**Using the code:**

All the new code for this assignment exists in “NamedEntityClassifier.java”.

Command Line Options

* Required:
  + –path <DATA PATH>
    - where <DATA PATH> is the location of the root data folder
      * MUC6: “../data/muc6/data/”
      * ConLL: “../data/conll/”
      * Twitter: “../data/twitter/”
  + –corpus <CORPUS>
    - where <CORPUS> is the named of the specific corpus
      * MUC6: “muc6”
      * ConLL Spanish: “conll-esp”
      * ConLL Dutch: “conll-ned”
      * Twitter: “twitter”
  + –test <TEST SET>
    - where <TEST SET> is the type of test set desired
      * Single Dev File: “dryrun-single”
      * Development: “dryrun”
      * Final: “formal”
* Optional:
  + –ngrams <NUM GRAM>
    - where <NUM GRAM> is the ngram n-value. (ex: 2, 3, 4, 5 …)
    - Defaults to n=2

Output:

* The output file is named with the extension “-tagged-<NUM GRAM>.txt”.
  + Example: “ned.testb-tagged-2.txt” is the formal, conll-ned, bigram test.

Running the Evaluator:

* cd data/
* perl bin/conlleval.txt –r < file.txt
  + Ex:
    - perl bin/conlleval.txt -r < twitter/ner-formal-tagged-2.txt