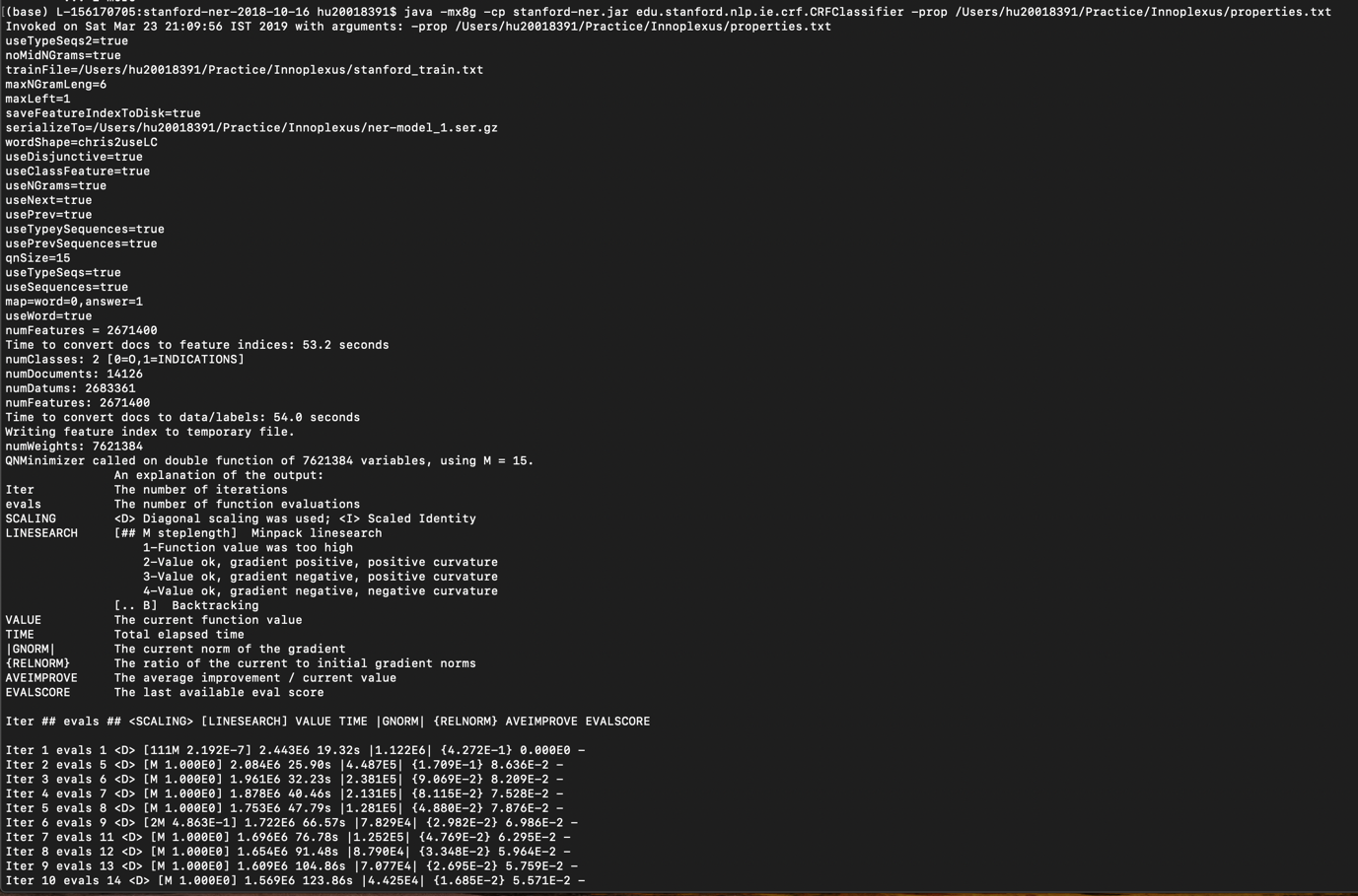
# Steps to Generate NER Model:

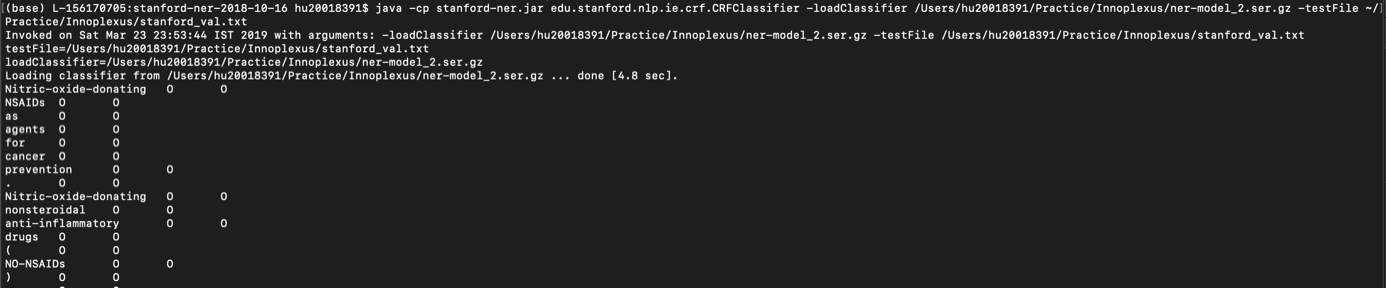
1. For the purpose of creating my model, I used **Standford CoreNLP** Library to implement CRF Classifier.
2. This is command line based tool, written in java.
3. To train a similar model, make sure Java is installed on your system.
4. Download the **Stanford CoreNLP** package. I’ve kept the package in my codes folder. Folder name is **stanford-ner-2018-10-16.**
5. To train a model, you’ll need a **properties.txt** file. This file contains the parameters which are used to initialize the model. Change the ‘**trainFile’** and ‘**serializeTo**’ parameters to the path of your train file and the path for the output model respectively.
6. **Stanford\_train.txt** file is the training file for the model. This file is created in **TSV** format. This is the standard format to train the CRF model. Code to generate this file is written in the iPython notebook.
7. To start training type following command in your terminal:

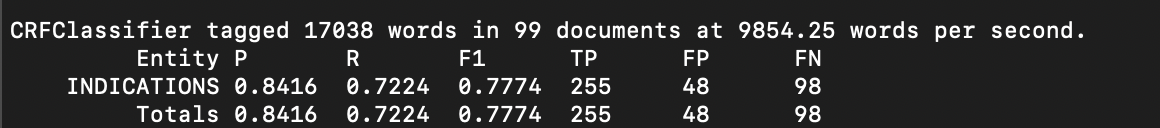
*java -mx8g -cp [path\_to\_stanford-ner-2018-10-16]/stanford-ner.jar [path\_to\_stanford-ner-2018-10-16]/edu.stanford.nlp.ie.crf.CRFClassifier -prop [path\_to\_properties.txt]*

****

1. I’ve created a **Stanford\_val.txt** to run cross validation on my generated model. You can do so by running the following command:

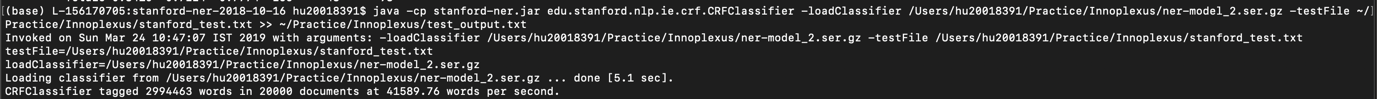
*Java -cp* *[path\_to\_stanford-ner-2018-10-16]/stanford-ner.jar [path\_to\_stanford-ner-2018-10-16]/edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier [path\_to\_classifier] -testFile [path\_to\_val\_file]*





1. To generate output, you can load the model into python and generate output as shown my iPython notebook. Or you can format the test file similar to validation file and generate the output using command line by running the following command:

*Java -cp* *[path\_to\_stanford-ner-2018-10-16]/stanford-ner.jar [path\_to\_stanford-ner-2018-10-16]/edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier [path\_to\_classifier] -testFile [path\_to\_test\_file] >> [path\_to\_prediction\_file]*



1. Funciton to convert the generated **INDICATIONS** tag into **BIO** format is written in the iPython notebook.
2. Refer to iPython Notebook for all the functions used to generate the train, test and validation files.