Survey of Named Entity Recognition Systems with respect to Indian and Foreign Languages

Nita Patil
School of Computer Sciences
North Maharashtra University,
Jalgaon (MS), India

Ajay S. Patil
School of Computer Sciences
North Maharashtra University,
Jalgaon (MS), India

B. V. Pawar
School of Computer Sciences
North Maharashtra University,
Jalgaon (MS), India

ABSTRACT
Named Entity Recognition (NER) is sub task of Information Extraction that includes identification of named entities and classification of them into named entity classes such as person, location and organization etc. NER can be used to preprocess textual information and convert it into structured form that can be useful for Information Retrieval, Machine Translation, Question Answering System and Text Summarization. This paper presents a survey regarding NER research done for various Indian and non Indian languages. The study and observations related to approaches, techniques and features required to implement NER for various languages especially for Indian languages is reported.

General Terms
NER (Named Entity Recognition), HMM (Hidden Markov Model), CRF (Conditional Random Fields), SVM (Support Vector Machine)

Keywords
NER tools, Information Extraction, Machine Translation

1. INTRODUCTION
Information on the web is increasing rapidly. Social networking applications are adding large volumes of information on web which is one important reason of information overload on the web. If a user requests for information from the huge collection of data on the web, the answer to the request is usually present in unstructured data sources such as text and images. Unstructured data is present in the form of spoken text, pictures, video, audio etc. and is computationally opaque. It is impossible for humans to process all data and fulfill request quickly because it is voluminous. Computers are also not able to directly query for the target information because it is not stored in structured format. Information Extraction (IE) helps to handle extraction of required information from huge unstructured collection of data. Information Extraction, the branch of Artificial Intelligence makes the natural language text more suitable for information processing task. IE adds meaning to raw data so that it can be easily processed by computers. IE plays significant role in information retrieval, data mining, machine translation and summarization. Deductive and inductive reasoning is used to build logical rules and inferences by distilling domain knowledge from propositions in text that are useful for text mining and knowledge discovery. IE is significant for extractive summarization which extracts complete document and summarizes it. Extracted information could be useful for information systems only if it is semantically classified, computationally transparent and semantically well defined. Question answering systems pinpoint relevant information by expressing question in natural language whose answers are extracted by the system from the texts in documents. Recognizing entities and semantically meaningful relations between entities is a key to provide focused information access. IE is one of the core technologies that facilitate highly focused information retrieval. Cross language information retrieval system allows query written in one language which is searched in document base in another language. IE combines Natural Language Processing (NLP) for focused information retrieval. NLP deals with processing of linguistic structure of the text. This includes morphological, syntactic, phonetic and semantic analysis of the human language. Subtasks of IE are named entity recognition, noun phrase coreference resolution, semantic role recognition, entity relation recognition, and date and time line recognition. Named entity recognition (NER) is a task that identifies proper nouns in the natural language text and classifies them to appropriate named entity classes. Person, location, organization, date, numbers, measurements are some common named entity classes considered in NER. This paper presents a survey of NER systems implemented for various languages. This paper is organized into four sections. First section discuses origin of the research problem and workshops, conferences and symposium dedicated to NER task. Second section reports tools available for NER, third reports techniques used for implementation of NER systems for non Indian languages. The fourth section reports approaches and techniques used for implementation of NER systems for Indian languages.

2. ORIGIN OF THE NER PROBLEM
Message Understanding Conference (MUC–6) was conducted in 1995 in US was sponsored by DARPA. The task in the conference was extracting company and defense related information in news papers. In this conference the concept extraction of named entities (NEs) and their recognition evolved. In 1998 NE task was independently evaluated for Chinese and Japanese in Multilingual Entity Task (MET). Person, location, organization and numeric were the four entities considered in MET. In 1998-1999 Information Retrieval and Extraction Exercise (IREX) was conducted outside the US. In IREX artifact NE category is added in the evaluation. After that Conference on Computational Natural Language Learning (CoNLL) 2002 and 2003 were conducted for Spanish, Dutch and English, German languages respectively. In shared task of CoNLL-2003 language independent named entity recognition evolved out. Automatic Content Extraction (ACE) was conducted for English. In ACE conference the name entity categories viz., geographical and political entities (GPE) were added. Mainly 7 to 10 basic categories of NEs were used. Automatic annotation systems, dictionaries, rules were developed. Use of supervised learning technique for NER was introduced. In 2004–2008 (HAREM) evaluation contest for named entity recognizers for Portuguese was conducted. Information Retrieval and Extraction Exercise (IREX) was conducted during 1998–1999 for Japanese. The other NER evaluation forums are ACL

Special Interest Group in Chinese (SIGHan), TAC Knowledge Base Population Evaluation (TAC/ KBP), Speech technology evaluation for the French language (ESTER/ETAPE), Evaluation of NLP and Speech Tools for Italian (EVALITA). NER task for south and south East Asian languages was conducted in IJCNLP-08 at IIIT Hyderabad (India) in which five languages; Hindi, Bengali, Oriya, Telugu and Urdu were focused. Table 1 shows the NER evaluation conferences acronyms, domain languages, year and sponsors.

Table 1 NER Evaluation Forums

Conference	Language(s)	Year(s)	Sponsor
MUC	English	1987–1999	DARPA
MET	Chinese, Japanese	1998	US
IREX	Japanese	1998–1999	
ACE	English	2000–2008	NIST
CoNLL	Spanish, Dutch, German, English	2002–2003	
HAREM	Portuguese	2004–2008	Linguateca
SIGHan	Chinese	2006	
EVALITA	Italian	2007, 09, 11	CELCT
IJCNLP	South East Asian	2008	IIIT
TAC	English	2009	NIST
ESTER	French	2009, 2012	AFCP/ISCA

3. NER TOOLS

Various tools as mentioned in Table 2 are available freely for recognition of name entities. Tools are developed by considering some specific languages as a domain for recognition of name entities. Stanford named entity recognizer is JAVA implementation of statistical algorithms based on conditional random fields and maximum entropy. Lingpipe is a toolkit used for computational linguistics based on dictionary lookup and hidden markov model. Yamcha is a generic, customizable, open source text chunker based on support vector machine. Sanchay is an open source platform that uses object oriented architecture with emphasis is on modularity, reusability, extensibility and maintainability. CRF++ is a simple, customizable, and open source implementation of conditional random fields useful for segmenting or labeling sequential data implemented using C++ with Standard Template Library. Mallet is a statistical package for statistical natural language processing which includes tools for named entity extraction based on linear chain conditional random fields.

Table 2: NER tools

NER Tool	Universal Resource Locator
Stanford NER	http://nlp.stanford.edu/software/crf-ner.shtml
Lingpipe	http://alias-i.com/lingpipe/
Yamcha	http://chasen.org/~taku/software/yamcha/
Sanchay	http://sanchay.co.in/
CRF++	http://crfpp.googlecode.com/svm/trunk/doc/
Mallet	http://mallet.cs.umass.edu

Named Entity recognition tools vary in terms of the language they can support. Each language has its own syntax and semantics that may affect the way the entities can be extracted. Frank Landsbergen [1] evaluated NER research and explored work of Palmer. Statistical methods were used for finding named entities in newswire articles for Chinese, English, French, Japanese, Portuguese and Spanish. The researchers reported that significant part of the task could be performed with simple methods but different difficulties are reported in NER for different six languages. The results were affected by low F-measure and an absence of mapping between entities to types. The state-of-the-art NER tools are not useful in practice without significant domain-specific modifications. Some authors have proposed a unit test for NER tools that explores many of the corner cases that cannot be handled by current NER tools [2].

4. NER APPROACHES

The main approaches for development of NER systems are linguistic paradigm based on handcrafted rules development and statistical paradigm based on data driven approaches.

Table 3: NER Development Paradigm

Features	Linguistic Paradigm	Statistical Paradigm
Resources Exhaustion	Well designed & tested language grammar, lexicons, tagset & test corpus	Well annotated training corpus with considerable amount of NEs
Accuracy of the tagger	Considerable time, expertise and efforts results in 95% precision and 99% recall.	Well designed tagset and tagger can disambiguate up to 95 – 97%. [3]
Portability to other domains	Easy to adopt grammar by little correction or improvement in some particular domain.	Taggers accuracy depends upon the coverage for NE's in training corpus for particular domain
Towards 100% output	Non linguistic methods can be used to resolve tagging remained by linguistic tagger.	Difficult for improvement after 97% accuracy.

In linguistic approach rules are designed by grammar expert with help of knowledge derived from language, observations of samples, dictionaries, thesaurus etc.

5. NER FOR FOREIGN LANGUAGES

Work in NER for English started in MUC, ACE and CoNLL. In CoNLL-2003 four NE classes such as Person, Location, Organization and Miscellaneous were considered. Sixteen systems participated in the task. Techniques AdaBoost, Conditional Random Fields (CRF), Hidden Morkov Models (HMM), Maximum Entropy (ME), Memory-Based Learning (MBL), Recurrent Neural Networks (RNN), Support Vector Machine (SVM), System Combination, Transformation-Based Learning (TBL), Voted Perceptrons etc. were used for NER task. In shared task of NER lexical information, part of speech tags, affix information, previous NE tags, orthographic information, gazetteers, chunk tags, orthographic patterns, global case information, trigger words, bag of words, quote information, global document information etc. features were used [4]. Hercules et. al developed NE tagger for Swedish using 1,08,000 news articles in training annotated by 100 NE categories. NER system developed using mixed approach by combining rules, lexicons and training strategies obtained 92% precision and 46% recall [5]. Guo Dong Zhou et. al. proposed HMM based chunk tagger to recognize names, times, numbers and quantities using internal, external NE evidences, capitalization-digitalization features, triggers, internal gazetteers and external macro context features for English obtained F-measure of 96.6% on MUC-6 data; training data (1320), held out development data (121) and held out test data (124)[6].

(Hwang et.al. 2003) gathered 68,000 person, 25,000 location and 10,000 organization names for constructing an IE (Independent Entity) dictionary, 92 location, 121 organization

names for constructing CE (Constituent Entity) dictionary and 114 person, 39 location and 33 organization names to construct AE (Adjacent Entity) dictionary [18]. Muntsa et. al. presented NER system using finite automata acquisition based on causal state splitting reconstruction algorithm. Authors have reported F-measure 89.01% on development and 89.42% on test data [7]. Wu et.al. designed a Chinese Name Entity tagger using character based model since Chinese words do not contain space and every character is meaningful. They used CityU & MSRS Chinese corpus and Maximum Entropy, CRF classifier, majority vote and memory based learner methods to combine results of the classifiers. This work indicates that the memory based methods can outperform the individual classifiers [8].

Bart Desmet used ensemble classifier based on Memory Based Learning (MBL), CRF and SVM trained using eight different features for Dutch, used genetic algorithm and received significant marginal F-Score[9].

Michailidis et.al. developed NER for Greek using three algorithms SVM, ME and onetime using 400 news articles consisting of 172,000 tokens. Results obtained by each algorithm were compared. SVM performed best among all with greater precision [10]. Duarte et. al. proposed NER using machine learning techniques HMM, TBL and SVM for Portuguese. 2100 sentences were annotated and preprocessed using tagging conventions. Annotated corpus consists of 3,325 NEs. It has been proved the approach that uses SVM gives better performance i.e. 88.11% F-score [11]. Louis et.al described probabilistic NER system based on gazetteers and Semantic features to classify NEs for South African language. Name gazetteer contains 5,930 names and surname gazetteer contains 90,221 surnames. NER using gazetteer and syntactic features based on Bayesian network improved performance ranges from 53.1% to 77.6% [12]. Mehdad et. al. developed a NER system based on YAHCHA classifier using SVM. The system uses 525 news stories from news paper as development data consisting of 180,000 words [13]. Padro et.al presented NER system for Spanish based on finite automata acquisition algorithm based on CSSR algorithm. The system obtained 89.01% F-score for development data and 89.42% for test data [7].

Table 4: NER for non-Indian languages

Year	Language	Algorithm(s)	F-Score (%)
2000	English	MaxEnt, HMM, handcrafted rules [19]	93.39
2001	Swedish	Hybrid[5]	59.00
2003	Korean	HMM[18]	83.80
2004	Thai	MaxEnt[20]	89.87
2005	Spanish	Finite Automata Acquisition[7]	67.15
2006	Chinese	ME, CRF[8]	88.72
2006	Hungarian	SVM, Artificial NN, C4.5 decision tree[17]	93.32
2006	South African	Dynamic Bayesian Network[12]	70.00
2006	Greek	SVM,ME[10]	91.06
2007	Portuguese	HMM, SVM & TBL[11]	88.11
2008	Serbian	Morphological processing [15]	
2008	Japanese	SVM + Viterbi [16]	87.72
2009	Italian	SVM[13]	81.09
2010	Dutch	SVM,CRF,MBL[9]	83.77
2010	Tibetan	Case-auxiliary Grammars[22]	86.91
2011	Turkish	CRF[23]	88.71
2012	Arabic	rule-based, decision-tree classifier[24]	88.77
2014	Nepali	HMM & rules[25]	85.15

Vitas and Lazetic analyzed NER for Serbian using lexical recognition based on morphological dictionary. Geographic entities from the dataset containing 10,000 entities in Serbia and Montenegro, 50,000 entities of Yugoslav and 1,00,000 from other regions were selected. The size of geographic name dictionary was 400 lemmas and dictionary of forms contain 40,000 words and dictionary of proper names contain 500 words. It is proved that retrieval performance depends upon the lexical resources describing the lexical fund[15]. Sasano & Kurohashi presented an approach that uses structural information like cache features, coreference relations, syntactic features and case frame features based on SVM for Japanese NER. CRF is trained with 18,677 NEs from 174 articles in Mainichi Newspaper, IREX formal test data with 1,510 NEs from 71 articles and Web NE data with 1,686 NEs from 354 articles. It was observed that the structural approach improved performance of the system [16]. Farkas & Szarvas have introduced multilingual NER using statistical modeling techniques for Hungarian text using Support Vector classifier, Artificial Neural Networks and c4.5 Decision Tree learning algorithm. The system has achieved 93.59% as a best F-measure at term level and 90.57% at phrase level evaluation [17]. Table 4 shows NER systems developed for various non-Indian languages, the techniques used to develop NE taggers and F-Score obtained.

6. NER FOR INDIAN LANGUAGES

Detection of NEs in raw information is not easy in Indian languages because Indian languages do not have capitalization. Indian languages have highly phonetic characteristics. Resources like gazetteers, dictionaries, POS taggers [14], morphological analyzers are not easily available. Lot of variations exists in spellings writing style [26]. Work on NER in Indian languages is a difficult and challenging task and also limited due to scarcity of resources, but it has started to appear [27].

A survey made by Shashidhar et.al [28] points out that research for NER on Indian languages is difficult because of different writing methodologies, writing style variations, difficult morphology, little availability of annotated corpora and agglutinative nature like in Telugu. Many researchers have concluded that rule based approach for NER gives satisfactory results with sufficient gazetteers list and language independent rules. Rule based approach is not very easy for NER system development in Indian languages and therefore language independent NER system using hybrid models is needed. Srikanth and Narayana have developed CRF based noun tagger, trained on manually tagged data of 13,425 words and test dataset of size 6,223 words. 92% of F-Score have been given by name tagger [29]. Raju et.al. described Telugu NER based on ME by using news articles form Eenadu Vaartha newspaper and data from Telugu Wikipedia using the roman forms of the articles. The system is evaluated with and

without using name list with ME and observed that ME using name list performs best for NER [30]. NER system development preferred news articles because news is rich source of NEs of almost all categories. The work presented by Ekbal and Bandyopadhyay [31] mentioned a useful technique to develop tagged Bengali news corpus from web for Bengali NER. Nayan et.al. used phonetic matching algorithm Editex and Fuzzy string matching technique Soundex to recognize NEs in Hindi. The system has reported 81%precision. It is observed that large set of annotated data is yet to be available for Indian Languages [32]. A novel NER approach which combines the global distributional characteristics with local context based on MEMM was presented by Gupta and Bhattacharya [33]. A hybrid machine learning approach by using MaxEnt and HMM was presented by Biswas et.al. [34] for NER in Oriya. 32 different rules were developed to identify numbers, measures and time. Gazetteers of specialized names were developed by translation into Oriya.

Table 5: NER for Indian Languages

Language	Year	Technique/Algorithm	F1
Hindi	1999	Morphological & contextual clues[38]	41.70
	2003	CRF, Feature Induction[39]	71.50
	2006	MEMM[40]	79.70
	2009	Editex, Soundex[32]	65.69
	2009	MaxEnt[41]	82.66
	2010	CRF[42]	78.29
	2010	CLGIN[33]	82.90
	2010	SVM[43]	77.17
	2011	CRF, MaxEnt, Domain Rules[44]	80.82
	2011	SVM [44]	80.21
Bengali	2007	HMM[45]	83.79
	2008	CRF[46]	90.70
	2009	MaxEnt[41]	85.22
	2009	CRF[42]	81.15
	2010	SVM[43]	84.15
	2011	SVM[45]	83.39
Telugu	2008	CRF + Majority Tag[29]	84.49
	2010	MaxEnt[30]	67.07
	2011	Survey[28]	-
Oriya	2010	MaxEnt + HMM [34]	84.08
Punjabi	2011	condition based list lookup[47]	86.25
	2012	Domain rule, list look up[48]	85.88
Tamil	2008	CRF[49]	80.44
	2008	E-M(HMM)[50]	72.72
Urdu	2010	MaxEnt[51]	74.67
Nepali	2014	HMM + Rule based[25]	85.15

Bhattacharya et. al. [35] developed hand-crafted rule based named entity recognizer for Marathi. The rules were constructed for extracting instances of NE classes using TILDE and WARMR techniques of inductive logic programming. TILDE is extension of traditional c4.5 decision tree learner to first order logic and WARMR is an extension of apriori algorithm to first order logic. Authors have used tagged data of 3,884 sentences in Marathi and 27,748 sentences in Hindi. NER system developed by using GATE (a framework and graphical development environment which enables users to develop and deploy language engineering components and resources in a robust fashion) [36]. Vasudev Verma et. al. proposed an approach to identify the NEs present in under resourced languages by utilizing the NEs present in English. Bisecting k-means algorithm is performed for clustering multilingual documents based on the identified NEs [37]. Table 5 shows NER work done for some Indian languages, year of publication, techniques used and F-score obtained. Many NER systems observed here are implemented using more than one technique and evaluated with more than one dataset. F-score value in table 4 and 5 is the best reported performance of that respective system. Some NER systems reported more than one F-score, in case of such systems the average of F-scores is presented in this survey. The performance of NER system is measured by metrics precision, recall and F-measure. Precision measures how many of the tokens tagged are tagged correctly. Recall measures how many of the tokens are tagged are indeed tagged [1]. F-Score is harmonic mean of precision and recall.

7. CONCLUSION

This paper has presented a literature review of named entity recognition and classification for Indian and non Indian languages. Significant NER work has been done for non Indian languages whereas NER work is in progress for Indian languages. Issues such as unavailability of annotated corpus, lack of capitalization feature, variations in writing style, difficult morphology, use of foreign words in text, free order and agglutinative nature makes named entity recognition a very challenging task for Indian languages. It is evident from the review that many authors have implemented NER systems using linguistic, machine learning or hybrid approaches. Multiple statistical techniques or combination of linguistic and statistical techniques are used for comparing results. It is observed that rule based approaches with some language independent rules and gazetteer lists combined together with statistical approach gives satisfactory results. It is found that combination of linguistics and statistical techniques is better combination to perform named entity recognition in Indian languages. Very less work on NER is reported for Indian languages like Marathi and Gujrathi is reported. Development of appropriate techniques, methods for NER for such languages is necessary.

8. ACKNOWLEDGEMENT

This research work is supported by grants provided to the School of Computer Sciences, North Maharashtra University, Jalgaon (MS), India under SAP-DRS (I) scheme of UGC, New Delhi.

9. REFERENCES

[1] Frank Landsbergen, Evaluation of Named Entity Work in IMPACT: NE Recognition and Matching, Technical Report, 2012.

[2] Robert Krovetz, Paul Deane and Nitin Madnani, "The Web is not a Person, Berners-Lee is not an Organization, and African-Americans are not Locations: An Analysis of the Performance of Named-Entity Recognition." in Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World (MWE 2011). Association for Computational Linguistics, Stroudsburg, 2011, PA, USA, pp. 57-64.

[3] Hans Van Halteren, "Syntactic Wordclass Tagging (Text, Speech, and Language Technology)", Springer, 1999.

[4] Language-Independent Named Entity Recognition, http://www.cnts.ua.ac.be/conll2003/ner/

[5] Dalianis, Hercules, and Erik Åström. SweNam—A Swedish Named Entity Recognizer. Technical Report.

Department of Numerical Analysis and Computing Science, Sweden <ftp.nada.kth.se/IPLab/TechReports/ IPLab189.pdf>, 2001.

[6] GuoDong Zhou, Jian Su, "Named Entity Recognition using an HMM-Chunk Tagger", Proceedings of 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, 2002, pp. 473-480.

[7] Muntsa Padro and Lluis Padro, "Named Entity Recognition System based on a Finite Automata Acquisition Algorithm", Journal Natural Language Processing, Vol. 1 No. 35, pp. 319 - 326, 2005.

[8] Chia-Wei Wu, Shyh-Yi Jan, Tzong-Han Tsai, Wen-Lian Hsu, "On Using Ensemble Methods for Chinese Named Entity Recognition", Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing, Sydney, July 2006, pp. 142–145.

[9] Desmet, Bart, and Véronique Hoste. "Dutch Named Entity Recognition using Classifier Ensembles." LOT Occasional Series 16, 2010, pp. 29-41.

[10] Ionas Michailidis, Konstantinos Diamantaras, Spiros Vasileiadis, Yannick Frere, "Greek Named Entity Recognition using Support Vector Machines, Maximum Entropy and Onetime", in Proceedings of the 5th International Conference on Language Resources and Evaluation, 2006, pp. 47–52.

[11] Julio Cesar Duarte, Ruy Luiz Milidiu, "Machine Learning Algorithms for Portuguese Named Entity Recognition", Journal of Artificial Intelligence Revista Iberoamericana", 2007, pp. 67-75.

[12] Anita Louis, Alta De Waal and Cobus Venter, "Named Entity Recognition in a South African Context", In Proceedings of SAICSIT 2006, pp 170-179.

[13] Yashar Mehdad, Vitalie Scurtu, Evgeny Stepanov, "Italian Named Entity Recognizer", in EVALITA 2009 Workshop, XIth International Conference of the Italian Association for Artificial Intelligence", Italy, 2009

[14] H B Patil, A S Patil and B V Pawar (2014) "Part-of-Speech Tagger for Marathi Language using Limited Training Corpora", International Journal of Computer Applications Proceedings on National Conference on Recent Advances in Information Technology NCRAIT Vol. 4, pp. 33-37.

[15] Dusko Vitas and Gordana Pavlovic Lazetic, "Resources and Methods for Named Entity Recognition in Serbian", In INFOTHECA-Journal of Informatics and Librarianship, Ng 1-2, vol. IX, p35a-42a May 2008.

[16] Sasano R, Kurohashi S, "Japanese Named Entity Recognition Using Structural Natural Language Processing", in Proceedings of IJCNLP 2008,pp. 607-612, 2008

[17] Richard Farkas, Gyorgy Szarvas, "Statistical Named Entity Recognition for Hungarian: Analysis of the Impact of Feature Space Characteristics", in Proceedings of CESCL 2006, Budapest, Hungary, 2006

[18] Hwang, Yi-Gyu, Eui-Sok Chung, and Soo-jong Lim. "HMM based Korean Named Entity Recognition." Organization 24, no. 11.3 (2003): 4-0.

[19] Srihari, Rohini, Cheng Niu, and Wei Li. "A Hybrid Approach for Named Entity and Sub-type Tagging." in Proceedings of the Sixth Conference on Applied Natural Language Processing, Association for Computational Linguistics, 2000, pp. 247-254.

[20] Chanlekha, Hutchatai, and Asanee Kawtrakul. "Thai Named Entity Extraction by Incorporating Maximum Entropy Model with Simple Heuristic Information", in Proceedings of the IJCNLP. 2004.

[21] David Nadeau and Satoshi Sekine, "Survey of Named Entity Recognition and Classification", Journal of Linguisticae Investigationes, Vol. 30, No. 1, 2007

[22] Hongzhi Yu, Tao Jiang and Ning Ma, "Named Entity Recognition for Tibetan Texts Using Case-auxiliary Grammars", In Proceedings of the International MultiConference of Engineers and Computer Scientists , Vol. I, IMECS March 2010, Hong Kong

[23] Yeniterzi, Reyyan. "Exploiting Morphology in Turkish Named Entity Recognition System", in Proceedings of the ACL 2011 Student Session, Association for Computational Linguistics, 2011

[24] Abdallah, Sherief, Shaalan, Khaled,Shoaib, Muhammad, "Integrating Rule-Based System with Classification for Arabic Named Entity Recognition" , in Computational Linguistics and Intelligent Text Processing Lecture Notes in Computer Science Vol. 7181, 2012, pp. 311-322

[25] Arindam Dey, Abhijit Paul, Bipul Syam Purkayastha," Named Entity Recognition for Nepali language: A Semi Hybrid Approach". International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 8, February 2014 pp. 21-25

[26] N. V. Patil, H. B. Patil, A. S. Patil and B. V. Pawar, "The State-of-the-Art of Named Entity Recognition for Natural Language Processing", National Conference on Emerging Trends in Computer Science and Computer Applications. Organized by DES's Fergusson College, Pune, on 7th–8th Dec. 2013 pp 1-8.

[27] Shilpi Srivastava, Mukund Sanglikar & D.C Kothari, ," Named Entity Recognition System for Hindi Language: A Hybrid Approach", International Journal of Computational Linguistics (IJCL), Volume (2) : Issue (1) : 2011 pp.10-23

[28] B. Sasidhar, P. M. Yohan, Dr. A. Vinaya Babu, Dr. A Goverdhan, "A Survey on Named Entity Recognition in Indian Languages with Particular Reference to Telugu", In IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, ISSN:1694-0814, 2011

[29] Srikanth P and Narayana Murthy Kavi,"Named Entity Recognition for Telugu", Proceedings of IJCNLP 2008,Workshop on NER for South and South East Asian Languages, IIIT, Hyderabad, India, 2008

[30] G. V.S. Raju, B. Shrinivasu, Dr. S. Viswanadha Raju and K. S. M. V. Kumar, "Named Entity Recognition for Telugu using Maximum Entropy Model", Journal of Theoretical and Applied Information Technology, 2005-2010.

[31] Ekbal, Asif, and Sivaji Bandyopadhyay, "Development of Bengali Named Entity Tagged Corpus and its Use in NER Systems." IJCNLP, 2008, pp. 1-8.

[32] Animesh Nayan, B. Ravi Kiran Rao, Pawandeep Singh, Sudip Sanyal and Ratna Sanyal, "Named Entity Recognition for Indian Languages", in Proceedings of the IJCNLP-08 Workshop on NER for South and South East Asian Languages, Pages 97-104, Hyderabad, India, 2008

[33] Shalini Gupta, Pushpak Bhattacharyya, "Think Globally, Apply Locally: Using Distributional Characteristics for Hindi Named Entity Identification" Proceedings of the 2010 Named Entities Workshop, ACL 2010, pages 116-125 Uppsala, Sweden, 2010

[34] Sitanath Biswas, S. P. Mishra, S Acharya and S Mohanty, "A Hybrid Oriya Named Entity Recognition System: Harnessing the Power of Rule", International Journal of Artificial Intelligence and Expert Systems (IJAE),Vol.1: Issue 1, 2010 pp.1-6

[35] Anup Patel, Ganesh Ramkrishana and Pushpak Bhattacharya, "Incorporating Linguistic Expertise using ILP for Named Entity Recognition in Data Hungry Indian Languages", in Proceedings of the 19th International Conference on Inductive Logic Programming ILP'09, 2009, pp 178-185.

[36] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "Gate: An Architecture for Development of Robust HLT Applications," in Recent Advances in Language Processing, 2002, pp. 168-175.

[37] N Kiran Kumar, GSK Santosh, Vasudeva Varma, "A Language-Independent Approach to Identify the Named Entities in Under Resourced Languages and Clustering Multilingual Documents", International Conference on Multilingual and Multimodal Information Access Evaluation (CLEF- 2011), pp 74-82.

[38] Cucerzan, Silviu, and David Yarowsky. "Language Independent Named Entity Recognition Combining Morphological and Contextual Evidence." Proceedings of the 1999 Joint SIGDAT Conference on EMNLP and VLC. 1999, pp. 90-99.

[39] Li, Wei, and Andrew McCallum, "Rapid Development of Hindi Named Entity Recognition Using Conditional Random Fields and Feature Induction", ACM Transactions on Asian Language Information Processing (TALIP) Vol. 2 No. 3, 2003, pp. 290-294.

[40] Kumar N. and Bhattacharyya Pushpak "Named Entity Recognition in Hindi using MEMM." In technical report IIT Bombay, 2006.

[41] Mohammad Hasanuzzaman, Asif Ekbal and Sivaji Bandyopadhyay, "Maximum Entropy Approach for Named Entity Recognition in Bengali and Hindi", International Journal of Recent Trends in Engineering, Vol. 1, No.1, May 2009.

[42] Ekbal, Asif and Bandyopadhyay, Sivaji, "A Conditional Random Field Approach for Named Entity Recognition in Bengali and Hindi", Linguistic Issues in Language Technology", Vol. 2, No. 1 November, 2009, pp. 1-44.

[43] Ekbal, Asif, and Sivaji Bandyopadhyay. "Named Entity Recognition Using Support Vector Machine: A Language Independent Approach", International Journal of Electrical and Electronics Engineering Vol. 4 No. 2 2010, pp. 155-170.

[44] Ekbal, Asif and Bandyopadhyay, Sivaji, "Named Entity Recognition in Bengali and Hindi Using Support Vector Machine", Journal of Lingvisticae Investigationes, Vol. 34, No. 1, 2011, pp. 35-67.

[45] Ekbal, Asif, Naskar, Sudip Kumar; Bandyopadhyay, Sivaji Named Entity Recognition and Transliteration in Bengali". Journal of Lingvisticae Investigationes Vol. 30, No. 1, 2007, pp. 95-114

[46] Ekbal, Asif, Rejwanul Haque, and Sivaji Bandyopadhyay. "Named Entity Recognition in Bengali: A Conditional Random Field Approach", IJCNLP. 2008.

[47] Vishal Gupta, Gurpreet Singh Lehal. "Named Entity Recognition for Punjabi Language Text Summarization" International Journal of Computer Applications (0975–8887) Vol. 33 No. 3, November 2011, pp. 28-32

[48] Kamaldeep Kaur, Vishal Gupta,"Name Entity Recognition for Punjabi Language", International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 2, No.3, June 2012, pp.561-567

[49] Vijayakrishna R and Sobha L., "Domain Focused Named Entity Recognizer for Tamil Using Conditional Random Fields", in Proceedings of the IJCNLP-08Workshop on NER for South and South East Asian Languages, Hyderabad, India. pp. 93–100,

[50] S. Lakshmana Pandian , Krishnan Aravind Pavithra , T. V. Geetha. Hybrid Three-stage Named Entity Recognizer for Tamil," INFOS2008 (2008), March 27-29, 2008 Cairo, Egypt, http://infos2008.fci.cu.edu.eg/infos/NLP_08_P045-052.pdf

[51] Smruthi Mukund, Rohini Shrihari and Erik Peterson,"An Information- Extraction System for Urdu- A Resource Poor Language", ACM Transactions on Asian Language Information Processing, Vol. 9, No. 4, Article 15, 2010.