KDM Assignment 2&3

Student id:16269904
Class id:4
Name:Lakshmi Korrapati

Task 1

Input
a. The dog saw John in the park NLP Tasks:
a. Part-of-speech (POS) tagger b. Named entity recognizer (NER) c. Co-reference resolution system

a. The dog saw John in the park.
i)Part-of-speech (POS) tagger.
The:DT, dog:NN, saw:VBD, John:NNP, in:IN, the:DT, park:NN, .:.
ii) Named entity recognizer (NER).
John- PERSON.
iii) Co-reference resolution system
Dog--->John in the park

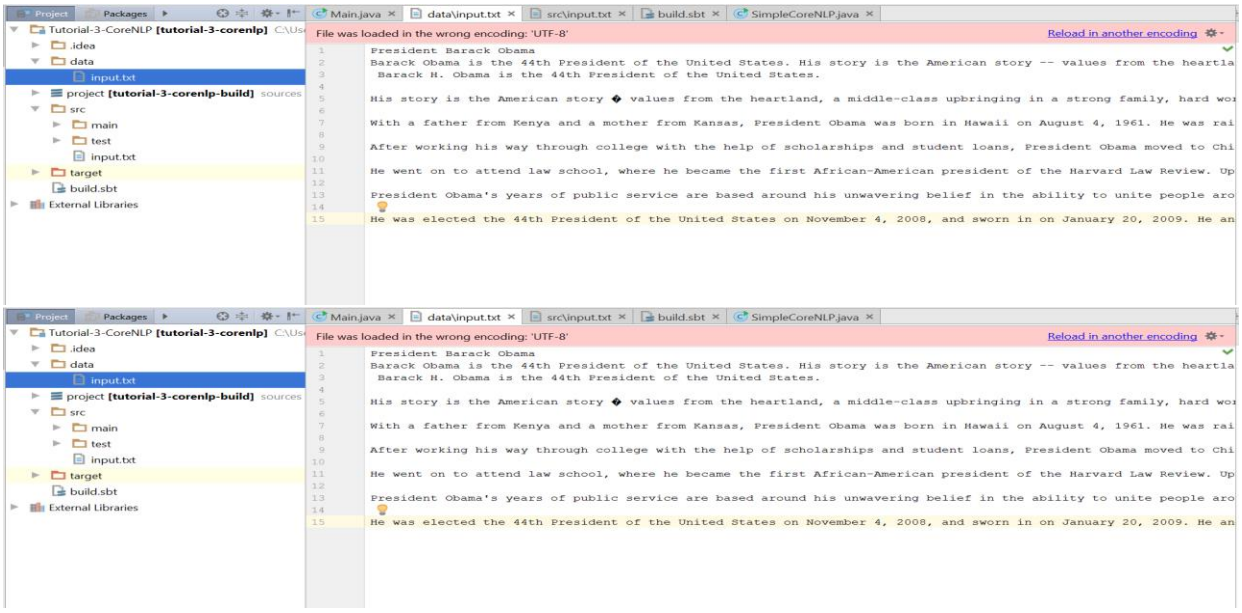
Task 2

Create a NLP project for the following tasks using CoreNLP.

What is Core NLP??

Stanford CoreNLP provides a set of natural language analysis tools. It can give the base forms of words, their parts of speech, whether they are names of companies, people, etc., normalize dates, times, and numeric quantities, mark up the structure of sentences in terms of phrases and word dependencies, indicate which noun phrases refer to the same entities, indicate sentiment, extract particular or open-class relations between entity mentions, get quotes people said, etc.

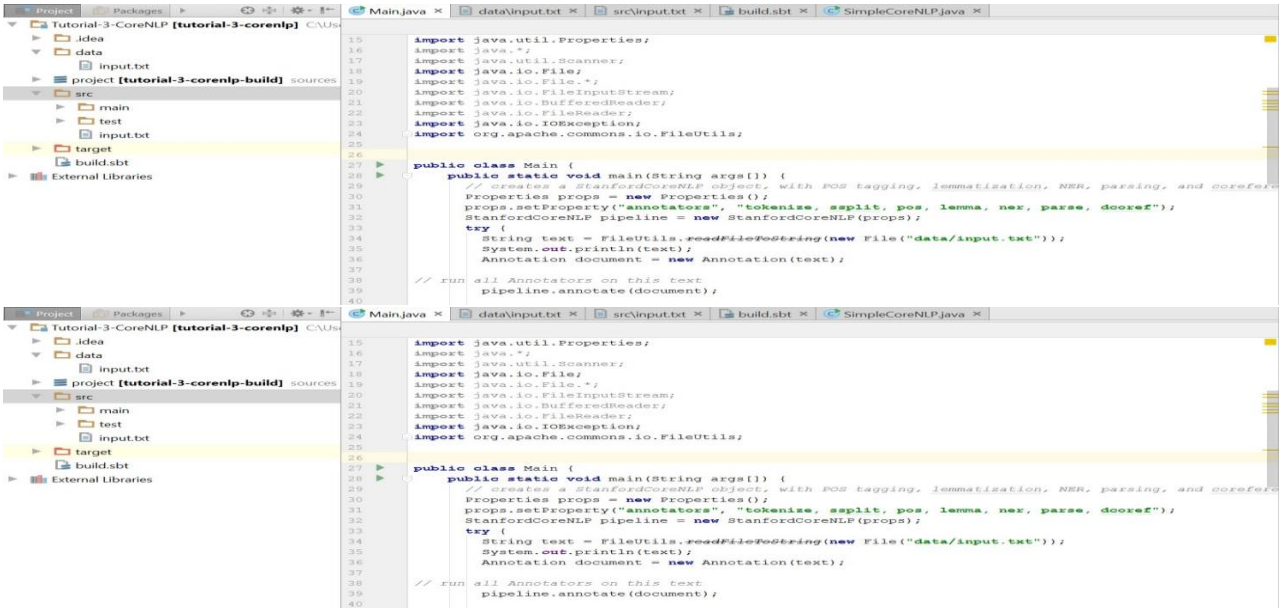
Input DataSet



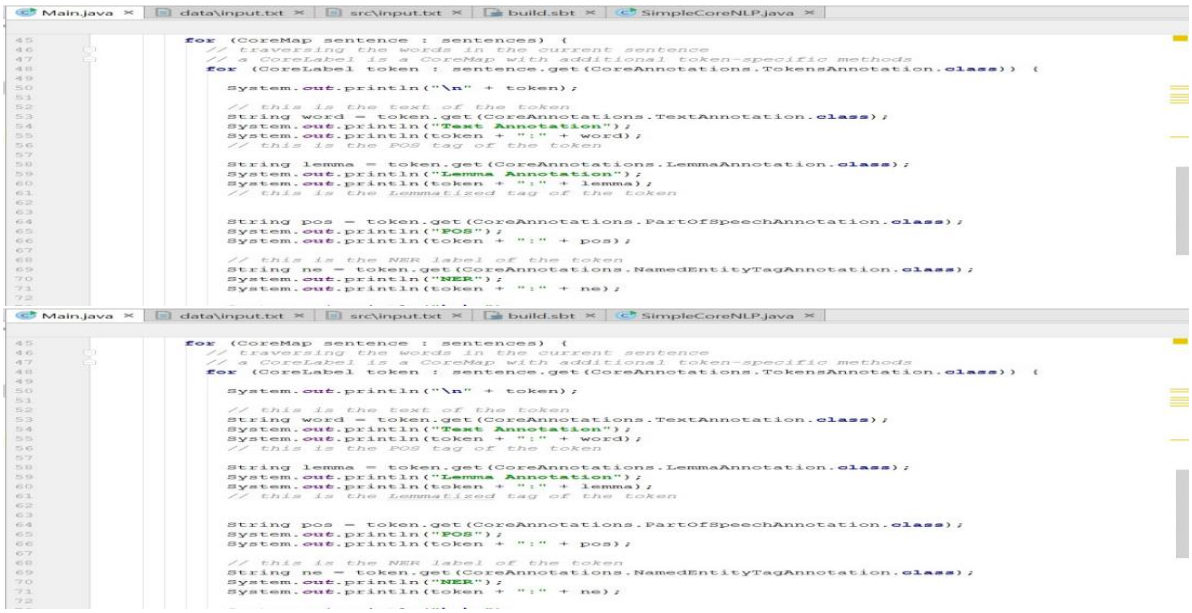
Program that performs the NLP tasks POS,NER,Coreferencing,Lemmatization

The below program reads an input dataset and performs POS,NER,Coreferencing,Lemmatization.

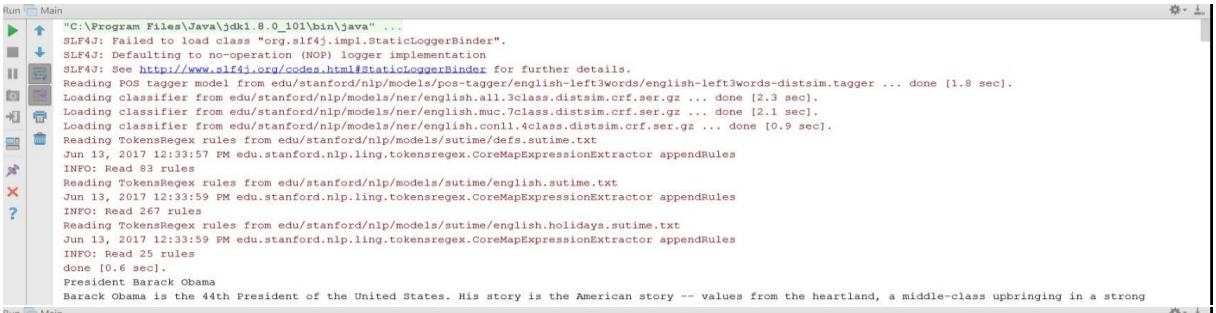
Parsing the Input



Program



Sucessful Program Execution

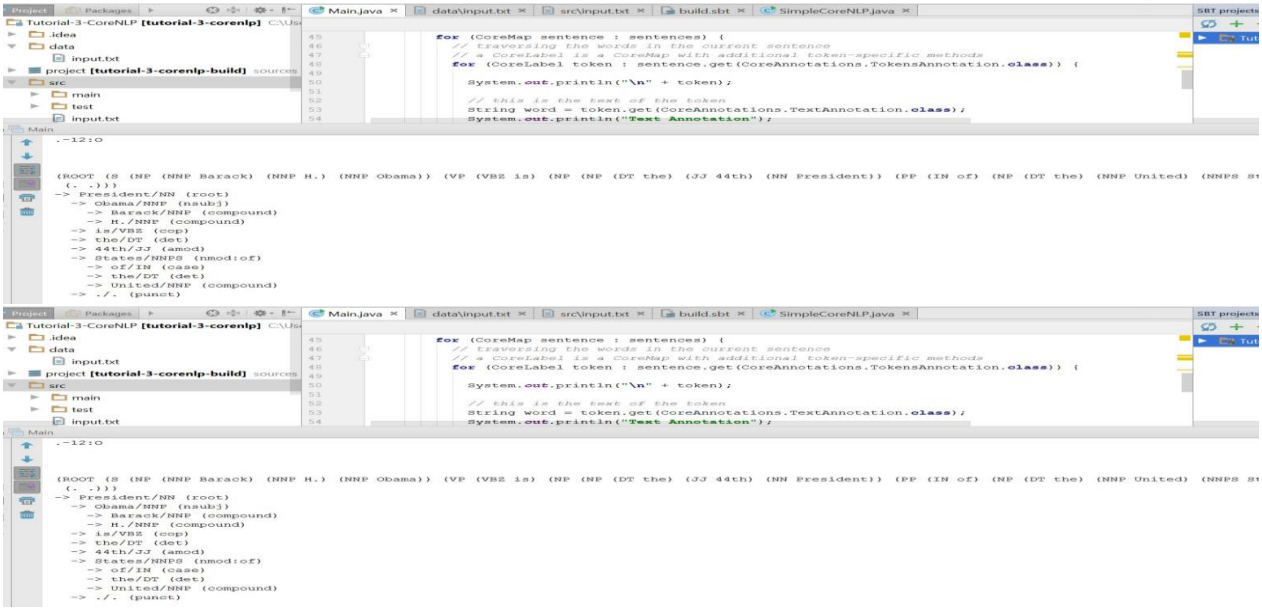


"C:\Program Files\Java\jdk1.8.0_101\bin\java" ...
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
Reading POS tagger model from edu/stanford/nlp/models/pos-tagger/english-left3words/english-left3words-distsim.tagger ... done [1.8 sec].
Loading classifier from edu/stanford/nlp/models/ner/english.all.3class.distsim.crf.ser.gz ... done [2.3 sec].
Loading classifier from edu/stanford/nlp/models/ner/english.muc.7class.distsim.crf.ser.gz ... done [2.1 sec].
Loading classifier from edu/stanford/nlp/models/ner/english.conll.4class.distsim.crf.ser.gz ... done [0.9 sec].
Reading TokensRegex rules from edu/stanford/nlp/models/sutime/defs.sutime.txt
Jun 13, 2017 12:33:57 PM edu.stanford.nlp.ling.tokensregex.CoreMapExpressionExtractor appendRules
INFO: Read 83 rules
Reading TokensRegex rules from edu/stanford/nlp/models/sutime/english.sutime.txt
Jun 13, 2017 12:33:59 PM edu.stanford.nlp.ling.tokensregex.CoreMapExpressionExtractor appendRules
INFO: Read 267 rules
Reading TokensRegex rules from edu/stanford/nlp/models/sutime/english.holidays.sutime.txt
Jun 13, 2017 12:33:59 PM edu.stanford.nlp.ling.tokensregex.CoreMapExpressionExtractor appendRules
INFO: Read 25 rules
done [0.6 sec].
President Barack Obama
Barack Obama is the 44th President of the United States. His story is the American story -- values from the heartland, a middle-class upbringing in a strong

POS,NER,Lemmatization Output

Tutorial-3-CoreNLP [tutorial-3-corenlp]
.idea
data
input.txt
project [tutorial-3-corenlp-build] sources
src
main
test
input.txt
for (CoreMap sentence : sentences) {
// traversing the words in the current sentence
// a CoreLabel is a CoreMap with additional token-specific methods
for (CoreLabel token : sentence.get(CoreAnnotations.TokensAnnotation.class
System.out.println("\n" + token);
// this is the text of the token
String word = token.get(CoreAnnotations.TextAnnotation.class);
System.out.println("Text Annotation");
President-1:O
Barack-2
Text Annotation
Barack-2:Barack
Lemma Annotation
Barack-2:Barack
POS
Barack-2:NNP
NER
Barack-2:PERSON

Dependency Tree Output



Coreferencing Output