

Assignment-1

REPORT

Using 36 Tags in the Penn Treebank dataset:

```
Accuracy of NN :75.10899722552516
Accuracy of IN :97.55970924195223
Accuracy of DT :98.73976055450535
Accuracy of NNS :70.5685618729097
Accuracy of MD :95.76719576719577
Accuracy of VB :91.4
Accuracy of JJ :74.11255411255411
Accuracy of NNP :79.83706720977597
Accuracy of CC :97.49430523917995
Accuracy of VBD :88.09523809523809
Accuracy of PRP :94.05099150141642
Accuracy of TO :96.60633484162896
Accuracy of CD :75.56497175141243
Accuracy of VBG :61.03448275862069
Accuracy of RB :75.29411764705883
Accuracy of JJR :78.37837837837837
Accuracy of VBP :79.6875
Accuracy of WDT :88.88888888888889
Accuracy of VBZ :88.19277108433735
Accuracy of EX :93.75
Accuracy of JJS :79.3103448275862
Accuracy of -RRB- :86.95652173913044
Accuracy of VBN :76.06635071090048
Accuracy of PRP$ :92.64705882352942
Accuracy of WP :96.07843137254902
Accuracy of NNPS :51.28205128205128
Accuracy of RBR :75.0
Accuracy of : :86.66666666666667
Accuracy of WRB :94.5945945945946
Accuracy of RP :66.66666666666666
Accuracy of -LRB- :95.0
Accuracy of # :100.0
Accuracy of PDT :60.0
Accuracy of RBS :50.0
Accuracy of LS :25.0
Accuracy of WP$ :100.0
Total Accuracy :83.68206017085552
```

Using 4 Tags in the Penn Treebank dataset:

```
Accuracy of N :85.31981824536875  
Accuracy of O :95.72649572649573  
Accuracy of V :83.32670115399921  
Accuracy of A :75.53133514986376  
Total Accuracy :87.81725888324873
```

The accuracy of the 4-tag model is better than the accuracy of the 36-tag model because as the number of tags increases the model needs to be more precise in terms of making distinctions between all the tags.

Also, in the case of 36 tags the model has fewer training cases for each tag as the overall data gets divided into 36 sections according to tag, but in the 4-tag model we have more data per tag in training.

In case of the 36-tag model more ambiguities and errors are possible because of the larger available choices at each point of time, also more amount of data would be needed to correctly capture the transition and emission probabilities.

Hence, the accuracy of the 4-tag system is better as compared to 36-tag system.