IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-3, March 2019 E-ISSN: 2347-2693

Parts of Speech Tagging for Indic Languages: A Survey

Floyd Avina Fernandes[1*], Kavita Asnani[2]

[1,2]Dept. of Computer Science, Goa College of Engineering, Goa University, Goa, India

[*]*Corresponding Author:floyd6555@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i3.729735 | Available online at: www.ijcseonline.org

Accepted: 13/Mar/2019, Published: 31/Mar/2019

Abstract—Natural language processing (NLP) comprises of various techniques addressing language text. Few to mention are Part of Speech (POS) tagger, Chunker, Morphological Analyzer, Spell-Checkers, Grammar Checkers, Machine translator, Transliterator etc. POS tagging is the basic building block in language processing which assigns part of Speech (POS) tag which is a peculiar label assigned to each and every token (word) in a text corpus to indicate the part of speech such as verb, pronoun, noun, adjective etc. POS tagging is useful and significant in pre-processing phase especially in the area of information retrieval, text to speech processing, word sense disambiguation and information processing. The methods of POS tagging are classified as rule-based POS tagging, transformation-based tagging, and stochastic tagging. Recent research reports various methods and approaches like Markov Model (MM), SVM (Support Vector Machine), ME (Maximum Entropy) etc used for POS tagging tested on several Indic languages like Hindi, Bengali, Manipuri, Assamese, Telugu, Kannada, Malayalam, Tamil, Punjabi. Since the performance of POS taggers is specific to context and language, there is a pressing need to carry out exhaustive survey. . This paper highlights a comprehensive study on two indic languages i.e. Hindi and Bengali. POS taggers with various approaches along with performance are reported.

Keywords: Ambiguity, Natural language processing (NLP), Named Entity Recognition (NER), Part of Speech(POS), Tagger

I. Introduction

India is a multilingual country with diverse cultures. It has thousands of spoken and written languages. 22 different languages are designated by the Constitution of India. The languages can be categorized into two linguistic categories i.e. Indo Aryan and Dravidian. These languages have important differences i.e. the way of developing the word and grammar is different, as it includes a lot of Sanskrit words. In addition, both have the same phraseology and construction. POS tagging plays a crucial role in tagging, each and every word according to the context. It gives the best grammar label/tag to a word which gives it a syntactic category like verb, adjective, adverb, preposition, conjunction etc. The challenging task is that the words in natural language represent more than one category. The tagger will create an annotated corpus as this annotated corpus is an initial step of information extraction, summarization, information retrieval, machine translation, speech recognition [1]. The work that has been carried out on Indian languages was rule based approach and the other was empirical based POS tagging approach. But rule-based approach requires proper linguistic language knowledge and hand written rules [2]. After rule based approach, researchers moved to stochastic based approach. Stochastic methods require large corpora to become effective. The main key issue for Indian languages is ambiguity. It is very tedious and time consuming process to assign a correct POS tag to different context words as different context words behave differently. Hence, it is a challenging problem for study in the field of NLP, to identify correct tag to a particular context. As this paper performs a detailed study of POS taggers based on the performance for the two languages.

The organization of paper is done in 4 sections. Section II gives related work carried out on POS tagging for the languages. Section III gives an insight into the techniques for POS tagging. Section IV gives the results. Section V concludes POS tagging.

II. Related Work

Research study is being carried out across the globe towards building of POS taggers for languages. Western languages have annotated corpora in a large volume and have been tested on many machine learning techniques. The accuracy ranges from 93-98% approximately [1]. Finite availability of annotated corpus for Indian languages is a challenging task.

II.I POS Taggers for Hindi Language

POS tagging methodology for resource poor languages is proposed by [3] . Here an annotated corpora of 15,562 words was used for higher coverage of lexicon and a decision tree based learning algorithm (CN2) in morphological analysis. Lexicon lookup was used by the system for identifying the other POS categories. 4-fold cross validation of the corpora was evaluated on the system with the accuracy of 93.45%.

Conditional random field (CRF) approach for Hindi language is proposed by [4]. It adopts morphological analyzer to provide information on root words and POS tags for training purpose. The evaluation of the system was done over a corpus of 21,000 words with 27 different POS tags and the system achieved the accuracy of 82.67%.

Maximum Entropy Markov Model (ME) for Hindi language is proposed by [5]. Annotated Hindi corpus is trained by the system and tags are assigned to unseen text. It depends upon a feature function which seizes the lexical and morphological feature of language and feature set is achieved after an in-depth analysis. Evaluation of the system was done over a corpus of 15,562 words with 27 different POS tags and the accuracy achieved was 94.81%.

Hidden Markov Model (HMM) for Hindi language is proposed by [6]. Its objective is to establish without making the use of tools like morphological analyzer or resources such as pre-compiled structured lexicon, as it equips the morphological richness of Indian languages. The roots of the words are found out by making use of naive stemmer as a preprocessor. Accuracy of the system achieved was 93.12% over 18 different POS tags.

The HMM based approach was determined to utilize the morphological richness of the languages without resorting to complex and expensive analysis [7]. The source idea of this approach was exploding the input so that it increases the length of the input and then encountering number of unique types during learning. This idea will increase the probability score of the correct choice and also decrease the ambiguity of the choices at every stage. Sparse data also decreases by new morphological forms for known base words. Evaluation on training and testing was performed with an exploded corpus size of 81,751 tokens which was divided into 80% and 20% parts respectively.

POS tagging for Hindi corpora is proposed by [8]. System scans the Hindi corpora and then abstracts sentences and words. The system inspects for the tagged pattern from database and displays the tag of each Hindi word like noun tag, adjective tag, number tag, verb tag etc.

A rule based approach to POS tagging is proposed by [9]. Hindi corpus is learned by the system and sentences are split into words according to the delimiter. The system searches the words in the database and assigns the appropriate tag to the words.

A rule-based POS tagger for Hindi is proposed by [10] . The testing of the system was done on the various domains of Hindi Corpora. Corpora of 26,149 words with 30 different POS tags achieved an accuracy of 87.55%.

II.II POS Taggers for Bengali Language

A substantial amount of research has been already done in POS tagger developments for Bengali language using different approaches.

HMM POS tagger, which takes input of a raw Bengali text and outputs POS Bengali tagged text, is proposed by [11]. Adapting to machine learning technique supervised Bengali trigram POS tagger was implemented. Bigram POS tagger was the baseline tagger for trigram tagger. Corpus of 895 words with 26 different POS tags achieved the accuracy for trigram tagger is 78.68% and bigram tagger is 74.33%.

Bigram Hidden Markov Model (HMM) i.e. a supervised and semi supervised and Maximum Entropy (ME) model is proposed by [12]. HMM-S is using supervised HMM model parameters and HMM-SS the uses semi supervised model parameters. Annotated corpus of about 40,000 words was used for supervised HMM and ME model. 5000 words are used for testing a set of randomly selected all three cases and the results showed that, the supervised learning model performs better over other models.

Named Entity Recognition (NER) system using Support Vector Machine (SVM) is proposed by [13]. The main goal of NER was to arrange each word to NE classes of predefined target. Words with the different features lead to predict the various named entity (NE) classes. 150K words which were manually annotated with the sixteen NE tags were used to train the system. Average Recall, Precision and F-Score of 94.3%, 89.4% and 91.8%, respectively were achieved for SVM based NER system.

An unsupervised POS tagger for the Bangla language, based on a Baum-Welch is proposed by [14]. Baum-Welch was trained on HMM approach and Brill tagger. The main objective was to test whether the phenomenon of rule based taggers is working better than stochastic taggers.

III. SEVERAL TECHNIQUES FOR POS TAGGING

POS taggers are widely classified into 3 classes i.e. rule based, empirical based and neural based. In rule based approach, rules are hand-written which will extricate the ambiguity of the tag. The empirical based approach is divided into stochastic taggers with HMM, maximum entropy, conditional random field, sector vector machine. Stochastic taggers are of 2 types i.e. supervised and unsupervised taggers. The Fig. 1 shows the approaches for parts of speech tagging.

III.I Supervised Models

The supervised approach to POS tagging requires human intelligence in the domain, for corpus that has been hand annotated by annotators. This is called as the training corpus. Therefore, training corpus will learn information about the tag set, word-tag frequencies, rules etc [14]. In supervised approaches, performance depends on the quality and size of annotation in training corpus.

1) Rule based POS tagging

In rule based POS tagging model, rules are hand written and human intelligence is used to assign appropriate tags to words in the training corpus. Grammatical knowledge and good experience are required to achieve the best results with the use of this method. The rules used in this method are called context frame rules. English POS-tagger is Brill's tagger based on rule-based approach .Its cost is high [2].

1. 1) Brill tagger

An effective tagger implemented for English and several other languages performed good results. Only drawback is that it requires a human-annotated corpus or set of rules [14].

2) Empirical Based POS tagging Approach

Due to the failure of rule-based approaches, there is a huge availability of machine readable text and thus increase in capability of hardware (CPU, memory, disk space) which leads to decline in cost so, researchers adopt to corpus based pos tagging. Empirical approach of parts of speech tagging is stochastic based approach [15] .

2.1) Stochastic based POS tagging

The stochastic approach uses a training corpus to pick the most probable tag for a given word on the basis of statistics i.e. frequency or probability [14] . It applies a set of rules for a specific word in the annotated training data. And then, the same information is used to tag that word in the unannotated text. The disadvantage of this approach is that it might yield a correct tag for a given word but it could not yield invalid sequences of tags [2]. The various stochastic approach methods are like n-grams, Maximum-Likelihood Estimation, (MLE) or Hidden Markov Models (HMM), Support Vector Machines (SVM), Conditional Random Fields (CRF). In order to train the corpus a large sized corpus is required for stochastic approach.

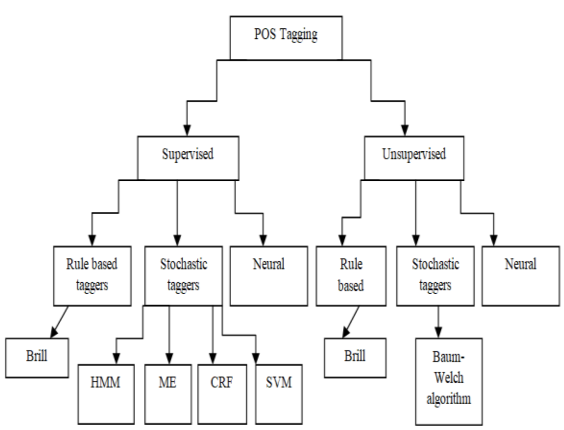

Figure 1. Various Techniques for POS tagging

2.1.1. Hidden Markov Model (HMM) based POS tagging

It measures the probability or frequency of a given catenation of tags. With the probability obtained for the most probable tag , there exists for each word or token of a sentence with n previous tags, where the value of n is set to 1, 2 or 3 for practical purposes [2]. The apt algorithm for implementation of an n-gram approach is the HMM's Viterbi Algorithm which tags new text.

2.1.2. Support Vector Machines Approach

SVM is a machine learning algorithm which has been applied to Natural Language Processing (NLP) and binary classification. SVM approach is used because it is simple, flexible, robust, portable and computationally very efficient as it meets all the requirements of modern NLP technology [2] .

2.1.3. Maximum Entropy Markov Model

MaxEnt stands for Maximum Entropy Markov Model (MEMM) [16] . It is called as a conditional probabilistic sequence model. As this model is used to represent numerous features of a word or token and can handle long term dependency. It is based on the principle of maximum entropy which states that the least biased model is the one which maximizes the entropy on the basis of all known facts [2] . The input to every source state for an exponential model takes the observation feature and the obtained output is in the form of distribution over next possible states.

2.1.4. Conditional Random Field Model

CRF stands for conditional random field. This model is called as discriminative probabilistic model. It bypasses the label bias problem and is similar to MEMMs. CRFs models are graphical models that are undirected and are used for calculation of conditional probability of values assigned to output nodes where by the values assigned to other assigned input nodes [17].

2.1.5. Transformation-based POS tagging Approach

In supervised tagging approach, a large size of pre-annotated corpus is needed but, in transformation –based tagging, it does not require any pre-annotated corpus. In this approach, in order to generate initial output, an untagged text is run through a tagging model. This is one approach for automatic rule induction after getting the output error correction is done. Two sets of data are compared by learning the correction rules. This process is repeated n number of times to achieve best results [15].

3) Neural Tagger
On the neural networks, neural taggers are located. From a training dataset, it learns the parameters of POS tagger [1]. The performance of neural taggers is better as compared to stochastic taggers.

III.II Unsupervised Models
The unsupervised POS tagging models does not require a pre-annotated corpus [14]. The Baum-Welch algorithm is used to determine the transformation rules automatically as they are advanced computational techniques. Once information is obtained, it generates the markov model required by stochastic taggers or the rule-based or transformation-based systems to produce the contextual rules.

IV. RESULTS

Performance of POS tagger is computed by comparing all the POS tagging approaches for Hindi and Bengali languages. An exhaustive study was made on all the approaches for respective languages.

Table1. Performance of POS tagging approaches for Hindi Language

APPROACH	NO. OF TAGS	TESTED DATA	ACCURACY	REFERENCE
Learning based (LB) tagger after 4-fold cross validation		15,562 words	93.45%	Smriti Singh et al.,(2006)
Maximum Entropy (ME)	27	15,562 words	94.81%	Aniket Dalal et al., (2006)
			75.69%	Sandipan Dandapat(2007)
	27	21000 words	78.96 %	Himanshu Aggarwal et al., (2006)
Conditional Random Fields (CRF)	27	15562	90.3%.	Aniket Dalal et al., (2006)
CRF using CRF++	27	21000 words	82.67%	Himanshu Aggarwal et al., (2006)
CRF+MA			82.67%	Agarwal et al.,(2006)
CRF+TBL			78.66%	Avinesh.P VS et al.,(2007)
Rule-Based Approach	30	26,149 words	87.55%	Navneet Garg et al.,(2012)
HMM	26		76.34%	Pattabhi R K Rao T et al.,(2007)
		21470 words	1]82.05%(Development set) 2]76.87%(Unannotated sets)	Asif Ekbal et al.,(2007)
			92.13%	Nisheeth Joshi et al.,(2013)
HMM +naive stemmer	18		93.12%	Manish Shrivastava et al., (2008)
HMM + error driven learning			using the TnT tagger-79.66% transformations in post processing-80.74%	Pranjal Awasthi et al., (2006)
HMM + rule based model			precision-92.56% accuracy - 87.55%	Vijeta Khicha et al.,(2017)
		27,151 words	82.05%	Asif Ekbal et al.,(2007)
HMM + TnT tagger			78.35%	G.M. Ravi Sastry et al.,(2007)
HMM using BrantsTnt		21000 words	79.64 %	Himanshu Aggarwal et al., (2006)
Shallow parsing			78.66%	G.M. Ravi Sastry et al.,(2007)
Shallow parsing+CRF			78.35%	Delip Rao et al.,(2007)
			79.64 %	Ravindran et. al.,(2006)
morphological analyzer			93.12%	Shrivastava et al.,(2008)
Artificial Neural Network(ANN)	27	11500 words	91.30%	Ravi Narayan et al.,(2014)

Table 2. Performance of POS tagging approaches for Bengali Language

APPROACH	No OF TAGS	TESTED DATA	ACCURACY	REFERENCE
HMM+Trigram tagger	26	895	78.68%	Kamal Sarkar et al.,(2012)
HMM+Bigram tagger	26	895	74.33%	Kamal Sarkar et al.,(2012)
Support Vector Machine		NLPAI-2006 contest	86.84%	A. Ekbal and S. Bandyopadhyay (2008)
	16	150K words	Recall-94.3% Precision-89.4% F-Score -91.8%	Asif Ekbal et al.,(2008)
Voted Approach method	27	57,341	92.35%	A. Ekbal et al.,(2009)
hybrid system			F-Score of 90.84%	M. M. Yoonus et al., (2011)
Global Linear Model			93.12%	S. Mukherjee et al., (2013)
ME		45,000 words	77.61%	Sandipan Dandapat(2007)
	26	72,341 words	88.2%	Asif Ekbal et al.,(2008)
HMM-S		Data 10K Data 20K Data 40K	57.53% 70.61% 77.29%	Sandipan Dandapat et al.,(2007)
HMM-S+suf		-same-	75.12 % 79.76 % 83.85%	-same-
HMM-S+MA		-same-	82.39 % 84.06 % 86.64 %	-same-
HMM-S+suf+MA		-same-	84.73 % 87.35 % 88.75 %	-same-
HMM-SS		-same-	63.40 % 70.67 % 77.16 %	-same-
HMM-SS+suf		-same-	75.08 % 79.31 % 83.76%	-same-
HMM-SS+MA		-same-	83.04 % 84.47 % 86.41 %	-same-
HMM-SS+suf+MA		-same-	84.41 % 87.16 % 87.95 %	-same-
ME+suf		-same-	77.38 % 82.63 % 86.78 %	-same-
ME+MA		-same-	82.34 % 84.97 % 87.38%	-same-
ME+suf+MA		-same-	84.13 % 87.07 % 88.4%	-same-
HMM+MA			95%	Sandipan Dandapat et al.,(2004)
HMM			84.5%	Asif Ekbal et al.,(2008)
			72.17%	Pattabhi R K Rao T et al.,(2007)
			1]90.9 %(Development set) 2]77.73 %(Unannotated sets)	Asif Ekbal et al.,(2007)
HMM + TnT tagger			74.58%	G.M. Ravi Sastry et al.,(2007)
HMM + rule based model		25,418 words	90.9%	Asif Ekbal et al.,(2007)
Shallow parsing			76.08%	G.M. Ravi Sastry et al.,(2007)
Shallow parsing+CRF			74.20%	Delip Rao et al.,(2007)
CRF			Recall-93.8 % F-score-87.8% Precision-90.7%	Asif Ekbal et al.,(2007)
CRF+TBL			76.08%	Avinesh.P VS et al.,(2007)
CRF+named entity(NE)	19	150K words	Recall-93.8%, Precision-87.8% F-Score-90.7%	Asif Ekbal et al.,(2008)
Deep Learning			93.33%	Md. Fasihul Kabir et al.,(2016)

Hybrid Approach			89.9%	Kanak Mohnot et al.,(2014)

V. CONCLUSION

In NLP, there is a progressive study towards development of POS taggers with high accuracy for better performance. In this paper, an exhaustive study is carried out on different POS taggers for Hindi and Bengali languages. This study shows which of the POS tagger have obtained better results as this will aid in determining the most prominent POS tagger during the training phase. Since there is insufficient availability of lexical resources for Indic languages, it becomes a tedious task for performing POS tagging.

REFERENCES

[1] Shambhavi.B.R, Dr.R. Kumar P., "*Current state of the art POS tagging For Indian Languages-A Study*", International Journal of Computer Engineering and Technology, Vol.1, No.1, pp.250-260, 2010.

[2] R. Kaur, L.S. Garcha, Dr.M.Garag,S. Singh, " *Parts of Speech Tagging for Indian Languages Review and Scope for Punjabi Language*", International Journal of Advanced Research in Computer Science and Software Engineering, April, Vol.7, Issue.4, pp.214-217, 2017.

[3] S. Singh, K. Gupta, M. Shrivastava, P. Bhattacharyya, "*Morphological Richness Offsets Resource Demand – Experiences in Constructing a POS Tagger for Hindi*", In the Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions, Sydney, July, pp.779–786, 2006.

[4] A. Himashu, A. Anirudh, "*Part of Speech Tagging and Chunking with Conditional Random Fields*", In the Proceedings of the NLPAI Contest, 2006.

[5] A. Dalal, K. Nagaraj, U. Sawant , S. Shelke, "*Hindi Part-of Speech Tagging and Chunking: A Maximum Entropy Approach*", In the Proceedings of the NLPAI Contest, 2006.

[6] M. Shrivastava, P. Bhattacharyya, "*Hindi POS Tagger Using Naive Stemming: Harnessing Morphological Information without Extensive Linguistic Knowledge*", In the Proceedings of ICON-2008: 6th International Conference on Natural Language Processing, Pune, India, and December, 2008.

[7] D. Kumar, G. Singh J. , " *Part of Speech Taggers for Morphologically Rich Indian Languages: A Survey*", International Journal of Computer Applications (0975 – 8887), September ,Vol. 6, No.5, pp.1-9, 2010.

[8] N. Mishra, A. Mishra, "*Part of Speech Tagging for Hindi Corpus*", In the Proceedings of 2011 International Conference on Communication Systems and Network Technologies, IEEE, pp. 554-558, 2011.

[9] S. Mall, U.C. Jaiswal, "*Hindi Part of Speech Tagging and Translation*", Int. J. Tech. 2011,Vol. 1, Issue.1, pp.29-32, 2011.

[10] N. Garg, V. Goyal, S. Preet, "*Rule Based Hindi Part of Speech Tagger*", In the Proceedings of COLING 2012: Demonstration Papers, Mumbai, December, pp.163–174, 2012.

[11] K. Sarkar, V. Gayen, "*A Practical Part-of-Speech Tagger for Bengali*", In the Proceedings of 2012 Third International Conference on Emerging Applications of Information Technology (EAIT), 2012.

[12] S. Dandapat, S. Sarkar, A. Basu, "*Automatic Part-of-Speech Tagging for Bengali: An Approach for Morphologically Rich Languages in a Poor Resource Scenario*" , In the Proceedings of ACL 2007 Demo and Poster Sessions, Prague, June, pp.221–224, 2007.

[13] A. Ekbal, S. Bandyopadhyay, *"Bengali Named Entity Recognition using Support Vector Machine*", In the Proceedings of IJCNLP-08 Workshop on NER for South and South East Asian Languages, Hyderabad, India, January, pp.51–58, 2008.

[14] H. Ali, "*An Unsupervised Parts-of-Speech Tagger for the Bangla language*", Department of Computer Science, University of British Columbia , 2010.

[15] Antony P J, Dr. Soman K P, *"Parts Of Speech Tagging for Indian Languages:A Literature Survey"*, International Journal of Computer Applications , 0975 – 8887, November ,Vol. 34, No. 8, pp-22-29,2011.

[16] K. Mohnot, N. Bansal, S. Pal Singh, A. Kumar, "*Hybrid approach for Part of Speech Tagger for Hindi language" ,* International Journal of Computer Technology and Electronics Engineering (IJCTEE), Vol. 4, Issue. 1, February ,pp.25-30, 2014.

[17] M. Kaur, M. Aggerwal, S. Kumar Sharma, *"Improving Punjabi Part of Speech Tagger by Using Reduced Tag Set"*, International Journal of Computer Applications & Information Technology, Vol. 7, Issue. II Dec 14- January, pp.142-148, 2015.

[18] A. Ekbal, et. al, "*Bengali part of speech tagging using conditional random field*" in Proceedings of the 7th International Symposium of Natural Language Processing(SNLP-2007), Pattaya, Thailand December, pp. 131-136, 2007.

[19] P. R K Rao T, V. S. Ram R, Vijayakrishna R , Sobha L *"A Text Chunker and Hybrid POS Tagger for Indian Languages",* In the Proceedings of IJCNLP-08 Workshop on NER for South and South East Asian Languages, Hyderabad, India, January, 2007.

[20] Md. F. Kabir, K. Abdullah-Al-Mamun, M. N. Huda, *"Deep Learning Based Parts of Speech Tagger for Bengali"*, In the Proceedings of 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV) ,IEEE, pp.26–29, 2016.

[21] M. M. Yoonus , S. Sinha, *"A hybrid pos tagger for indian languages."* Language in India, Vol. 11, No. 9, 2011.

[22] A. Ekbal and M. Hasanuzzaman, *"Voted approach for part of speech tagging in bengali." ,* 2009.

[23] S. Mukherjee, S. Das Mandal, *"Bengali parts-of-speech tagging using global linear model"* in India Conference (INDICON), 2013 Annual IEEE, Dec, pp. 1–4, 2013.

[24] G.M. Ravi Sastry, S. Chaudhuri, P. N. Reddy, *"An HMM based Part-Of-Speech tagger and statistical chunker for 3 Indian languages"*, In the Proceedings of IJCNLP-08 Workshop on NER for South and South East Asian Languages, Hyderabad, India, January, pp. 13-16, 2007.

[25] D. Rao, D. Yarowsky, *"Part of Speech Tagging and Shallow Parsing of Indian Languages"*, In the Proceedings of IJCNLP-08 Workshop on NER for South and South East Asian Languages, Hyderabad, India, January, pp. 17–20, 2007.

[26] P. Rutravigneshwaran, *"A Study of Intrusion Detection System using Efficient Data Mining Techniques",* Int. J. Sc. Res. in Network Security and Communication IJSRNSC, December , Vol. 5, Issue. 6, pp.5-8, 2017.

[27] Avinesh.PVS, Karthik G, *"Part-Of-Speech Tagging and Chunking using Conditional Random Fields and Transformation Based Learning",* In the Proceedings of IJCNLP-08 Workshop on NER for South and South East Asian Languages, Hyderabad, India, January, pp. 21-24, 2007.

[28] N. Joshi , H. Darbari, I. Mathur, *"Hmm based pos tagger for Hindi",* In the Proceedings of International Conference on Artificial Intelligence Soft Computing , pp. 341–349, 2013.

[29] P. Awasthi, D.Rao, B.Ravindran, *"Part of Speech Tagging and Chunking with HMM and CRF"*, In the Proceedings of NLPAI

MLcontest workshop, National Workshop on Artificial Intelligence, 2006.
[30] G. Kaur, K. Kaur, *"Sentiment Detection from Punjabi Text using Support Vector Machine",* International Journal of Scientific Research in Computer Science and Engineering, December, Vol. 5, Issue. 6, pp.39-46, 2017.
[31] A. Ekbal, S. Mandal, S. Bandyopadhyay, *"POS Tagging Using HMM and Rule-based Chunking"*, In the Proceedings of Workshop on shallow parsing in South Asian languages, pp. 25-28, 2007.
[32] V. Khicha, M. Manna, *"Part-of-Speech Tagging of Hindi Language Using Hybrid Approach",* International Journal of Engineering Technology Science and Research IJETSR, Vol. 4, Issue. 8 , pp. 737–741, 2017.
[33] R. Narayan, V. P. Singh , S. Chakraverty, *"Quantum Neural Network based Parts of Speech Tagger for Hindi"*, International Journal of Advancements in Technology, July, Vol. 5, No. 2, pp. 137-152, 2014.

Authors Profile

Miss. Floyd Avina Fernandes is pursing her masters in Computer Science and Engineering from Goa university. Her research interest is data mining.

Dr. Kavita Asnani is a faculty in Department of Goa College of Engineering and an active researcher in the domain of Sentiment Analysis, in specific to Aspect extraction. Broadly, her areas of interest include Natural Language Engineering, Text Analytics, Machine Learning. Presently, she has been actively working on various forms of machine's creativity including linguistics creativity on code-mixing in social media text and Computational Social Sciences. She has contributions by research publications in reputed conferences and journals. She has experience in mentoring projects, conducting several workshops and conferences in the area of NLP and funded research projects like CIDA.

