Shivank Bajpai (A20364833)

Results of the Assignment 01 for Natural Language Processing :

UNIGRAM TAGGER :

Overall Accuracy :0.8548755299489075
Average accuracy per sentence :0.853959495149
Precision for UniGram Tagger is
0.814798467335
Recall for UniGram Tagger is
0.919550733928
The Tags with highest accuracy for Unigram Tagger are :
['PRP$', '``', '"""', 'WP', '#', '$', 'FW', ',', '.', 'TO', 'PRP', '-LRB-', ':', 'LS', 'WP$', '-RRB-', 'SYM', 'UH']

BIGRAM TAGGER:

Overall Accuracy :0.1334927709533645
Average accuracy per sentence :0.162655970576
Precision for BiGram Tagger is
3.07503075031e-05
Recall for BiGram Tagger is
0.5
The Tags with highest accuracy for Bigram Tagger are :
['#', 'FW', 'LS', 'SYM', 'UH']

UNIBITRI BACKOFF TAGGER:

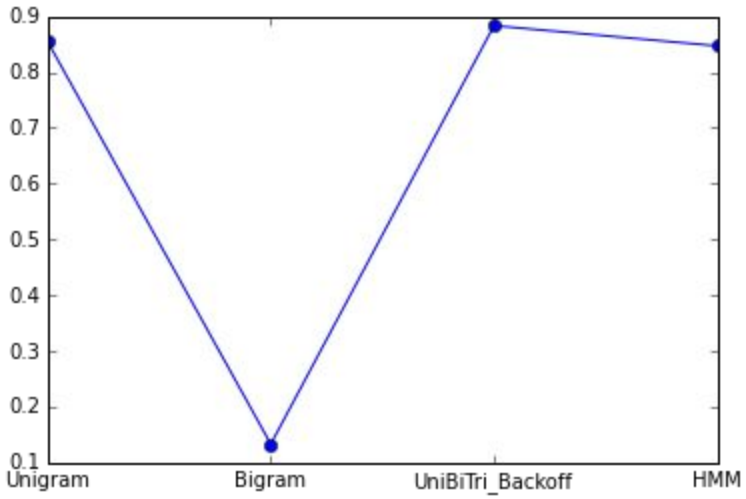
Overall Accuracy :0.8833568866181106
Average accuracy per sentence :0.883890037152
Precision for UniBiTriBackOff Tagger is
0.856859727351
Recall for UniBiTriBackoff Tagger is
0.888589259721
The Tags with highest accuracy for Unigram Tagger are :
['PRP$', '``', 'WP', '#', '$', 'FW', ',', '.', 'TO', 'PRP', '-LRB-', ':', 'LS', 'WP$', '-RRB-', 'SYM', 'UH']

HMM TAGGER:

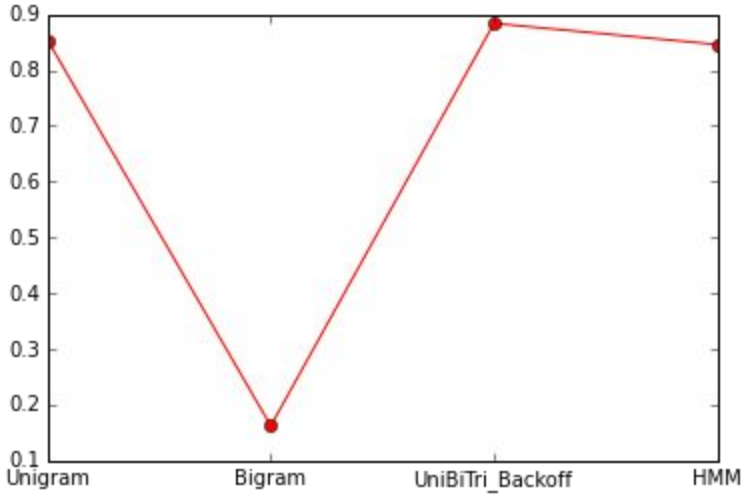
Overall Accuracy :0.8470485922382868
Average accuracy per sentence :0.846032939731

Precision for HMM Tagger is
0.681337389338
Recall for HMM Tagger is
0.839371717452
The Tags with highest accuracy for HMM Tagger are :
['PRP$', '#', 'FW', 'LS', 'SYM', 'UH']

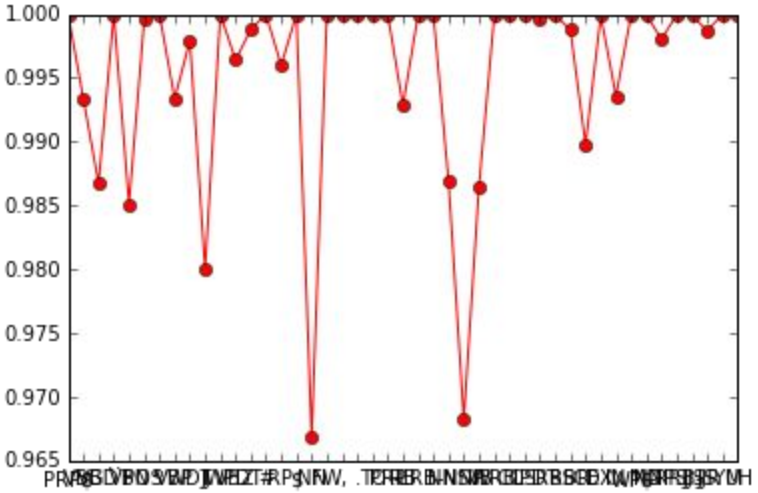
Graph to Compare the overall accuracies for the Unigram , Bigram , UniBiTri backoff and HMM tagger:



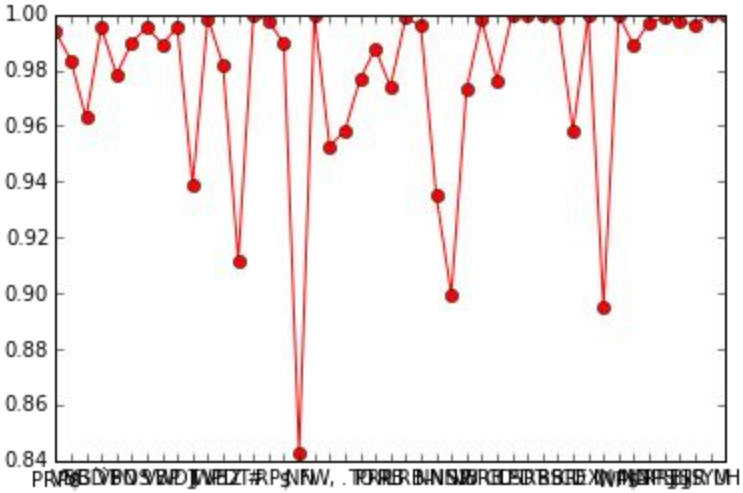
Graph to compare the average accuracies per sentence using Unigram , Bigram , UniBiTri and HMM:



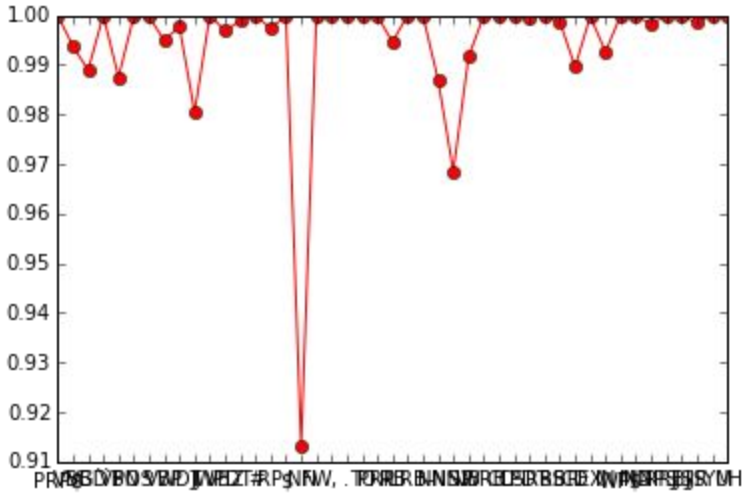
The Tags with highest accuracy for Unigram Tagger are :
['PRP$', '``', "''", 'WP', '#', '$', 'FW', ',', '.', 'TO', 'PRP', '-LRB-', ':', 'LS', 'WP$', '-RRB-', 'SYM', 'UH']



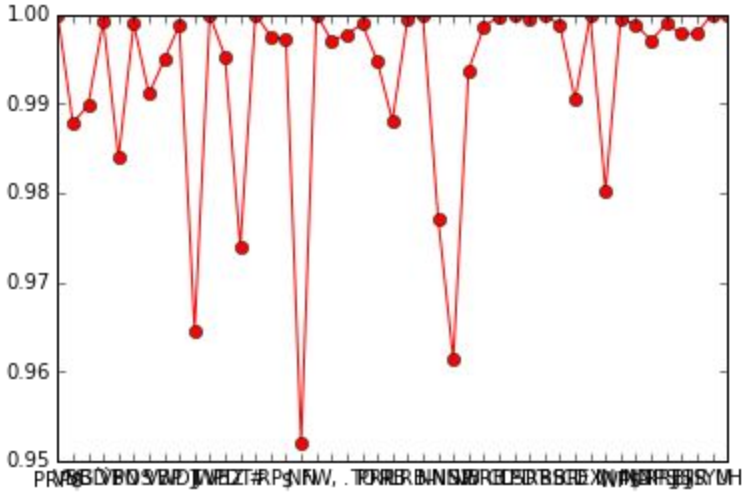
The Tags with highest accuracy for Bigram Tagger are :
['#', 'FW', 'LS', 'SYM', 'UH']



The Tags with highest accuracy for UniBiTriBackoff Tagger are :
['PRP$', '``', 'WP', '#', '$', 'FW', ',', '.', 'TO', 'PRP', '-LRB-', ':', 'LS', 'WP$', '-RRB-', 'SYM', 'UH']



The Tags with highest accuracy for HMM Tagger are :
['PRP$', '#', 'FW', 'LS', 'SYM', 'UH']



Results Analysis :

- The highest accuracy is obtained from the UniBiTri Backoff Tagger as the Tri tagger if not able to tag the words , then the subsequent taggers will try to tag the words which increases the probability of the word being tagged correctly .
- This also results in highest average accuracy per sentence as the UniBiTri Backoff tagger will tag the sentences with highest probability and will hence result , in more words of a sentence being tagged correctly.
- As the precision is defined by the total number of words which have been tagged correctly in the test set and hence , the UniBiTri tagger as explained above is able to identify the highest number of tags correctly in test set and hence resulting with highest precision.

- As recall is defined by the total number of positive cases which were identified by the Tagger , the UniGram tagger was able to assign the tags to the maximum number of words as the Unigram Tagger identifies most of the word in the corpus.
- The most effective tagger identified is the UniBiTri Tagger which identifies the words with the correct tag and has the highest overall accuracy , average accuracy per sentence and highest precision . The tagger has interdependencies between Tri , Bi and Uni gram Tagger and uses Backoff , to tag the words correctly , if not able to tag using the previou tagger.
- The tags occurring the most using every tagger as have been displayed in the above diagrams , this happens as the the tag are the most occurring tags and used extensively in the sentences present in the corpus .