



COMPARATIVE ANALYSIS OF VARIOUS APPROACHES BASED ON NAMED ENTITY RECOGNITION-A SURVEY

Sakshi¹, Shailendra Singh², Dharam Vir³, Shashank Singh⁴

^{1,2,4}Department of Computer Science & Engineering
Punjab Engineering College (Deemed to be University)
Chandigarh, India³ Department of Otolaryngology
PGIMER, Chandigarh, India

Abstract: Extraction of advantageous information from the data has turned out to be the most decisive activity across all domains because of the increase in the availability of data. Information Extraction goes into the more challenging task due to the availability of data in the form of documents written in the natural language. Named Entity Recognition (NER) is the part of Information Extraction which is used to extract important information from the code-mixed and informal data and then classifies these extracting named entities into its pre-characterized classes. For example: person, location, organization, city, state, and country etc. NER is acknowledged as the dominant task in the field of Natural Language Processing (NLP). This paper provides a survey of various methods and techniques which are being used in the extraction of proper nouns appeared in the document. This paper also outlines the knowledge of various challenges which are being faced while extracting the named entities. It also provides some research directions which various researchers can explore.

Keywords: Entity Extraction, Machine Learning, Linguistic approaches, Named Entity Recognition.

1 INTRODUCTION

NER is the process of identification of the named entities from the code-mixed and informal data. Code mixed data refers to mixing of two or more languages. NER is the process to detect the proper nouns from the code-mixed and informal data. It is the procedure to extract the named entities from the text and categorize these extracted named entities into its predefined classes [1]. Named Entities are basically the proper nouns that typify the Name of person, location, organization, river, percentage, quantity and time etc. NER is inherently the sub field of NLP. NER is a sub-effort of Information Extraction [2].

Some of the tasks of NER are automatic summarization, machine translation, question answering system, information extraction, information retrieval etc. Indian languages NER is still consulted as germinating field of NLP [2]. One can recognize the named entities only if one complies summations on the natural languages. NER can be divided into two sub tasks: named entity identification (NEI) and named entity classification (NEC). Named entities are the minute aspects in text assets to predefined classes such as name of a person, organization, location etc. [1].

The main efforts of NER include on natural entities like location, person, organization, time, date, measurement and number. The function of named entity recognition and classification can be narrated by recognition of proper nouns in computer interpreted document by means of annotation and classification of tags for information extraction. NER hit a decisive role in other kind of disambiguation's and reference resolutions.

The NER project was first arrived in the sixth message understanding conference (MUC- 6) sunheim (1995) and commit to identification of proper nouns (people and

organizations), place names, temporal expressions and numerical expressions [3]. According to various conferences named entities have different kind of categorizations and classifications. According to MUC-6 proper nouns were classified into three different labels. Proper nouns and their labels are: ENAMEX: location, organization, and person. TIMEX: time, date. NUMEX: quantity, money, percentage. According to DARPA's message understanding conference proper nouns were classified into three top-level categorizations: temporal expressions, number expressions and entity names. As we know NER is one of the rustling fields of research from the past 20-25 years. A huge amount of progression has been found in disclosing the named entities but NER is still enduring an immense problem at large.

2 DETECTION OF PROPER NOUNS FROM THE CODE-MIXED AND INFORMAL DATA AND VARIOUS APPROACHES BASED ON MACHINE TRANSLATION.

Different approaches and different methods have been used to detect the named entities which further classify these extracted named entities into its predefined classes[3]. Various approaches include: supervised, semi supervised and unsupervised learning methods as shown in Figure 1: Supervised learning methods include various machine learning approaches like hidden markov model, maximum entropy model, conditional random fields based Models, decision based Model, and support vector machine based model. Semi supervised learning methods include bootstrapping based models. Some rule based approaches has been used to detect the named entities from code mixed and informal data: linguistic approach and list lookup approach[4]

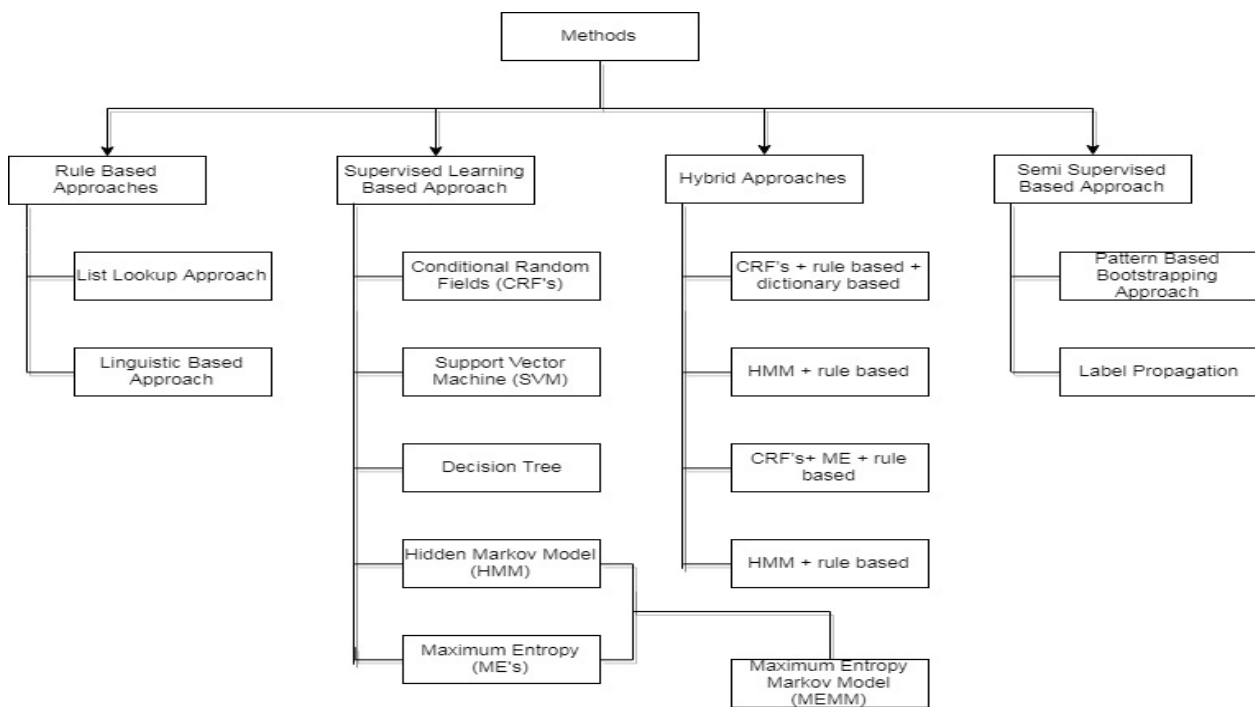


Figure 2: Different Methods Regarding NER

Various researchers have tried to translate text from source language to target language and have tried to extract named entities or proper nouns from the code-mixed and informal data.

2.1 Approaches based on Machine Translation

A research work was mentioned by Dhariya et al. [5] and developed the boosted mixed approach of phrase based statistical machine translation system (SMT), example based machine translation (EBMT) and rule based machine translation system (RBMT). The main aim to combine these approaches is to get the better accuracy while translating the text from Hindi to English language. The hybrid model consists of four basic steps: segmentation, translation, part of speech tagging, and rearrangement. The comparisons of our proposed hybrid approach with the respected available online translators like google, Babylonian and Bing is sightseen that states that our proposed hybrid approach is much enhanced or well outlined than these available online translators. Another work on machine translation was proposed by Chakrawarti et al. [6] which may concludes the controversy regarding ambiguities and translation divergences. The proposed approach comprises of seven steps. 1. source language i.e. Hindi, 2. target language i.e. English, 3. phases of NLP and ambiguities. 4. models and techniques which are based on machine translation: 5. techniques which are based on word sense disambiguation. 6. concept map construction based languages. 7. corpuses based on both Hindi and English languages like word net and Hindi word net. System architecture regarding translation from Hindi=>English consists of many steps. The result of this ongoing research is high précised, accurate and large quality translation of Hindi poetry literature into

English. Kumar et al. [7] tried to propose another work on machine translation and developed syntax directed translator approach which translates text from English to Hindi language without the provision of human translator. In this technique syntax is checked during Translation and many challenges like word order, word sense and ambiguity are resolved. This technique is used to check the correctness of grammar.

2.2 Rule Based Approaches applied in the work of NER

This section provides a survey of the various researches done in Indian languages. Mathur et al. [8] developed the hybrid approach using rule based approach and statistical based approach to extract the named entities from the document. The main agenda of this approach stands to increase the accuracy while translating the proper nouns and named entities. To extract the named entities from the English sentence, a StanfordNER tool is used. In the training corpus six named class entities are predefined and on the basis of these six predefined named class entities further named entities are extracted. Phonification algorithm is applied to get the respective phonemes. The system reported the recall of 83.16%, precision of 83.65% and F-measure of 83.40% respectively. Singhal et al. [9] proposed the hybrid approach regarding English-Hindi transliteration. This paper is discussing about the ongoing issues regarding named entities while translating the text from source language to target language. Syllabification and uni-gram model is proposed to resolve the issues concerning named entities. Primarily knowledgebase with English-Hindi named pairs is maintained then English name is confirmed from the knowledgebase in case if it is located then the interrelated

Hindi name is selected, else the Syllabification is used in the further process of named entity transliteration. A suitable corpus has been constructed for the entire feasible combination of English alphabets for these phonemes and their related Hindi aksharas are matched with it. The system reported the recall of 84.23%, precision of 84.23% respectively.

Chopra et al.[10] developed an approach of handling unknown words while recognizing the named entities. Unknown words are handled through transliteration approach. Transliteration approach involves the procedures of both the training data as well as testing data. The hidden markov model and viterbi algorithm is used for the better detection of named entities. The system reported the recall of 95.80%, precision of 96.13% and F-measure of 96.04% respectively. Morwal et al.[11] developed an approach name of transliteration for the better identification of named entities and proper nouns in natural language. In this process separated codes are designed for the training phase as well as for the testing phase which helps in the implementation of Transliteration and give the accurate named entities. Results are based on the performance metrics. The system reported the recall of 70.6%, precision of 70.6% and F-measure of 70.6% respectively.

Nayan et al.[12] developed an approach named phonetic matching for the NER. Phonetic matching involves the matching of strings of various languages on the basis of same sounding property. Architecture is proposed which consists of various parts: crawler, parser and phonetic matcher and editex algorithm. Various Transliteration rules are applied to do the task of phonetic matching. Baseline task has been selected which comprises of different parts: abbreviations check, first letter matching, preprocessing and editex Score. System achieves the Precision around 80%. Different research work was reported by Chopra et al.[13] and developed an approach to handle ambiguities and unknown words in NER by applying anaphora resolution. It is an approach to identify what an individual noun phrase and a pronoun at a given time indicates to. Anaphora resolution is a task of determining the antecedent of an anaphor. NER task is completed through 2 phases: training phase and testing phase. Different named entities, ambiguities and various unknown words are identified during the training as well as testing Phase. The system reported the recall of 95.80%, precision of 96.3% and F-measure of 96.04% respectively. Comparison of different rule based techniques used in NER as shown in Table 1:

Table 2: COMPARISON OF DIFFERENT RULE BASED TECHNIQUES USED IN NER

Language	Technique	Recall	Precision	F-measure	Reference No.
English- Hindi	Rule based + statistical based	83.16%	83.65%	83.40%	[8]
English-Hindi	Grapheme based + phoneme based	84.23%	84.23%	--	[9]
Hindi	Transliteration + hidden markov model + viterbi algorithm	95.80%	96.13%	96.04%	[10]
English	Transliteration	70.6%	70.6%	70.6%	[11]
English	Phonetic matching	81.39%	81.40%	--	[12]
English-Hindi	Anaphora resolution	95.80%	96.3%	96.04%	[13]
English-French	Anaphora resolution	95.80%	96.3%	96.04%	[13]
English-Urdu	Anaphora resolution	95.80%	96.3%	96.04%	[13]
English-Telugu	Anaphora resolution	95.80%	96.3%	96.04%	[13]
English-Bengali	Anaphora resolution	95.80%	96.3%	96.04%	[13]

2.3 Supervised Learning Based Approaches applied in the work of NER

Chopra et al.[14] developed an approach name of HMM for the NER. In this paper a brief comparison of using HMM over various methods is done. It is very difficult to manage the list of open class words while adopting the gazetteers. The following phases are involved in named entity recognition and classifications are annotation phase, train HMM, and test HMM. First step includes the annotation module which translates the input text to trainable data, second step includes the train module in which some parameters like start probability, transition probability and emission probability are calculated and the last step includes

the test module in which optimal tag sequence is detected by applying viterbi algorithm. The limitation regarding this approach is unknown names are not handled in the précised way. The system reported the recall of 97.14%, precision of 97.14% and F-measure of 97.14% for Hindi language.

Morwal et al.[15] developed an approach for the identification and classification of named entities in Indian languages like Hindi, Bengali, and Telugu in the purpose to increase and improve the accuracy. This approach follows the HMM to extract the named entities or proper nouns. For the implementation part data from the various resources has been collected. Hindi data collection is done from the tourism domain corpus and NLTK Indian corpus, the

Bengali and Telugu data is taken from the NLTK corpus as well. The following phases are involved in named entity recognition and classifications are annotation phase, train HMM, and test HMM. Viterbi algorithm is involved to compute the optimal state sequence for the provided test sentences. The system reported the recall of 96%, precision of 96% and F-measure of 96% for Hindi language.

Chopra et al.[16] developed an approach named HMM for the NER in English language. Treebank corpus is taken from the online available NLTK tool. Total of 6680 words are taken from the corpus. After collecting words, these words are trained for the NER task. HMM have 3 kinds of Probabilities: start probability, transition probability and emission probability and on the basis of these probabilities NER task is performed. System achieves the F- Measure of 73.8%respectively.

Eqbal et al.[17] developed an approach named support vector machine (SVM) for the task of NER. The main agenda of using this technique is that it helps in the detection of various vectors. It states even if an individual vector is an element of distinct target class or not. The main idea of using this approach is that both the training and the testing data coincide with the individual vector space. [17] System based on SVM achieves the recall, precision and F-Score of 94.3%, 89.4% and 91.8% respectively. Another work reported by Eqbal et al.[18] and developed another approach named conditional random fields (CRF's). The applied system which is based on CRF's reported the recall of 93.8%, precision of 87.8% and F-measure of 90.7% respectively. %. Comparison of different supervised learning based techniques used in NER as shown in Table 2:

Table 3: COMPARISON OF DIFFERENT SUPERVISED BASED LEARNING TECHNIQUES USED IN NER

Language	Technique	Recall	Precision	F-measure	Reference No.
Hindi	Hidden Markov Model	97.14%	97.14%	97.14%	[14]
Hindi	Hidden Markov Model	96%	96%	96%	[15]
Bengali	Hidden Markov Model	98.5%	98.5%	98.5%	[15]
Telugu	Hidden Markov Model	98.6%	98.6%	98.6%	[15]
English	Hidden Markov Model	--	--	73.8%	[16]
Bengali	Support Vector Machine	94.3%	89.4%	91.8%	[17]
Bengali	Conditional Random Fields	93.8%	87.8%	90.7%	[18]

2.4 Hybrid Approaches applied in the work of NER

Gupta et al.[1] developed a mixed approach to extract the proper nouns from the code mixed and informal data. It is very difficult to extract the proper nouns from the code mixed and informal data. Some feature set has applied to the CRF's classifier which further helps in the task of labeling the sequence of tokens. This process is done in two steps: entity extraction and entity classification. There are 22 different types of predefined entities are characterized in the training set and on the basis of these predefined entities further more entities or proper nouns are extracted. Three steps are being followed by this approach. First step is pre-processing that further consist of sub steps like tokenization and token encoding. Since the data are coming from the Twitter so there is a need of tokenizer which can provide the tokens to the words and further need a CMU tagger which provides the POS tagging of words. IOB encoding is used to tag the tokens in the chunking task. Second step involves the task of sequence labeling to label the sequence of tokens. Third step involves post processing which includes rule based and dictionary based approaches. The limitation of using proposed system is that this system gives the less accuracy. The proposed approach gives the recall of 50.39%, precision of 81.15%, F-measure of 62.17% respectively.

Amarappa et al.[19]developed hybrid model consist of HMM and rule based model. This task is based on NERand classification thenextraction of entities from the unstructured documents. The whole process is based on the recognition of proper nouns and then categorizes these

identified proper nouns into its pre-formed categories. Root words are also identified on the basis of classified pre-defined categories. The proposed approach gives the recall of 94.61%, precision of 95.10%, F-measure of 94.85% respectively.

Srivastava et al.[20] developed mixed approach which is made up of machine learning based approaches like maximum Entropy and CRF's and linguistic approaches like rule based approaches. To amaze the drawback of statistical models, linguistics approaches are used. Voting criteria is used to increase the accuracy of the applied system. The system reported the recall of 84.88%, precision of 81.11% and F-measure of 82.95% respectively.

Kumar Saha et al.[21] developed the mixed approach that comprises of maximum entropy (ME) model, language specific rules and gazetteers for the effort of NER. Baseline NER system is designed using named entity annotated corpora and some set of features. Few language specific rules are defined to the system to identify few specific named entity classes. To improve the achievement some gazetteers and context patterns are added to the system. ME model is used for the better extraction of named entities. Once the one-level NER system is developed then a set of rules is added to detect the nested named entities. The proposed approach is capable to identify 12 classes of named entities. Results are based on the performance metrics. The reported system attains the F-measure of 65.13% in Hindi, F-measure of 65.96% in Bengali, F-measure of44.65% in Oriya, F-measure of18.74% in Telugu and F-measure of35.47% in Urdu.

Laishram et al. [22] developed mixed approach which is made up of CRF's and rule based approach for the task of NER in Manipuri language. Different exclusive word features are characterized by rule based approaches that are further used to classify the proper nouns by the CRF classifier. System attains the better accuracy with minor corpus scope. The system reported the recall of 92.26%, precision of 94.27% and F-measure of 93.3% respectively. Jahan et al. [23] describes the different approaches which are used for the task of NER. This paper shows some results based on HMM and gazetteer method. Some comparisons are done on the basis of combined approaches. Results are based on the performance metrics and system achieves the F-Measure of 98.37% respectively.

Singh Bajwa et al. [24] developed the mixed approach consist of rule based approach and supervised learning based approach i.e. HMM. Proper nouns are not automatically tagged which directed to the generation of training and testing dataset as no dataset is accessible. The applied system is adequate to recognize different types of entities like name of person, location, date, time, facility, number etc. This paper is based on two different types of interpretations. The first interpretation is based on HMM only and another one is based on the combination of rule based and HMM approaches. Comparison of simple HMM and hybrid approach is done. System based on HMM achieves the recall, precision and F-Score of 76.27%, 72.92% and 74.56% respectively. Comparison of different hybrid techniques used in NER as shown in Table 3:

Table 4: COMPARISON OF DIFFERENT HYBRID TECHNIQUES USED IN NER

Language	Technique	Recall	Precision	F-measure	Reference No.
English-Hindi	CRF's + rule based + dictionary based	50.39%	81.15%	62.17%	[1]
English-Tamil	CRF's + rule based + dictionary based	30.47%	79.92%	44.12%	[1]
Hindi	HMM + rule based	--	--	94.61%	[2]
Kannada	HMM + rule based	94.61%	95.10%	94.85%	[19]
Hindi	CRF's + ME + rule based	84.88%	81.11%	82.95%	[20]
Hindi	ME + rule based	53.69%	82.76%	65.13%	[21]
Bengali	ME + rule based	70.07%	62.30%	65.96%	[21]
Oriya	ME + rule based	39.40%	51.51%	44.65%	[21]
Telugu	ME + rule based	14.91%	25.23%	18.74%	[21]
Urdu	ME + rule based	33.58%	37.58%	35.47%	[21]
Manipuri	CRF's + rule based	92.26%	94.27%	93.3%	[22]
Indian languages(Hindi)	HMM + gazetteer	--	--	98.37%	[23]
Punjabi	HMM + rule based	76.27%	72.92%	74.56%	[24]
Hindi	Rule based + list lookup	--	--	96%	[25]

2.5 Semi supervised Learning Based Approaches applied in the work of NER

Thenmalar et al. [26] developed a semi-supervised based NER approach named pattern based bootstrapping approach. This approach uses the small set of training data. The technique is initialized through the detection of named entities. The context features are designed to determine the pattern through the detected entities. The respective named entities in the testing data are characterized by using already defined pattern of each named entity class. Formation of the naïve patterns to recognize the named entity classes is done by the pattern scoring and tuple value. Survey is done by using two different corpuses: tagged corpus and untagged corpus. The system reported the F-measure of 75% respectively.

Another work based on semi-supervised approach was reported by Liu et al. [27] and developed semi-supervised learning framework with the combination of K-nearest neighbors (KNN) and CRF's based model for the better identification of named entities and to face the threats like inadequate data in the tweets and inaccessible training data. Pre-labeling is done by the KNN based classifier to gather the universal coarse proof across the tweets and sequential labeling is done by CRF's model. Threats like in adequacy of data is resolved by the semi supervised technique and gazetteers. The proposed system reported the F-measure of 78.5% respectively.

Chen et al. [28] proposed a graph based semi-supervised learning based technique named label propagation. The proposed approach shows both the labeled data and unlabeled information. Labeling functions are accessed

through the proposed approach in order to entertain two different kinds of compulsions like considerable function has to be strongly fixed on the labeled nodes and has to be continuous on the entire graph. The proposed system reported the better accuracy than SVM. Comparison of different semi-supervised learning based techniques used in NER as shown in Table 4:

Table 4: COMPARISON OF DIFFERENT SEMI-SUPERVISED BASED LEARNING TECHNIQUES USED IN NER

Language	Technique	F-measure	Reference No.
Tamil	Pattern based bootstrapping approach	75%	[26]
Non-English	Semi-supervised learning framework = KNN + CRF + gazetteers	78.5%	[27]
English	Label propagation	77.7%	[28]

3 CHALLENGES ARE FACED WHILE EXTRACTING NAMED ENTITIES

The challenges in the reinforcement of NER systems for Indian languages emerge due to various aspects.

- 1. Inadequacy of capitalization:** Capitalization hit a vital aspect in detecting proper nouns in English and some other European languages. On the other hand Indian languages do not contain the theory of capitalization [29].
- 2. Limited resources:** Maximum of the Indian languages is deficient in resources. There are no mechanized engines are free to perform preprocessing functions which are essential for NER such as chunking, parts of speech tagging etc. [30].
- 3. Spell diversities:** One of the extensive issues that various people spell the same proper noun distinctly. For example: Tamil is one of the Indian languages contains proper nouns like person name, organization, location, river, percentage, quantity and time etc. and in Tamil person name – Roja is spelt as “Rosa”, “Roza” [30].
- 4. Various ambiguities:** One of the leading challenges in detecting proper nouns is dialect. Another leading challenge is classifying identical words from texts [31].
- 5. Unfamiliar words:** Unfamiliar words are like foreign words and are that word which are not being used regularly these days and are not caught by the lot of people, is another challenge. Words like location name, person’s name [31].
- 6. Indian languages are free word order:** Many of the Indian languages are free word order language i.e. inadequacy of subject –verb –object [32].

- 7. Contention of nested entities:** Nested entities are the extensive challenge while extracting named entities or proper nouns. Nested entities such as NEWYORK UNIVERSITY comprises of two proper nouns which will give the ambiguity while extracting proper nouns in the text [29].
- 8. Various abbreviations:** Most of the words and sentences are dictated in various modes. Abbreviations are being used for the comfort of writing and interpreting. Another extensive challenge is words which frequently want some labels for the detection [31].
- 9. Common nouns vs. proper nouns:** Common nouns consistently appear as proper nouns. For example Priya means lovely in Hindi generate ambiguities between common nouns and proper nouns [29].
- 10. Agglutinative essence:** Agglutination calculates the added features to the root word to generate complicated meaning [29].

CONCLUSION:

NER has evolved tremendously in recent years. NER is not considerable as a solved task but we can take it as a solved task when we have a huge quantity of named entity types and a document assortment. In this paper we briefly reviewed rule based approaches (linguistic and list lookup based), supervised learning based approaches (machine learning based approaches), and hybrid approaches (machine learning based approaches and rule based approaches) for the task of NER. All the proposed approaches and models have tried to improve the accuracy in recognition module and compactness in recognition domain, as mentioned before. This paper also surveys various challenges which are being faced while extracting named entities from the code mixed and informal data and the comparative analysis of different approaches which are used in the task of NER.

REFERENCES:

- [1] D. Gupta, S. Tripathi, A. Ekbal, and P. Bhattacharyya, "A Hybrid Approach for Entity Extraction in Code-Mixed Social Media Data", In Proceedings of the Workshop on Shared Task on Code Mix Entity Extraction in Indian Languages in conjunction with Forum of Information Retrieval Evaluation, 2016.
- [2] D. Chopra, N. Jahan, and S. Morwal, "Hindi Named Entity Recognition by Aggregating Rule Based Heuristic and Hidden Markov Model", International Journal of Information, vol. 2, no. 6, pp. 43–52, 2012.
- [3] R. Sharnagat, "Named Entity Recognition: A Literature Survey", Center For Indian Language Technology, 2014.
- [4] D. Chopra and S. Morwal, "Detection and categorization of named entities in Indian languages using Hidden Markov Model", International Journal of Computational Science and Information Technology, vol. 1, no. 1, pp. 25–32, 2013.
- [5] O. Dhariya, S. Malviya, and U. S. Tiwary, "A Hybrid Approach For Hindi-English Machine Translation", IEEE International Conference of Information Networking, Da Nang, Vietnam, pp. 389–394, 2017.
- [6] R. K. Chakrawarti and P. Bansal, "Approaches for Improving Hindi to English Machine Translation System", Indian Journal of Science and Technology, vol.

- 10, no. 16, 2017.
- [7] P. Kumar, S. Srivastava, and M. Joshi, "Syntax directed translator for English to Hindi language", In Proceedings of IEEE International Conference on Research in Computational Intelligence and Communication Networks, Kolkata, India, pp. 455–459, 2015.
- [8] S. Mathur and V. P. Saxena, "Hybrid approach to english-hindi name entity transliteration", In Proceedings of IEEE Students Conference on Electrical, Electronics and Computer Science, Bhopal, India, pp. 1–5, 2014.
- [9] V. Singhal and N. Tyagi, "A Hybrid Approach Of English-Hindi Named-Entity Transliteration" International Journal of Advanced Technology in Engineering and Science, vol. 3, pp. 1489–1550, 2015.
- [10] D. Chopra and S. Morwal, "Handling Unknown Words In Named Entity Recognition Using Transliteration", International Journal of Advanced Technology in Engineering and Science, vol. 2, pp. 87–93, 2013.
- [11] S. Morwal, D. Chopra and G.N. Purohit, "Named Entity Recognition In Natural Languages Using Transliteration", International Journal on Natural Language Computing, vol. 2, no. 3, 2013.
- [12] A. Nayan, B. R. K. Rao, P. Singh, S. Sanyal, and R. Sanyal, "Named Entity Recognition for Indian Languages", In Proceedings of the International Joint Conference on Natural Language Processing Workshop on Named Entity Recognition for South and South East Asian Languages, Hyderabad, India, pp. 97-104, 2008.
- [13] D. Chopra and G.N. Purohit, "Handling Ambiguities and Unknown Words in Named Entity Recognition Using Anaphora Resolution", International Journal on Computational Sciences and Applications, vol. 3, pp. 456–63, 2013.
- [14] D. Chopra, N. Joshi, and I. Mathur, "Named Entity Recognition in Hindi Using Hidden Markov Model", IEEE Second International Conference on Computational Intelligence & Communication Technology, Ghaziabad, India, pp. 581-586, 2016.
- [15] S. Morwal and D. Chopra, "Identification and Classification of Named Entities in Indian Languages", International Journal on Natural Language Computing, vol. 2, pp. 37–43, 2013.
- [16] D. Chopra and S. Morwal, "Named Entity Recognition In English Using Hidden Markov Model", International Journal on Computational Sciences & Applications, vol. 3, no. 1, 2013.
- [17] A. Ekbal and S. Bandyopadhyay, "Bengali Named Entity Recognition using Support Vector Machine", In Proceedings of the International Joint Conference on Natural Language Processing Workshop on Named Entity Recognition for South and South East Asian Languages, Hyderabad, India, pp. 51-58, 2008.
- [18] A. Ekbal, R. Haque, and S. Bandyopadhyay, "Named entity recognition in Bengali: A conditional random field approach", In Proceedings of the Third International Joint Conference on Natural Language Processing, vol. 2, 2008.
- [19] S. Amarappa and S. V Sathyanarayana, "A Hybrid approach for Named Entity Recognition, Classification and Extraction (NERCE) in Kannada Documents", in Proceedings of International Conference on Multimedia Processing, Communication and Information Technology, 2013.
- [20] S. Srivastava, M. Sanglikar, and D. Kothari, "Named Entity Recognition System for Hindi Language: A Hybrid Approach", International Journal of Computational Linguistics, vol. 2, no. 1, pp. 10–23, 2011.
- [21] S. K. Saha and W. Bengal, "A Hybrid Approach for Named Entity Recognition in Indian Languages", In Proceedings of the International Joint Conference on Natural Language Processing Workshop on Named Entity Recognition for South and South East Asian Languages, Hyderabad, India, pp. 17-24, 2008.
- [22] L. Jimmy and D. Kaur, "Named Entity Recognition in Manipuri: A Hybrid Approach", Language Processing and Knowledge in the Web, Springer, Verlag Berlin Heidelberg, pp. 104-110, 2013.
- [23] N. Jahan, S. Morwal, and D. Chopra, "Named Entity Recognition in Indian Languages Using Gazetteer Method and Hidden Markov Model Hybrid Approach", International Journal of Computer Science & Engineering Technology, vol. 3, no. 12, pp. 621–628, 2012.
- [24] K. S. Bajwa and A. Kaur, "Hybrid Approach for Named Entity Recognition", International Journal of Computer Applications, vol. 118, no. 1, pp. 36–41, 2015.
- [25] Y. Kaur and E. R. Kaur, "Named Entity Recognition (NER) System for Hindi Language Using Combination of Rule Based Approach and List Look Up Approach", International Journal of scientific research and management, vol. 3, no. 3, pp. 2300–2307, 2015.
- [26] S. Thenmalar, J. Balaji, and T. V. Geetha, "Semi-supervised Bootstrapping approach for Named Entity Recognition", CoRR, abs/1511.06833, 2015.
- [27] X. Liu, S. Zhang, F. Wei, and M. Zhou, "Recognizing Named Entities in Tweets", In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL): Human Language Technologies, Portland, Oregon, vol. 1, pp. 359–367, 2011.
- [28] J. Chen, D. Ji, C. L. Tan, and Z. Niu, "Relation extraction using label propagation based semi-supervised learning", In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, Sydney, Australia, pp. 129–136, 2006.
- [29] P. Sharma, U. Sharma, and J. Kalita, "Named Entity Recognition in Assamese: A Hybrid Approach", IEEE International Conference on Advances in Computing, Communications and Informatics, Jaipur, India, pp. 2114–2120, 2016.
- [30] P. R. K. Rao, C. S. Malarkodi, R. Vijay Sundar Ram, and S. L. Devi, "ESM-IL: Entity Extraction from Social Media Text for Indian Languages @ FIRE 2015 – An Overview", In Forum of Information Retrieval Evaluation, 2015.
- [31] A. Mansouri, L. S. Affendey, and A. Mamat, "Named Entity Recognition Approaches and Challenges", International Journal of Advanced Research in Computer and Communication Engineering, vol. 6, no. 2, 2017.
- [32] A. Mansouri, L. S. Affendey, and A. Mamat, "Named Entity Recognition Approaches: A Study applied to English and Hindi Language", IEEE International Conference of Circuit, Power and Computing Technologies, Nagercoil, India, pp. 1–4, 2015.