕಳ ಮನಸ್ಸಿನಲ್ಲಿ ವಿಜ್ಞಾನಪ್ರೇಮ ಅಂಕುರಿಸುವಂತಾಗಬೇಕು. ಈಚೆಗೆ ಕನ್ನಡದಲ್ಲಿ, ಹಲವಾರು ವಿಜ್ಞಾನಿಗಳ ಹೊರಬಿದ್ದಿರುವುದು ಸಂತೋಷದ ವಿಷಯ. ಈ ಗ್ರಂಥದ ಪ್ರಕಾಶನದಲ್ಲಿ ನೆರವಾದ ಎಲ್ಲರಿಗೂ ನಾನು ಮುಖೆ. ಬ್ರಿಷ್ಲನ್ ಬಗರ ಡಾ. ಎನ್.ಎಸ್. ವೀರಪ್ಪ ಮಾರ್ಚ್ 31, 1957

Page 7

ಸಾಹಿತ್ಯದಲ್ಲಿ, ಜೀವನಚರಿತ್ರೆ ದೇಶಪೂರ್ವಕವಾಗಿ ಎಲ್ಲ ಪೂರ್ವಕಾಂಕ್ಷೆಗಳನ್ನೂ ಒಳಗೊಂಡ ಮಿಶ್ರಾಹಾರ ಅಗತ್ಯವಷ್ಟೆ. ಹಾಗೆಯೇ ನಮ್ಮ ಜೀವನದ ಪೂರ್ವಕಗಳ ಜೀವನದ ವಿವಿಧ ಪ್ರಾಕಾರಗಳನ್ನು ವಿವಿಧ ರೀತಿಯಲ್ಲಿ ಚಿತ್ರಿಸುವ ಸಾಹಿತ್ಯ ಅಗತ್ಯ. ಕಾವ್ಯ, ಕಥೆ, ಕಾದಂಬರಿ, ನಾಟಕ, ವಿಮರ್ಶೆ, ಜೀವನಚಿತ್ರ, ಅತ್ಯಕಥೆ, ಜೀವನ ಚರಿತ್ರೆ - ಇವು ಅಂಥ ಸಾಹಿತ್ಯದ ವಿವಿಧ ಮುಖಗಳು. ಅವುಗಳಲ್ಲಿ ಜೀವನ ಚರಿತ್ರೆಯನ್ನು ಕುರಿತು ಹಲವಾರು ಅಂಶಗಳನ್ನು ಪರಿಶೀಲಿಸುವುದು ಪ್ರಸ್ತುತ. ಲೇಖನದ ಉದ್ದೇಶ ಜೀವನಚರಿತ್ರೆ ಸಾಹಿತ್ಯವೋ, ಇತಿಹಾಸವೋ ಎಂಬ ಸಮಸ್ಯೆ ಸ್ವಾಭಾವಿಕವಾಗಿ ಎದ್ದುಕಾಣುತ್ತದೆ. ಮಾನವನ ಅರ್ಥಿಕ, ಧಾರ್ಮಿಕ, ರಾಜಕೀಯ ಮುಂತಾದ ಸಮಸ್ಯೆಗಳನ್ನೊಳಗೊಂಡಂತೆ ಅವನ ಸಂಸ್ಕೃ ಲಮನ್ಯದೆಯನ್ನು ಚಿತ್ರಿಸುವುದು ಇತಿಹಾಸದ ಕಲಸ. ಆ ವಿಶಾಲ ಚಿತ್ರಣದ ಬಾಂಧವ್ಯದಲ್ಲಿ ಮಾನವ ವ್ಯಕ್ತಿಗಳು ನಕ್ಷತ್ರಗಳಂತೆ ಚದುರಿಹೋಗಿರುತ್ತವೆ. ಎಂದರೆ, ವ್ಯಕ್ತಿಗಳು ಇತಿಹಾಸದಲ್ಲಿ ಕಾಣಿಸಿಕೊಂಡು ಸಂಸ್ಕೃತಿ - ನಾಗರಿಕತೆಗಳ ಮನ್ನಡೆಯಲ್ಲಿ ಅವರು ಸಂಕೀರ್ತನು ಮಾತ್ರ. ಅವರಿಗೆ ಅಷ್ಟು ಪ್ರಾಧಾನ್ಯವಿಲ್ಲ. ಸಾಹಿತ್ಯದಲ್ಲಿ ವ್ಯಕ್ತಿಗಳ ಪ್ರಾಧಾನ್ಯ; ಸಂಸ್ಕೃತಿ-ನಾಗರಿಕತೆಗಳ ಚರಿತ್ರೆ ಕೇವಲ ವ್ಯಕ್ತಿತ್ವ ಪರಿಪೂರ್ಣತೆಯ ಹಿನ್ನೆಲೆ ಮಾತ್ರ. ಒಂದು ಚಿತ್ರಣದಲ್ಲಿ ಒಂದೇ ವ್ಯಕ್ತಿಗೆ ಪ್ರಾಧಾನ್ಯವಿರಬಹುದು; ಇನ್ನೊಂದರಲ್ಲಿ ಹಲವಾರು ವ್ಯಕ್ತಿಗಳಿಗೆ ಆ ಪ್ರಾಧಾನ್ಯ ಹಂಚಿಹೋಗಿರಬಹುದು; ಮತ್ತೊಂದರಲ್ಲಿ ವ್ಯಕ್ತಿಯ ಸುಳಿವೇ ಕಾಣದಿರಬಹುದು. ಅದರೂ ಅಲ್ಲಿ ಪ್ರತ್ಯಕ್ಷವಾಗುವ ಪರೋಕ್ಷವಾಗಿಯೂ, ವ್ಯಕ್ತಿಪ್ರಧಾನವಾದಿಗಿರುತ್ತದೆ. ನಿರ್ಗತದ ಸನ್ನಿವೇಶವೊಂದರ ವರ್ಣನೆಯಾಗಲಿ, ಸಾಮಾಜಿಕ ಸನ್ನಿವೇಶವೊಂದರ ಚಿತ್ರಣವಾಗಲಿ, ಭಾವನೆಯೊಂದರ ನಿರೂಪಣೆಯಾಗಲಿ - ಅಷ್ಟು ಸಾಹಿತ್ಯದಲ್ಲಿ, ಬರುವುದು ಬಿರಬಲ್ಲ ಪ್ರಭಾವವಾಗಿ ಸಾಹಿತ್ಯದಲ್ಲಿ ಪ್ರತ್ಯಕ್ಷವಾಗಿಯೂ, ಪರೋಕ್ಷವಾಗಿಯೂ, ಮಾನವತ್ವಕ್ಕೆ, ವ್ಯಕ್ತಿತ್ವಕ್ಕೆ, ಪ್ರಾಧಾನ್ಯವಿರುವುದು ಸಹಜವೇ. ಅಷ್ಟೇ ಅಲ್ಲ, ಇತಿಹಾಸಕ್ಕೂ ಸಾಹಿತ್ಯಕ್ಕೂ ಉದ್ದೇಶದಲ್ಲೂ ಭಿನ್ನತೆಯಿದೆ. ಎಂದೆ

Fig. 2: Indicates the dataset consisting of Named Entities like name of a person, place, recognition of awards, Designation of a person and various other named entities shall be found in the above dataset

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,

$$TPR = \frac{\text{Predicted result}}{\text{Total number of words}} \quad (1)$$

$$FPR = \frac{\text{Incorrect Prediction}}{\text{Total number of words}} \quad (2)$$

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.

ಡಾ.. ಎನ್.ಎಸ್.ವೀರಪ್ಪ - **designation found**

Fig. 2: Designation of a person found from the proposed method

The result of applying the proposed method over a bench mark dataset shall be seen in fig.2. As we were intending to recognize the Kannada Named Entities from a Kannada text documents, we followed two stages of case studies. The case study has focused on finding the designation of a person.

Similarly, the other case studies shall be performed using the proposed R-KNER. So that the Kannada Named Entities like Place, and Organization shall be found from a Kannada Text.

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 dataset consisting of names of a person shall be seen in fig.3.

ಚಾರ್ಲ್ಸ್ ಡಾರ್ವಿನ್- **name found**
 ಐಸಾಕ್ ನ್ಯೂಟನ್ - **name found**
 ಮೇಡಂ ಕ್ಯೂರಿ- **name found**
 ಲೂಯಿ ಪ್ಯಾಶ್ಚರ್- **name found**
 ಸರ್ ಎಂ. ವಿಶ್ವೇಶ್ವರಯ್ಯ- **name found**
 ಸರ್ ಸಿ.ವಿ. ರಾಮನ್- **name found**
 ಭಾಗ್ಯಶಿಲ್ಪಿ- **name found**
 ಲೆನಿನ್- **name found**

Fig. 3: The result of applying R-KNER over a dataset of CIIL resulted in obtaining the names of a person from text files of CIIL is shown above.

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.

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

Method	TPR	FPR	Overall
Nymble	75.46	24.54	75.46
MENE	56.51	43.49	56.51
Association Rule Mining	66.36	33.64	66.36

Maximum Entropy	67.54	32.46	67.54
Proposed R-KNER	89.32	10.68	89.32

ಡಾ.. ಎನ್.ಎಸ್ ವೀರಪ್ಪ - **designation found**
 ಚಾರ್ಲ್ಸ್ ಡಾರ್ವಿನ್ - **name found**
 ಸಿ.ಕೆ. ವೆಂಕಟರಾಮಯ್ಯನವರ - **name found**
 ಡಿ.ಆರ್. ರಾಮಯ್ಯನವರ - **name found**
 ಜಗದೀಶ ಚಂದ್ರ ಬೋಸ್- **name found**
 ಚಾರ್ಲ್ಸ್ ಡಾರ್ವಿನ್- **name found**
 ಐಸಾಕ್ ನ್ಯೂಟನ್ - **name found**
 ಮೇಡಂ ಕ್ಯೂರಿ- **name found**
 ಲೂಯಿಸ್ ಪ್ಯಾಶ್ಚರ್- **name found**
 ಸರ್ ಎಂ. ವಿಶ್ವೇಶ್ವರಯ್ಯ- **name found**
 ಸರ್ ಸಿ.ವಿ. ರಾಮನ್- **name found**
 ಭಾಗ್ಯಶಿಲ್ಪಿ- **name found**
 ಲನಿನ್- **name found**

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

1. Deepti Chopra1, Sudha Morwal, "Named Entity Recognition in Punjabi Using Hidden Markov Model," International Journal of Computer Science & Engineering Technology (IJCSSET), Vol. 3 No. 12 pp. 616-621, Dec 2012, (references)
2. Kanwalpreet Singh Bajwa, Amardeep Kaur, "Hybrid Approach for Named Entity Recognition" International Journal of Computer Applications (0975 - 8887), Volume 118 - No. 1, May 2015
3. Virat V. Giri, Dr.M.M. Math and Dr.U.P. Kulkarni, "A Survey of Automatic Text Summarization System for Different Regional Language in India Bonfring International Journal of Software Engineering and Soft Computing, Vol. 6, Special Issue, October 2016.
4. Baldi A. "Computational Approaches for Drug Design and Discovery: An Overview." *Systematic Reviews in Pharmacy* 1.1 (2010), 99-105. Print. doi:10.4103/0975-8453.59519
5. Alireza Mansouri, Lilly Suriani Affendey, Ali Mamat, "Named Entity Recognition Approaches," International Journal of Computer Science and Network Security, VOL.8 No.2, February 2008.
6. S Amarappa1and S V Sathyanarayana, "A Hybrid approach for Named Entity Recognition, Classification and Extraction (NERCE) in Kannada Documents," Proc. of Int. Conf. onMultimedia Processing, Communication and Info. Tech., MPCIT.
7. Ms. Maithilee L. Patawar, Mrs. M. A. Potey, "Approaches to Named Entity Recognition: A Survey International Journal of Innovative Research in Computer and Communication Engineering Vol. 3, Issue 12, December 2015.
8. Kamaldeep Kaur, Vishal Gupta, "Name Entity Recognition for Punjabi Language" International Journal of Computer

- Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 Vol. 2, No.3, June 2012.
9. S Amarappa 1 Dr. S V Sathyanarayana, "Named Entity Recognition and Classification in Kannada Language," International Journal of Electronics and Computer Science Engineering, Vol.2, No.1, 2015.
10. Kommaluri VIJAYANAND and R. P. Seenivasan, "Named Entity Recognition and Transliteration for Telugu Language", Problems of Parsing in Indian Languages, 5 May 2011.
11. Bhuvaneshwari C Melinamath, "Rule based Methodology for Recognition of Kannada Named Entities," International Journal of Latest Trends in Engineering and Technology, Vol. 3 Issue 4 March 2014.
12. S Amarappa, S V Sathyanarayana, "Kannada Named Entity Recognition and Classification using Support Vector Machine," Transactions on Machine Learning and Artificial Intelligence, Vol.5, Issue1, 2017.
13. Leena Jain, Prateek Agrawal, "English to Sanskrit Transliteration: an effective approach to design Natural Language Translation Tool," International Journal of Advanced Research in Computer Science, Vol 8, No.1. Jan-Feb-2017.
14. Hinal Shah, Prachi Bhandari, Krunal Mistry, Shivani Thakor, Mishika Patel and Kamini Ahir," Study of Named Entity Recognition," International Journal of Information Sciences and Techniques (IJIST) Vol.6, No.1/2, March 2016.
15. Daljit Kaur, Ashish Verma, "Name Entity Recognition by New Framework Using Machine Learning Algorithm," IOSR Journal of Computer Engineering (IOSR-JCE), Volume 16, Issue 5, Ver. IV (Sep - Oct. 2014).
16. Dikshan N. Shah , Harshad Bhadk, "A Survey on Various Approach used in Named Entity Recognition for Indian Languages", International Journal of Computer Applications (0975 - 8887) Volume 167 - No.1, June 2017.
17. Varinder Kaur, Amandeep Kaur Randhawa," A Brief Survey about Work Done in Named Entity Recognition for Indian Languages" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 7, July 2015.
18. Mamatha Hosalli Ramappa, Srikantamurthy Krishnamurth, "A Comparative Study of Different Feature Extraction and Classification Methods for Recognition of Handwritten Kannada Numerals", International Journal of Database Theory and Application Vol. 6, No. 4, August, 2013.
19. Mamatha H.R, Sucharitha S and Srikanta Murthy K, "Handwritten Kannada Numeral Recognition based on the Curvelets and Standard Deviation" International Journal of Signal Processing Systems Vol. 1, No. 1 June 2013.
20. Kalyani, R., Thej, M., Prabhakar, K., Kiran, J.Single dominant left coronary artery: An autopsy case report with review of literature(2011) Journal of Cardiovascular Disease Research, 2 (2), pp. 130-132. DOI: 10.4103/0975-3583.83038