

Tagged test file – result.txt

Postag.py – tagger, has a baseline function as well which is not called in the final version. (If need be, replace the `hmm(train,test)` with `baseline(train,test)`)

Acc.py – compares the original tagged file with the file tagged by postag

scripty.py – takes a tagged file and writes just the words without tags to a file – to test accuracy.

Accuracy:

While I was not able to determine how accurate the program was for the test file since there was not a complete file to test the results against, I tested the accuracy with 20% of the training file. While the accuracy for the baseline method which assigned the words the highest frequency tag if they were in the training data and NP if they were not (the simplest method possible) was 78%, the accuracy for the HMM method was 98%. What is still baffling for me is the fact that it really does not seem to like barbecue. I'm really not sure what happened.

The rest of the information is in the comments.