Serial Assignment - N1

The purpose of this assignment is to build a HMM based/TBL based POS tagger. The tools and their documentation will be provided to you. The data to be used has been split into two parts - training and testing. Use the training data to train the POS tagger and test it on the testing data.

The steps to be followed are -

1. Convert the data into the format used by the tagger (Will be there in the documentation).
2. Train the POS tagger on the training data.
3. Test the tagger using the test data.
4. Report results and do an error analysis. The error analysis should include the number of tags which were marked incorrectly, overall precision and recall figures, tag-wise precision and recall figures.
5. (Optional) See how the precision and recall figures vary as you change the size of the training corpus. Make a table in the following format

Training size	#Correct Tags	#Errors	Precision
10,000	-	-	-
20,000	-	-	-
30,000	-	-	-
40,000	-	-	-
…	-	-	-

6. (Optional) Suggest steps to improve the precision of your tagger. Implement these steps and see their effect on the precision figures.

You have to do tasks 1 to 4 as well as one of the optional tasks.