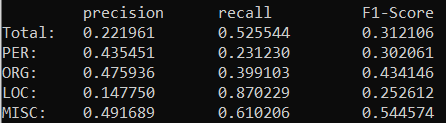
Statistical NLP Assignment 4 Report

Jiyuan Lu jl11046

Question 2:

Performance of the model:



Observations:

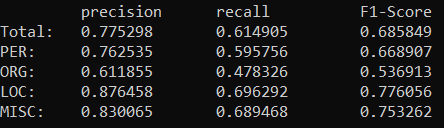
LOC has the lowest precision but the highest recall and it has the lowest F1-score.

MISC has the highest precision, and it has the highest F1-score.

PER has the lowest recall.

Question 4:

Performance of the model:



Observations:

For PER, ORG and MISC, their precision, recall and F1-score all improved.

For LOC, its precision greatly improved but its recall slightly degraded.

ORG now has the lowest precision, recall, and F1-score.

LOC now has the highest precision, recall, and F1-score.

By using an HMM tagger instead of a simple emission tagger, the improvement is the least for ORG, and the most for LOC.