# Report

## The description of your tagger, including the architecture, the hyperparameters, etc.

For LSTM tagger, the architecture of the model is similar to the one introduced on PyTorch website (https://pytorch.org/tutorials/beginner/nlp/sequence\_models\_tutorial.html). During the data pre-processing, each sentence and its respecting tags is tokenized into two lists. After tokenization, every word and tag is then represented by an index. Therefore, we have 2 2-d lists for indexed words and indexed tags respectively. For the training part, indexed words are put into the LSTM model as input, while indexed tags are taken as the output.

Architecture of the LSTM model:  
1 embedding layer – 64 dimensional  
1 hidden layer – 64 dimensional  
1 output layer – 8 dimensional  
learning rate = 0.5  
no. of epoch = 30  
loss function = NLL loss  
optimizer = SGD

For transformer model, it is also similar to the one introduced on PyTorch website (https://pytorch.org/tutorials/beginner/transformer\_tutorial.html). However, the objective of the tutorial is different from our assignment. So a little modification is done on the data pre-processing and tokenization stage.

Architecture of the transformer model:  
1 embedding – 16 dimensional  
2 encoder layer – 16 dimensional  
1 decoder layer – 8 dimensional  
learning rate = 10  
learning rate decay rate (gamma) = 0.95  
no. of epoch = 30  
loss function = cross entropy loss  
optimizer = adam  
dropout = 0.2  
no. of attention heads = 2  
training batch size = 10  
evaluation batch size = 10

## The description of your settings, including the hyperparameters you have tried, the performance of your tagger on the dev set, etc.

The hyperparameters are designated for this task. The embedding size for both models and the no. of attention heads are small because each input sentence is quite short. Other hyperparameters are set by trying different values and see which one yields the highest testing accuracy. Due to time limitations, only few values could be experimented.

## The discussion and table of your results, including the performance of your tagger on the test set, the analysis of the results, etc.

The accuracy of the LSTM model and the transformer model are and respectively.