

CS 6320 Natural Language Processing

Homework 3 POS Tagging

Report

```
E:\Books\NLP\Assignment\assignment3>evaluation.py -ref reference.txt -sys output
Test Words: 63346
Tagged Words: 63346
Unknown Words: 1238
Known Words: 62108
Unknown Good Words: 231
Known Good Words: 59921
Good Count:60152
Known Good Count:59921
Unknown Good Count:231
Accuracy, Overall: 0.949578505352
Accuracy for Known words: 0.964787144973
Accuracy for Unknown words: 0.186591276252

E:\Books\NLP\Assignment\assignment3>
```

Notes

1. In the tagged output file (i.e., the output from the second phase which is testing) has '^' as a delimiter (the known ones have '/' by default) between word and tag for UNKNOWN words.
2. The probability of occurrence of an UNKNOWN word is assumed to be 0.1 in the testing file.