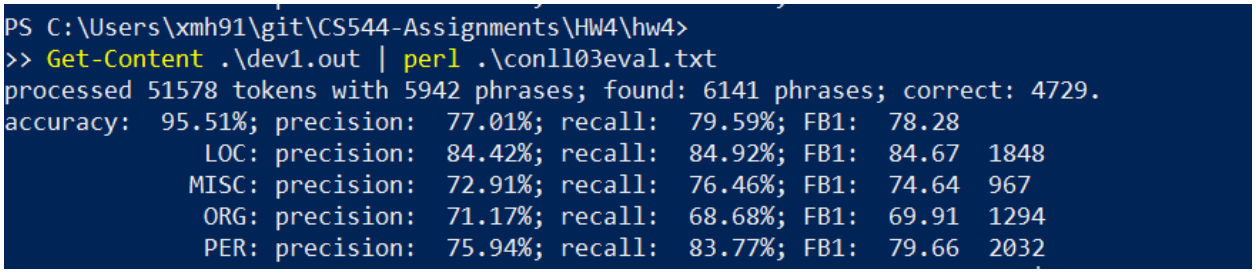
Q1. What are the precision, recall, and F1 score on the dev data? (hint: the reasonable F1 score on dev is 77%.

According to my dev1.out result, the precision is 77.01%, the recall is 79.59% and F1 is 78.28

Screenshot, same dev1.out included in the submission:

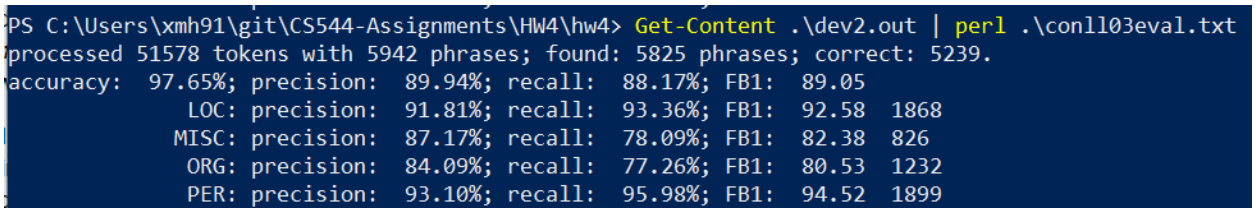


My Q1 vocab is based on training data only, I filtered out words that appear less than 3 times, and made them as “<unk>” and included it in the vocab.

For training, my epoch is 100, and my learning rate is 0.05, with no other hyperparameters involved. I’m feeding one sentence at a time for training, and no padding was used.

Q2 What are the precision, recall, and F1 score on the dev data? (hint: the reasonable F1 score on dev is 88%.

According to my dev2.out result, the precision is 89.94%, the recall is 88.17% and F1 is 89.05



My Q2 vocab is generated based on the train, dev, and test data, however, during the implementation, I chose to directly look up the embedding of a given word from the GloVe word embedding loaded from glove.6B.100d.txt, if the word is not in the glove key set, try lower case version, if so, use the lowercase embedding, if the lower case version of the word is still not in the key set, I picked “unk” embedding for the unknown word.

For training, my epoch is 50, and my learning rate is 0.05, with no other hyperparameters involved. I’m feeding one sentence at a time for training, and no padding was used.