

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

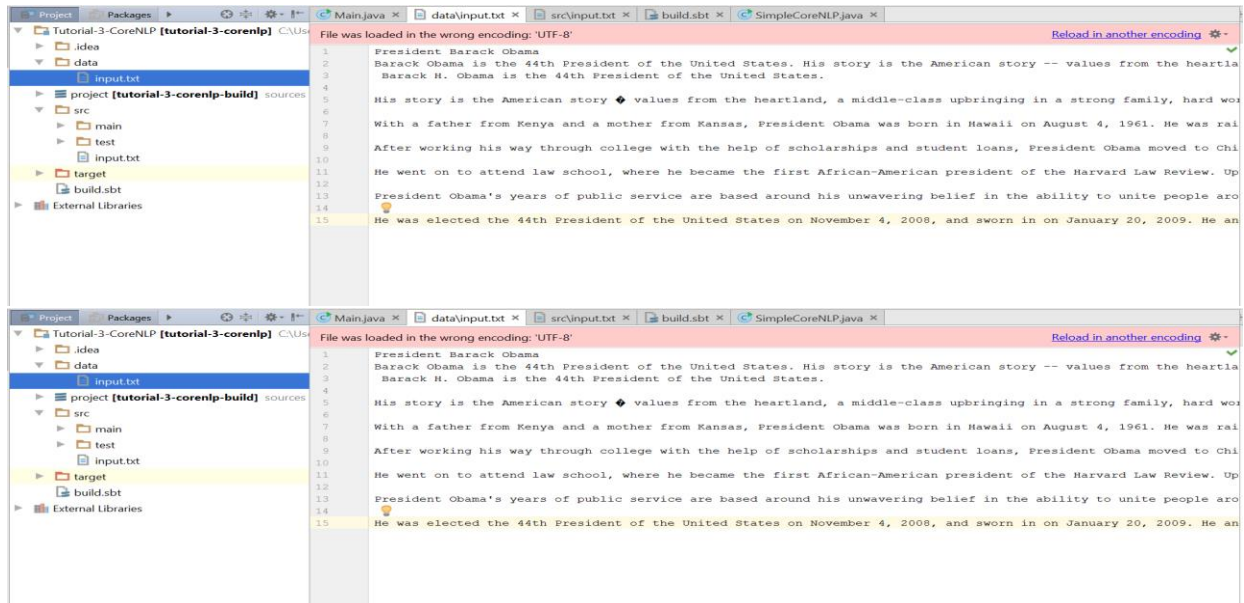
Task 2

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



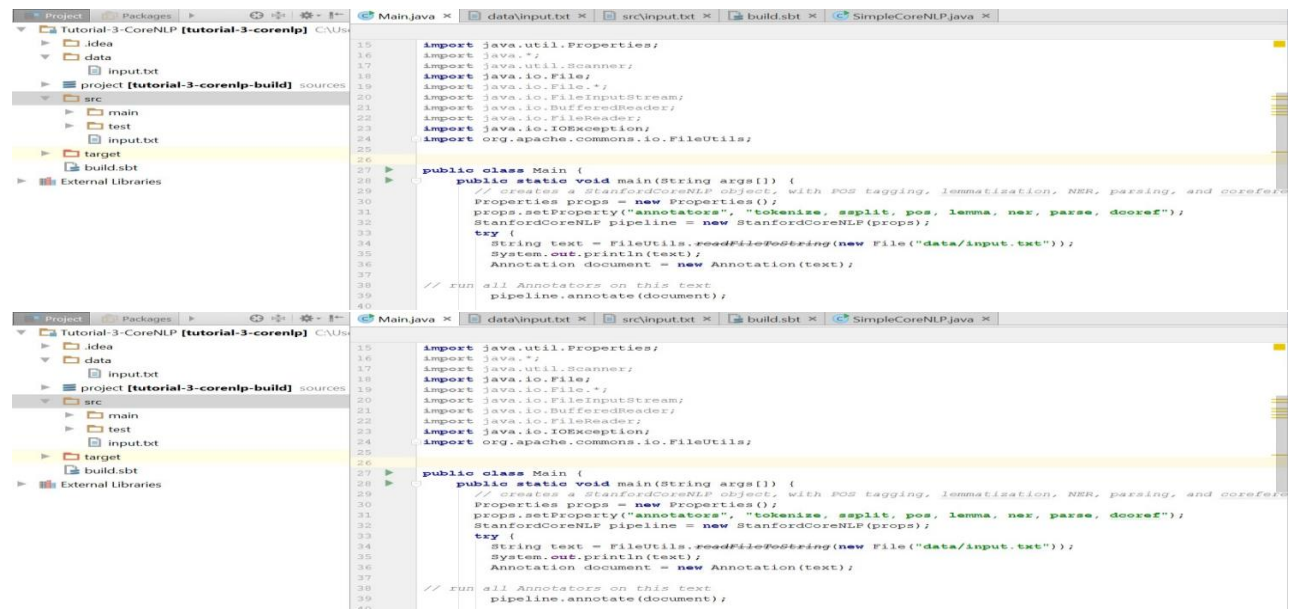
The screenshot shows an IDE window with a project named "Tutorial-3-CoreNLP [tutorial-3-corenlp]". The project structure includes a "data" directory containing "input.txt". The "input.txt" file is open, showing text about Barack Obama. A warning message at the top of the editor indicates "File was loaded in the wrong encoding: 'UTF-8'". The text in the file is as follows:

```
1 President Barack Obama
2 Barack Obama is the 44th President of the United States. His story is the American story -- values from the heartla
3 Barack H. Obama is the 44th President of the United States.
4
5 His story is the American story ♦ values from the heartland, a middle-class upbringing in a strong family, hard wor
6
7 With a father from Kenya and a mother from Kansas, President Obama was born in Hawaii on August 4, 1961. He was rai
8
9 After working his way through college with the help of scholarships and student loans, President Obama moved to Chi
10
11 He went on to attend law school, where he became the first African-American president of the Harvard Law Review. Up
12
13 President Obama's years of public service are based around his unwavering belief in the ability to unite people aro
14
15 He was elected the 44th President of the United States on November 4, 2008, and sworn in on January 20, 2009. He an
```

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

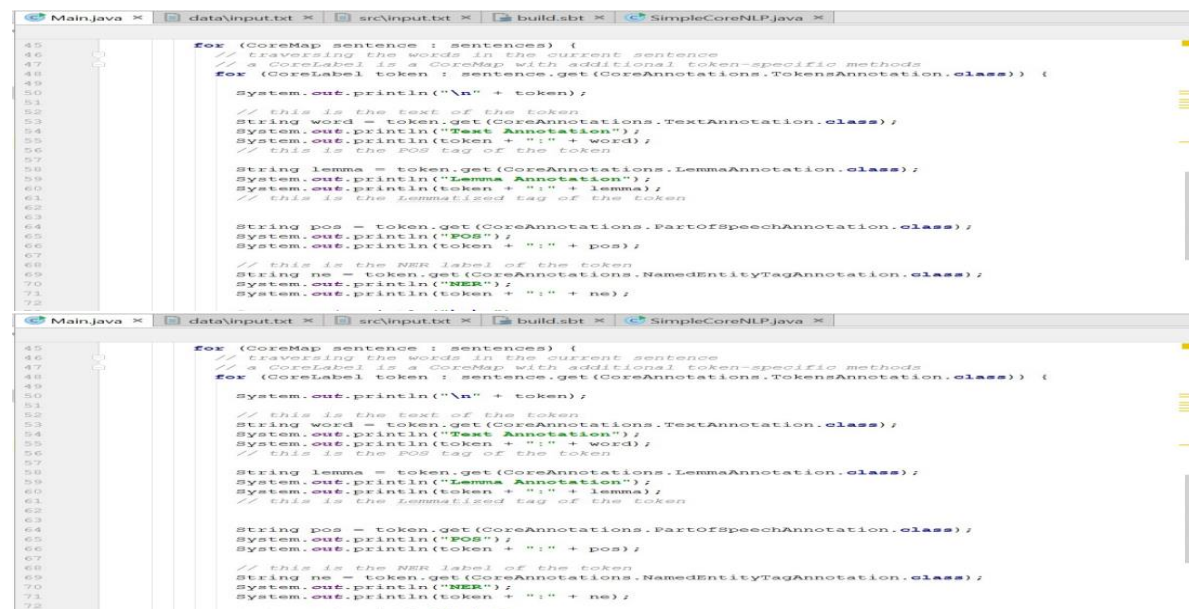
The below program reads an input dataset and performs POS,NER,Coreferencing,Lemmatization.

Parsing the Input



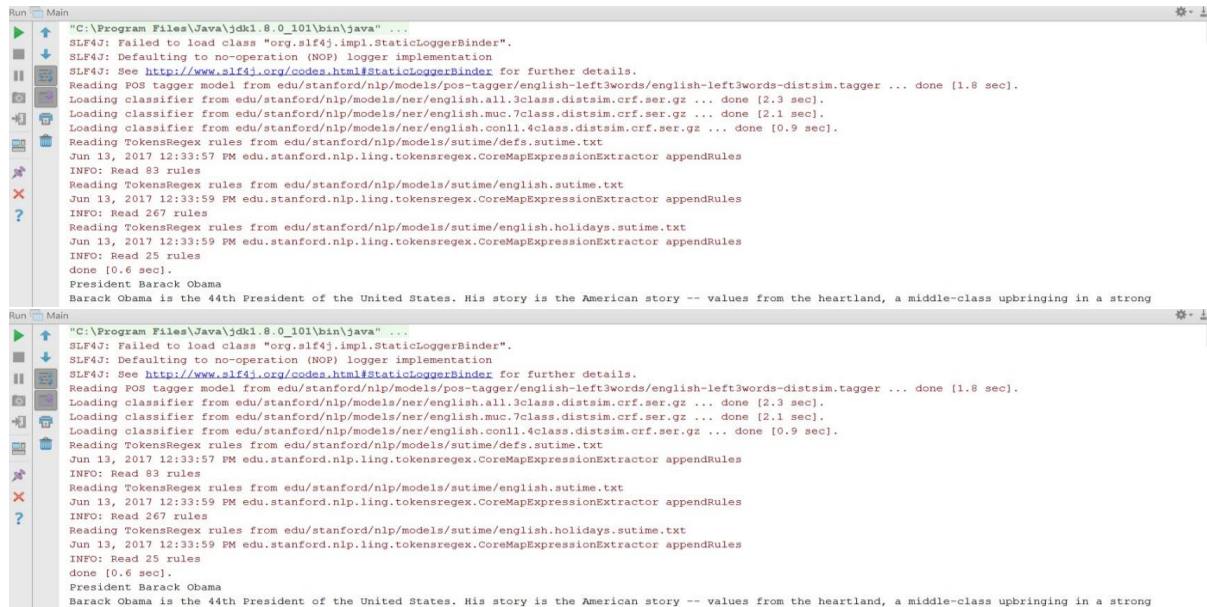
```
15 import java.util.Properties;
16 import java.*;
17 import java.util.Scanner;
18 import java.io.File;
19 import java.io.File.*;
20 import java.io.FileInputStream;
21 import java.io.BufferedReader;
22 import java.io.IOException;
23 import org.apache.commons.io.FileUtils;
24
25
26 public class Main {
27     public static void main(String args[]) {
28         // creates a StanfordCoreNLP object, with POS tagging, lemmatization, NER, parsing, and coreference
29         Properties props = new Properties();
30         props.setProperty("annotators", "tokenize, pos, lemma, ner, parse, decoref");
31         StanfordCoreNLP pipeline = new StanfordCoreNLP(props);
32         try {
33             String text = FileUtils.readFileToString(new File("data/input.txt"));
34             System.out.println(text);
35             Annotation document = new Annotation(text);
36             // run all Annotators on this text
37             pipeline.annotate(document);
38         } catch (IOException e) {
39             e.printStackTrace();
40         }
41     }
42 }
```

Program



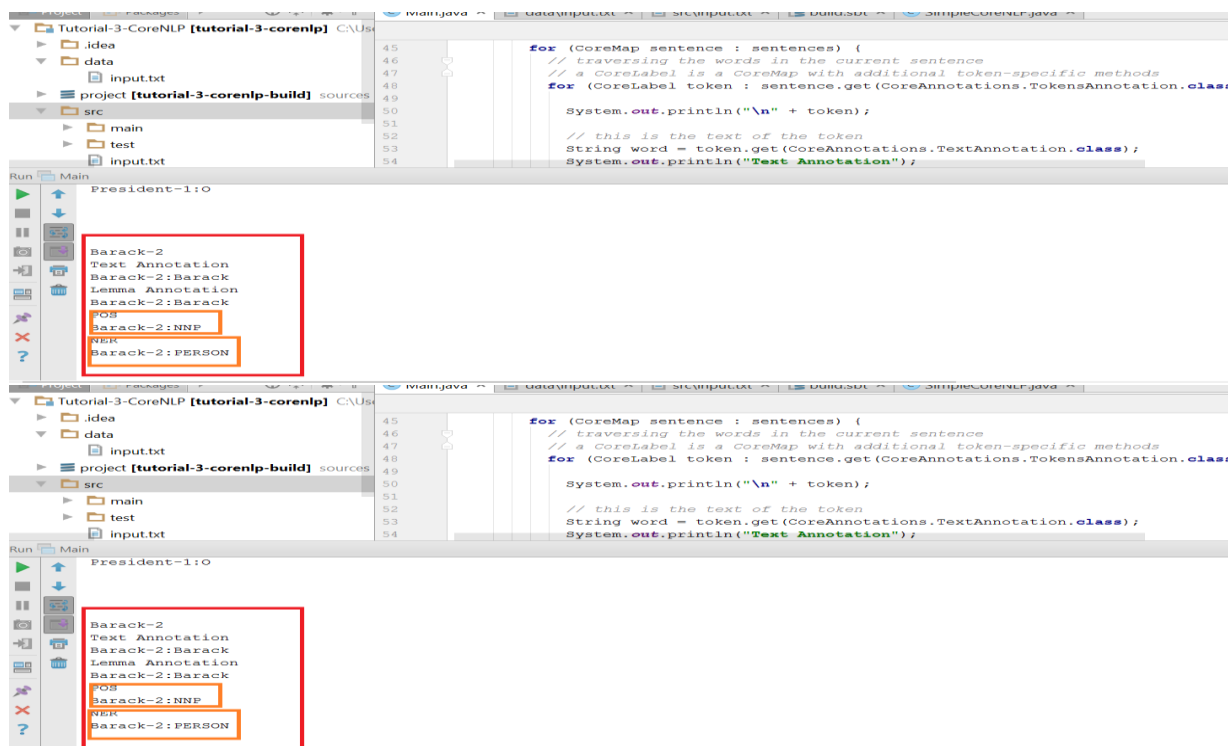
```
45 for (CoreMap sentence : sentences) {
46     // traversing the words in the current sentence
47     // a CoreLabel is a CoreMap with additional token-specific methods
48     for (CoreLabel token : sentence.get(CoreAnnotations.TokensAnnotation.class)) {
49         System.out.println("\n" + token);
50         // this is the text of the token
51         String word = token.get(CoreAnnotations.TextAnnotation.class);
52         System.out.println("Text Annotation");
53         System.out.println(token + " | " + word);
54         // this is the POS tag of the token
55         String pos = token.get(CoreAnnotations.PartOfSpeechAnnotation.class);
56         System.out.println("POS");
57         System.out.println(token + " | " + pos);
58         // this is the NER label of the token
59         String ne = token.get(CoreAnnotations.NamedEntityTagAnnotation.class);
60         System.out.println("NER");
61         System.out.println(token + " | " + ne);
62     }
63 }
64
65
66 for (CoreMap sentence : sentences) {
67     // traversing the words in the current sentence
68     // a CoreLabel is a CoreMap with additional token-specific methods
69     for (CoreLabel token : sentence.get(CoreAnnotations.TokensAnnotation.class)) {
70         System.out.println("\n" + token);
71         // this is the text of the token
72         String word = token.get(CoreAnnotations.TextAnnotation.class);
73         System.out.println("Text Annotation");
74         System.out.println(token + " | " + word);
75         // this is the POS tag of the token
76         String pos = token.get(CoreAnnotations.PartOfSpeechAnnotation.class);
77         System.out.println("POS");
78         System.out.println(token + " | " + pos);
79         // this is the NER label of the token
80         String ne = token.get(CoreAnnotations.NamedEntityTagAnnotation.class);
81         System.out.println("NER");
82         System.out.println(token + " | " + ne);
83     }
84 }
```

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



```
President-1:0
Barack-2
Text Annotation
Barack-2:Barack
Lemma Annotation
Barack-2:Barack
POS
Barack-2: NNP
NER
Barack-2: PERSON
```

Dependency Tree Output

```
(ROOT (S (NP (NNE Barack) (NNP H.) (NNE Obama)) (VP (VBE is) (NP (NP (DT the) (JJ 44th) (NN President)) (PP (IN of) (NP (DT the) (NNE United) (NNPS States) (NN the))))) (PUNCT .)))
  → President/NN (root)
  → Obama/NNP (nsubj)
  → Barack/NNP (compound)
  → H./NNP (compound)
  → is/VBE (cop)
  → the/DT (det)
  → 44th/JJ (amod)
  → States/NNPS (nmod:of)
  → of/IN (case)
  → the/DT (det)
  → United/NNP (compound)
  → ./. (punct)
```

Coreferencing Output

```
the United States" in sentence 1, "his" in sentence 2, "the 44th President of the United States" in sentence 3, "President Obama" in sentence 5, "President Obama" in sentence 7, "President Obama" in sentence 10, "his" in sentence 11, "the 44th President of the United States" in sentence 13, CHAIN1["the United States" in sentence 1, "the United States" in sentence 3, "United States" in sentence 12, "the United States" in sentence 13], CHAIN2["this story" in sentence 2, "the American story -- values from the heartland, a middle-class upbringing in a strong family, hard work and education as the means of getting ahead" in sentence 2, "this story" in sentence 4], CHAIN3["values from the heartland" in sentence 2], CHAIN4["the heartland" in sentence 2, "the heartland" in sentence 4], CHAIN5["a middle-class upbringing in a strong family" in sentence 2, "a middle-class upbringing in a strong family" in sentence 4], CHAIN6["a strong family" in sentence 2, "a strong family" in sentence 4], CHAIN7["the means" in sentence 4], CHAIN8["a life" in sentence 2, "a life" in sentence 4], CHAIN9["othera" in sentence 2], CHAIN10["othera" in sentence 2], CHAIN11["Barack H. Obama" in sentence 3, "his" in sentence 4], CHAIN12["the American story values from the heartland, a middle-class upbringing in a strong family, hard work and education as the means of getting ahead" in sentence 4], CHAIN13["othera" in sentence 4], CHAIN14["Kenya" in sentence 5], CHAIN15["Kenya" in sentence 5], CHAIN16["his" in sentence 6, "his" in sentence 6], CHAIN17["he" in sentence 7, "he" in sentence 7, "he" in sentence 8, "he" in sentence 9, "he" in sentence 9, "his" in sentence 9], CHAIN18["a mother from Kansas" in sentence 5], CHAIN19["August 4, 1961" in sentence 5], CHAIN20["Patton" in sentence 6], CHAIN21["this grandfather, who served in Patton's army, and his grandmother, who worked her way up from the secretarial pool to middle management at a bank" in sentence 6], CHAIN22["this grandfather" in sentence 6], CHAIN23["Patton's army" in sentence 6], CHAIN24["this grandmother, who worked her way up from the secretarial pool to middle management at a bank" in sentence 6], CHAIN25["her" in sentence 6], CHAIN26["her way" in sentence 6, "his way" in sentence 7], CHAIN27["the secretarial pool" in sentence 6], CHAIN28["a bank" in sentence 6], CHAIN29["Chicago" in sentence 7, "Chicago" in sentence 9], CHAIN30["the help of
```

