

**writeup:**

1. Describe how you handled unknown words in hmm1.

We create for each word variations of: number, cnumber, ing, 2-unk, 4-unk, and saved the tag of word regard the taf of the variations.

if we see unknown word, we check the variations of the word and tag according the variation tag.

2. Describe your pruning strategy in the Viterbi hmm.

The pruning strategy is not to check all the tags for a word. We have checked all the possible tags by using the e.mle file - the tags we already seen before. For unknown words the tags will be only the tags that was seen for the variations.

3. Report your test scores when running each tagger on each dataset:

- a. The POS results:

1. GreedyTag:	92.79%
2. HMMTag:	94.02%
3. GreedyMaxEntTag:	96.44%
4. MEMMTag:	95.53%

- b. The NER results:

In the ner.txt file

4. Is there a difference in behavior between the hmm and maxent taggers? discuss.

The hmm is faster than the maxent tagger, but the accuracy of maxent tagger is better than the hmm tagger.

5. Is there a difference in behavior between the datasets? discuss.

The data-set of the ner gives better precisions from the POS data set.  
0.5%-1.0% better than the POS data set.

6. What will you change in the hmm tagger to improve accuracy on the named entities data?

Most of the tags in the Ner tag were O-tag. We notice that the mistakes when trying to tag I-tag but tagging O-tag instead.

Our improvement is to try first to tag I-tag first and not the O-tag or to normalize the amount of O-tag.

7. What will you change in the memm tagger to improve accuracy on the named entities data, on top of what you already did?

We train the model with features that related to the POS data, and so we used for tagging the data. The features are more fitting to the POS and not to the ner. We can change the features to be more related to the ner data. In the ner data we can notice of capital letters, and we can create feature with capital letters.

8. Why are span scores lower than accuracy scores?

fill this Q in the ner.txt file.