Current Trends in Technology and Sciences
Volume : 1, Issue : 2 , (Sept.-2012) ISSN : 2279-0535

Computational Analysis of Part of Speech Tagging

Name: Richa Garg
Designation: Student, M. Tech, Organization: BIST, Bhopal, India, Email ID: richaagar20@gmail.com

Name: Akhilesh A. Waoo
Designation: Assistant Professor, Organization: BIST, Bhopal, India, Email ID: akhilesh_waoo@rediffmail.com

Name:P.S. Patheja
Designation: HOD, Organization: BIST, Bhopal, India, Email ID: pspatheja@gmail.com

Abstract —In order, to make text a suitable input to an automatic method of information extraction it is usually transformed from unstructured source of information into a structured format. Part of Speech Tagging is one of the preprocessing steps which assign one of the parts of speech to the given word. In this paper we had discussed various models of supervised and unsupervised technique shown the comparison of various techniques based on accuracy, and experimentally compared the results obtained in models of Condition Random Field and Maximum Entropy model. We had deployed a model of part of speech tagger for which we had compared the results with other models. The developed is based on HMM approach and had shown good results in terms of efficiency in comparison with other models.

General Terms — Accuracy. Part of Speech Tagging, Supervised Technique, Unsupervised Technique.

Keyword — CRF, MaxEnt, NLP, POS.

1. INTRODUCTION

There is a wide range and focus areas in Human Language Technology (HLT). These include areas such as Natural Language Processing (NLP), Speech Recognition, Machine Translation, Text Generation and Text Mining.

A natural language understanding system must have knowledge about what the words mean, how words combine to form sentences, how word meanings combine to from sentence meanings and so on [49].

Text documents are greatest source of information from which user extract information depending upon his interest [18]. So in order to extract meaning and relevant information in text document focus lies towards passage or sentence.

Retrieving relevant passage as compared to whole document helps in filtering out irrelevant information that improves accuracy [3]. It runs into many stages, namely tokenization, lexical analysis, syntactic analysis, semantic analysis, pragmatic analysis and discourse analysis.

As text is an unstructured source of information, to make it a suitable input to an automatic method of information extraction it is usually transformed into a structured format. This preprocessing involves multiple steps namely sentence segmentation, tokenization, part of speech tagging, entity detection, relation detection [21]. We in this paper are focusing on one of the preprocessing step i.e. part of speech tagging.

Parts of Speech Tagging is an approach to perform Semantic Analysis and include the process of assigning one of the parts of speech to the given word. Parts of speech include nouns, verbs, adverbs, adjectives, pronouns, conjunction and their sub-categories.

Part of Speech Tagging has been broadly divided upon Supervised and Unsupervised Techniques having further classification of each type. In the remainder of this paper detailed classification of both Supervised and Unsupervised Techniques are described further stating the best techniques resulted based on accuracy achieved so far.

We had then shown experimental results obtained for two best of art Part of Speech Tagging techniques based on their execution time.

The results of a model based on Part of Speech tagger has been demonstrated which has been developed taking WordNet as a lexicon. Finally we had discussed the issues occurring in supervised system of tagging.

2. CLASSIFICATION OF PART OF SPEECH TAGGING

Tagging in natural language processing (NLP) refers to any process that assigns certain labels to linguistic units. It denotes the assignment of part-of-speech tags to texts. A computer program for this purpose is called a tagger. Part of speech tagging includes the process of assigning one of the parts of speech to the given word.

For example, the english word rust for instance is either a verb or a noun. Part of speech tagging can be categorized as follows:

2.1 Supervised Tagging and Unsupervised Tagging

Supervised Technique use a pre-tagged corpora (structured collection of text) which is used for training to learn information about the tagset, word-tag frequencies, rule sets etc. As compare to Supervised

Classification Unsupervised Part of Speech (POS)

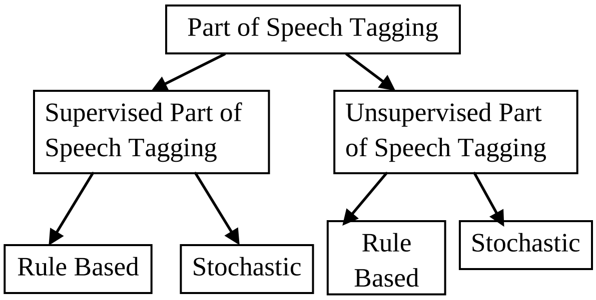

Fig. 1. Broad Classification of Part of Speech Tagging Techniques

Operates by assuming as input POS lexicon, which consists of a list of possible POS tags for each word. Both supervised and unsupervised tagging can be of two types, Rule based and stochastic [29]. The successful work to date has used supervised learning techniques. Unsupervised algorithms that can learn from raw linguistic data, as humans can, remain a challenge [30].

2.2 Rule Based Technique and Stochastic Technique

Stochastic tagging is the phenomena, which incorporates frequency or probability, i.e. statistics. Rule based techniques use contextual and morphological information to assign tags to unknown or ambiguous words. These rules are often known as context frame rules for example: If an ambiguous/unknown word X is preceded by a determiner and followed by a noun, tag it as an adjective [36].
These rules can be either automatically induced by the tagger or encoded by the designer. Eric Brill designed the best-known rule-base part of speech tagger, which was the first one to be able to achieve an accuracy level comparable to that of stochastic taggers i.e. 95%-97% [28].

3. Classification of Supervised and Unsupervised Tagging Techniques

Supervised and Unsupervised tagging techniques can be classified into following categories.

3.1 Decision Tree Model

A decision tree is a predictive model with a tree structure that recursively partitions the training data set. Each internal node of a decision tree represents a test on a feature value, and each branch represents an outcome of the test.
A prediction is made when a terminal node (i.e., a leaf) is reached. Tree tagger is able to achieve the accuracy of 96.36% on Penn Treebank better than of trigram tagger (96.06%) [12].

3.2 Condition Rand

3.3 om Field Model

J.Lafferty explores the use of Condition Random Field (CRF) model for building probabilistic models and labeling sequence data. They are a probabilistic framework for labeling and segmenting structured data, such as sequences, trees and lattices. Conditional random fields (CRFs) for sequence labeling offer advantages over both generative models like Hidden Markov Model (HMM) and classifiers applied at each sequence position. CRFs do not force to adhere to the independence assumption and thus can depend on arbitrary, non-independent features, without accounting for the distribution of those dependencies [17].CRF achieves accuracy of 98.05% in close test and 95.79% in open test [15].

3.4 Hidden Markov Model

In Hidden Markov Models, (HMM) state transitions are not observable. HMM taggers require only a lexicon and untagged text for training a tagger. Hidden Markov Models aim to make a language model automatically with little effort. Disambiguation is done by assigning more probable tag. The classic method of training HMMs for part of-speech induction is the Baum-Welch [22]. The most common stochastic tagging technique uses a Hidden Markov Model (HMM) [32]. Hidden Markov Models (HMMs), an important special case of DBNs, are a classical method for speech recognition [26]. A Hidden Markov Model (HMM) consists of the states which correspond to the tags, it has an alphabet which consists of the set of words, the transition probabilities P ($Tag_i|Tag_i$-1) and the emission probabilities P($Word_i|Tag_i$).In HMM, for a given (word, tag) pair we have the probability formulae[19]:

$$P(w, t) = \Pi\, P(Tagi|Tagi - 1) * P(Wordi|Tagi).$$

Different work carried out under HMM are that of Merialdo, 1994, Elworthy, 1994, Banko and Moore 2004, Wang and Schuurmans 2005 [2]. Maximum accuracy obtained is 95%-97% [5].

3.5 Maximum Entropy Model

Maximum Entropy Tagging thrives to find a model with maximum entropy. Maximum entropy is the maximum randomness. The outputs of the maximum entropy tagging are tags and their probabilities. In contrast to HMMs, in which the current observation only depends on the current state, the current observation in an MEM may also depend on the previous state. The term, maximum entropy here means maximum randomness or minimum additional structure. The MaxEnt model is trained from labelled data and has access to any predefined attributes of the entire word sequence and to the labels of previous words [25]. Best accuracy reported in maximum entropy model is by Stanford tagger of 96.9% [24].

Current Trends in Technology and Sciences
Volume : 1, Issue : 2 , (Sept.-2012) ISSN : 2279-0535

3.6 Clustering Model

This model focuses distributional properties and co-occurrence patterns of text (similar words occur in similar contexts) by computing context vectors for each word to cluster words together in groups, groups which can then be assigned Part of Speech tags or word classes as groups. The key characteristics are how the context vectors are defined, size of the context vectors (number of dimensions), metric used to compute vector similarity (i.e. make clusters), and how the tags or word classes are induced on the clusters. Schutze, 1995 and Clark, 2000 had shown results in this Category of clustering model.
Best accuracy is reported as 59.1% [31].

3.7 Prototyping Model

Prototypes posses better evaluation (since small size) and more meaning than clusters. In this model a few examples or prototypes are collected (one for each target tag) and then propagated across the corpus of unlabeled data. No lexicon is required in this model. A gradient based search with the forward-backward algorithm. This model is used to maximize the log linear model parameters. Accuracy achieved in this model is 80.5% [41].

3.8 Bayesian Model

Bayesian learning models for Part of Speech tagging integrates over all possible parameter values as compare to finding a parameter set which maximizes the probability of tag sequences given unlabeled observed data. Work done in Bayesian Model is shown by Toutanova and Johnson, 2007, Goldwater and Griffiths, 2007, Johnson, 2007 Accuracy achieved is 93.4% [23].

3.9 Neural Networks

A neural network (NN) is an interconnected group of natural or artificial neurons that uses a computational model for processing data pairs of input feature and desired response where data pairs are input to the learning program. Input features partition the training contexts into a number of overlapping sets corresponding to the desired responses. Best accuracy achieved in neural network is 96.9% [21].

Table (1). Comparative Performance Results

Models	Accuracy
Decision Tree [37]	96.36%
Max Entropy [24]	96.97%
HMM [7]	“95-97”%
CRF [2]	“95.79-98.05”%
Clustering [31]	59.10%
Prototyping [41]	80.50%
Bayesian [23]	93.40%
Neural Network [21]	96.90%
Rule Based [28]	“95-97”%

4. Data Sources Referred During Part of Speech Tagging

Knowledge is a fundamental component of part of speech tagging. Knowledge sources provide data which are essential to associate senses with words. Knowledge sources can be divided into following types.

4.1 Machine Readable Dictionaries (MRD)

MRD are dictionaries in electronic format which are most utilized resource for word sense disambiguation in English.WordNet encodes a rich semantic network of concepts and defined as a computational lexicon [21].

4.2 Tagset

Apart from corpora, a well chosen tagset is also important. So, for deciding upon a tagset, following properties should be considered i.e. Fineness Vs Coarseness, Syntactic function Vs Lexical category, New Tags Vs Tags from a Standard Tagger.
Tagset can be divided into coarse gain and fine grain tagset. Most work has focused on POS-tagging for English using the Penn Treebank. The Penn Treebank tagset contains 36 POS tags and 12 other tags [27]. This generally involves working with the standard set of 45 POS-tags employed in the Penn Treebank [10].

4.3 WordNet

WordNet build a general lexical database without defining a certain domain but covering all possible topics. So it can be widely used for Natural Language Processing (NLP) and Word Sense Disambiguation (WSD) tasks in any possible context. It is a lexical-conceptual model and database consisting of both lexical units and the relations between such units [3].

5. Comparison of Different Models of POS Tagging Technique

Comparative results have been shown in table 1 for different models of Part of Speech Tagging Technique based on data obtained from different reference papers and sources and correspondingly best accuracy results had been demonstrated by two supervised tagging technique i.e. CRF and Maximum Entropy model.

6. Experimental Results

6.1 Condition Random Field Model and Maximum Entropy Model

A maximum entropy based tagger has been proposed in Ratnaparkhi, 1996.The tagger learns a log linear conditional probability model from tagged text, using maximum entropy method [39].We had used Stanford part of speech tagger, which is an extension of the paper Ratnaparkhi, further incorporating log linear concept in maximum entropy model. This tagger uses Penn Treebank tagset, comprising a set of around 48 tags for tagging[40].

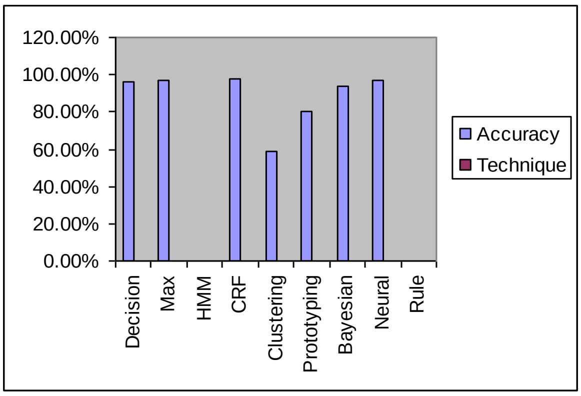

Fig. 2 Graphical representation of different tagging Techniques based on accuracy.

The tagger has three modes tagging, training, and testing. Tagging allow us using a pretrained model to assign part of speech tags to unlabeled text.

Training allows us saving a new model based on a set of tagged data. Testing allows us to see how well a tagger performs by tagging labeled data and evaluating the results against the correct tags. We had experimented the results with varying number of tokens and the correspondingly execution rate achieved.

As stated in [24] the best accuracy reported in Maximum Entropy model ranges from 96.97%-97%.Correspondingly the performance of tagger in terms of efficiency is demonstrated in Table 2.

CRF tagger had been helping in dealing with the label bias problem present in Maximum Entropy Markov Model [16].

CRF model address the problem by using a single exponential model for entire label sequence given a observation sequence. Table 2 shows the performance of CRF model based tagger in terms of the efficiency achieved for same dataset as used for Maximum Entropy model. When demonstrating tagging results with Condition Random field model we had used Penn Treebank tagset. CRF tagger is unable to demonstrate accurate results for small number of tokens.

6.2. Proposed Work: New Tagger Developed Deploying WordNet as Lexicon

Since one of the goals of tagging is to have a fast implementation for tagging large amounts of data quickly with significantly faster at runtime[14], so system proposed to deploy a HMM supervised stochastic approach while implementing POS tagger. Define tags using coarse tagset. The major steps/algorithms used during the development of a POS tagger shall be as follows:

1. Segmentation of text into word and sentence units through the technique of tokenization.
2. Initial (non-contextual) part-of-speech assignment by adopting following mathematical formulae.

   $$P\ (w \mid t) = \sum P\ (wi \mid ti) + \sum\ (\ ti \mid ti\text{-}1\ ) \qquad (1)$$

3. Deploying WordNet lexical database for tag inference.
4. Using Penn Treebank tagset for comparison aspect.
5. Deducing accuracy as degree of confidence.
6. Representing output in terms of degree of confidence, ambiguity and elapsed execution time of proposed tagger.

So, we had developed a model of part of speech tagger using WordNet as a computational lexicon the tagger derives a probability formulae where w = Word, t = Tag

$$\mathbf{P\ (w \mid t) = \sum P\ (wi \mid ti) + \sum\ (ti|ti\text{-}1)} \qquad \mathbf{(1)}$$

There are two aspects of the efficiency i.e. amount of time required to execute the algorithm and the memory space it consumes. The time complexity of Stanford tagger is $O(TN^2)$ while that of developed tagger is O(NT+T) where T is the length of state sequence and N is the total number of tags or states. The space complexity of Stanford tagger is $O(T^2)$ while that of developed tagger is O(T) where T is the length of state sequence. The results of proposed tagger based time and space complexity is represented in figure 4, figure 5, and figure 6. Ambiguity is the issue in which a word tag relationship is ambiguous i.e. multiple tags are associated with same word with same probability of occurrence. The developed tagger is unable to demonstrate good average accuracy due to the ambiguity encompass within correspondingly its run time performance evaluation based on ambiguity, accuracy and execution time is demonstrated in figure 7, 8, 9 and 10 .

Table (2). Results obtained for Stanford Tagger and CRF Tagger

S.No	Number of Tokens	Execution Rate (Stanford, per s)	Stanford Execution Time (in s)	CRF Execution Time (in s)
1	04	85.11	0.0469	0.0
2	08	170.21	0.0470	0.0
3	12	255.32	0.0479	0.0
4	16	361.70	0.0442	0.0
5	32	680.85	0.0470	0.0
6	513	2052.00	0.25	0.031
7	989	2108.74	0.469	0.063
8	3896	2770.98	1.406	0.203
9	15584	5167.11	3.015	0.781
10	31168	5334.25	5.842	1.1531

Current Trends in Technology and Sciences
Volume : 1, Issue : 2 , (Sept.-2012) ISSN : 2279-0535

Table (3). Comparison of Results for Stanford Tagger and Developed Tagger

S.No	Tokens	Stanford Execution Rate (per s)	Stanford Execution Time (in s)	Developed Tagger Execution Time (in s)
1	04	85.11	0.0469	0.022
2	08	170.21	0.0470	0.031
3	12	255.32	0.0479	0.040
4	16	361.70	0.0442	0.049
5	32	680.85	0.0470	0.082

Table (4). Results obtained For Developed Tagger for Ambiguity, Memory Used and Accuracy.

S.No	Number of Tokens	Ambiguity (%)	Memory Used (KB)	Accuracy (%)
1	04	1.13	221	98.86
2	08	5.47	220	94.52
3	12	6.02	265	93.97
4	16	7.10	312	92.89
5	32	6.58	768	93.41

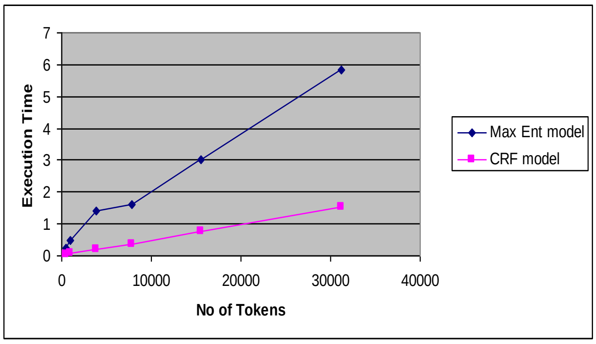

Fig. 3. Comparison of CRF and Stanford Tagger

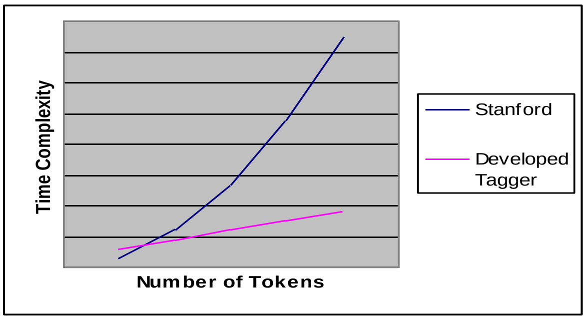

Fig. 4. Comparison of Time Complexity of Developed Tagger and Stanford Tagger for fixed values of T.

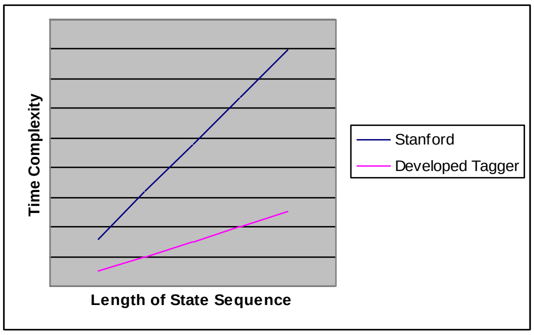

Fig. 5. Comparison of Time Complexity of Developed Tagger and Stanford Tagger for fixed values of N.

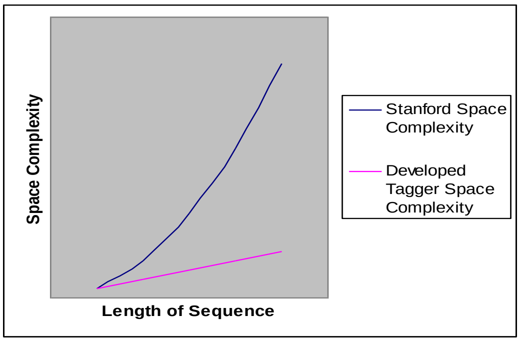

Fig. 6. Comparison of Space Complexity of Developed Tagger and Stanford Tagger

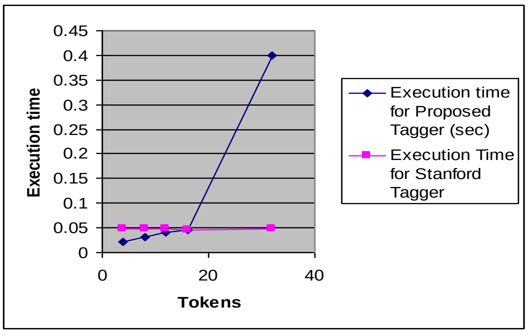

Fig. 7. Graphical Comparisons for Stanford Tagger and New Tagger for Execution Time

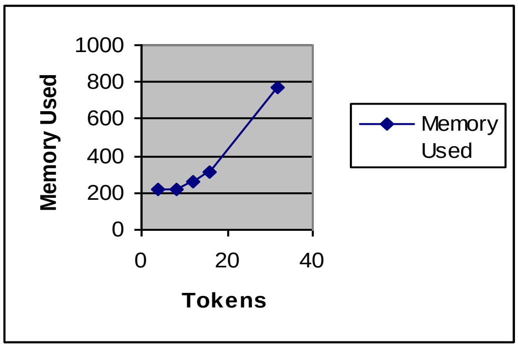

Fig. 8. Graph shows used memory results for Proposed tagger

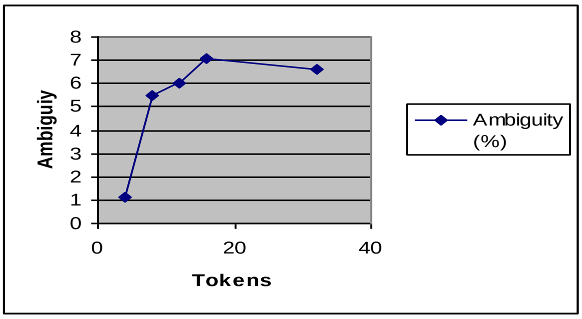

Fig. 9. Graph shows ambiguity results for Proposed tagger

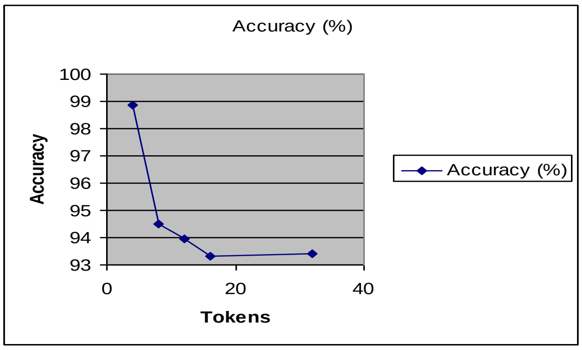

Fig. 10. Graph shows accuracy results for proposed tagger

7. Discussion

Condition Random Field (CRF) based model attains good performance results as compared to Maximum Entropy Model as shown experimentally in Figure.3. Also the model of tagger using WordNet as lexicon has demonstrated sufficiently good performance when compared with Maximum Entropy model as shown in Figure. 7.

Though supervised technique had shown good performance results in terms of accuracy, yet it suffers from the problem of data sparcity. Data Sparcity is the issue in which words appearing in test set are unavailable in test set due to large size of dictionaries. Research is going on to solve this issue of data sparcity with the help of CRF model.

8. Limitation

Every research is inculcated with some limitation which provides a scope of future work. For example E-rater software used by GMAT cannot detect humor, spelling error or grammar. Similarly the proposed tagger have demonstrated good results for very small number of tokens but due to some memory issues during coding fail to demonstrate better results for large number of tokens. Also for some tokens the performance results demonstrate by the tagger increases due to the issue of ambiguity. A limitation also lies with the CRF tagger i.e. this tagger fails to give results for very small number of tokens due to which we have shown comparative results with Maximum Entropy based tagger only.

9. Conclusion

After comparing the experimental results of both CRF and MEM based models it is found that for a dataset of around 31k tokens the average execution time obtained for Maximum Entropy model is 2.0675 sec and for Condition Random Field model is 0.49733 sec respectively. Thus it is proved that CRF model achieve better performance results both in terms of accuracy and execution time than Maximum Entropy model as shown in Figure.3.

There are two aspects of the efficiency i.e. amount of time required to execute the algorithm and the memory space it consumes. The time complexity of Stanford tagger is $O(TN^2)$ while that of developed tagger is O(NT+T) where T is the length of state sequence and N is the total number of tags or states.

As shown in Figure 4 and Figure 5 that time complexity of both the developed tagger and of Stanford tagger increases with number of tokens and length of sequence, however in both the figures the time complexity of developed tagger hold lesser and better result than Stanford tagger.The space complexity of Stanford tagger is $O(T^2)$ while that of developed tagger is O(T) where T is the length of state sequence and as shown in figure 6 that space complexity i.e. memory requirement of both the developed tagger and of Stanford tagger increases with length of sequence, However space complexity of developed tagger hold lesser and better result than Stanford tagger.

The model of implemented tagger using WordNet as lexicon has demonstrated better better efficiency i.e. for small number of tokens than Maximum Entropy based tagger i.e. for a dataset of around 32 tokens the average execution time obtained for Maximum Entropy model is .0496 sec and and for proposed Tagger is 0.0355 sec respectively. Hence demonstrated sufficiently good efficiency when compared with Maximum Entropy model as shown in Figure.7.

The accuracy of tagger decreases with increase in number of tokens as shown in figure 10. The developed model has taken well care for efficiency issue but due to ambiguity issue, it had shown its limitation in obtaining best average accuracy.

The ambiguity encompassed with the tagger increases with increase in number of tokens from 4 to 20 and then decreases slightly for 32 tokens as shown in figure 9.The memory usage increases with increase in number of tokens as shown in figure 8. The graph show less variation for small number of tokens but for number of tokens greater or equal to 32, it shows drastic memory usage. As system has proved better space complexity theoretically for developed tagger, hence the reason can be coding issues. Since the aim of system while developing this tagger is to obtain an improvement in the performance parameters of tagger in terms of memory space, time, accuracy and ambiguity. So the system is able to obtain better results in the field of time. But due to the copyright issue of Stanford tagger system had failed to demonstrate comparison of run time performance issues of accuracy, ambiguity and memory space results obtained for new tagger with that of Stanford Tagger.

REFERENCES

[1] M.Suneetha, S. Sameen Fatima, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-3, July 2011.
[2] Myint, C, A Hybrid Approach for Part of Speech Tagging, IEEE 978-1-42244-92982-427,May 2011,
[3] Genoveva Galarza Heredero, Subhadip Bandyopadhyay, Arijit Laha, ACM 978-1-4503-0750-5/11/03, Copyright 2011.
[4] Mehmet Ali Yatbaz, Deniz Yuret, Coling 2010: Poster Volume, pages 1391–1398, Beijing, August 2010.
[5] Lichi Yuan,Sch. of Inf. Technol, IEEE ,Jiangxi Univ. of Finance & Econ, Nanchang, China , 978-1-4244-6892-8, 2010 .
[6] Michael Connor, Yael Gertner, Cynthia Fisher, Dan Roth, Association for Computational Linguistic, July 2010.
[7] Taesun Moon, Katrin Erk, and Jason Baldridge, Conference on Empirical Methods in Natural Language Processing, Proceedings of the 2010.
[8] Shane Bergsma, Emily Pitler, Dekang LinAssociation for Computational Linguistics, 2010.
[9] Jonathan K. Kummerfeld, Jessika Roesner,Tim Dawborn, James Haggerty, James R. Curran, Stephen Clark, Association for Computational Linguistics, 2010.
[10] Sujith Ravi, Jason Baldridge, Kevin Knight, Association for Computational Linguistics, 2010.
[11] Michael Lamar, Yariv Maron, Mark Johnson, Elie Bienenstock, Association for Computational Linguistics, July 2010.
[12] XIAOJUN WAN and JIANGUO XIAO, ACM Transactions on Information Systems, Publication date: May 2010.
[13] Ryan T. K. Lin, Hong-Jie Dai, Yue-Yang Bow,Justing Lian-Te Chiu ,Richardg Tzon-Han Tsa, Published in: Journal Integrated Computer-Aided Engineering, December 2009.
[14] Drahom ıra johanka Spoustova ,Jan Hajic ,Jan Raab ,Miroslav Spousta, Association for Computational Linguistic,2009.
[15] Xiaofei Zhang,Heyan Huang,Zhang Liang ,The application of CRF in Part of Speech Tagging , IEEE, 9778-0-7695-3752-8, Nov 2009.
[16] Gerard de Melo, Gerhard Weikum, Hong Kong, China, November 2–6, Copyright 2009 ACM.
[17] Cosmin Munteanu, Gerald Penn, Xiaodan Zhum, ACL and AFNLP, August 2009. .
[18] Zhongqiang Huang, Vladimir Eidelman, Mary Harper, Association for Computational Linguistics Proceedings of NAACL HLT, June 2009.
[19] Fei Huang, Alexander Yates, ACL and AFNLP, 2009.
[20] Sujith Ravi and Kevin Knight,ACL and AFNLP, 2009.
[21] Roberto Navigli, ACM Computing Surveys, Vol. 41, No. 2, Article 10, Publication date: February 2009.
[22] William P. Headden III, David McClosky, Eugene Charniak, Proceedings of the 22nd International Conference on Computational Linguistics, August 2008.
[23] Ankit k Srivastava,Unsupervised Approaches to Part of Speech Tagging, March 19,2008.
[24] Jinghui Xiao, Xiaolong Wang, Bingquan Liu, ACM Trans. Asian Language Information Processing, Pub. Date: September 2007.
[25] Eric Ringger, Peter McClanahan, Robbie Haertel, George Busby, Marc Carmen,James Carroll, Kevin Seppi, Deryle Lonsdale, Association for Computational Linguistics, June 2007.
[26] Charles Sutton, Andrew McCallum, Khashayar Rohanimanesh , Journal of Machine Learning Research Published 3/07, 8 (2007) 693-723.
[27] Elliott Franco Drabek, David Yarowsky Association for Computational, June 2005.
[28] Garnett Wilson and Malcolm Heywood, Washington, DC, USA, Copyright 2005 ACM 1-59593-010-8/05/0006.
[29] http://language.worldofcomputing.net.
[30] Sharon Goldwater and Mark Johnson, 2005 ACL.
[31] Valentin I. Spitkovsky, Hiyan Alshawi, Angel X. Chang, Daniel Jurafsky, Unsupervised Dependency Parsing without Gold Part-of-Speech Tags.
[32] Qing Ma, Bao-Liang Lusahara, 2002 IEEE ,0-7803-5731-0 .
[33] Nuno C. Marques, Gabriel Pereira Lopes, Neural Networks, Part-of-Speech Tagging and Lexicons, Technical Report DI-FCT/UNL n. 6/98.
[34] Steven Abney, 1996, Part-of-Speech Tagging and Partial Parsing, 1996.
[35] Adwait Ratnaparkhi, Maximum Entropy Model for Part-Of-Speech Tagging,IEEE 1996.
[36] Eric Brill Department of Computer Science Johns Hopkins University, Unsupervised Learning of Disambiguation Rules for Part of Speech Tagging, 1996.
[37] Helmut Schmid, Probabilistic Part of speech Tagging Using Decision Trees, Manchester, UK, 1994.

[38] Eric Brill, A Simple Rule Based Part of Speech Tagger, ACLTrento Italy, 1992, Proceedings of the Third Conference on Applied Computational Linguistics .
[39] Kristina Toutanova, Christopher D. Manning, Stanford, CA 94305–9040, USA.
[40] http://www.800score.com/awac4.html.
[41] Aria Haghighi, Dan Klein, Prototype-driven learning for sequence models, Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, 06 .
[42] http://nlp.stanford.edu.