Submitted by

Prudhvi Suggula|SADANAND KALLAKURI|NIKITHA KONA|HARSHA |

Intelligent QA System

Increment 1

# **1.Motivation**

Humans are always in quest of knowledge. Information retrieval has become a form of knowledge discovery where we search on web to get relevant answers. Question Answering can be considered as an intersection of Natural Language Processing, Machine learning and artificial intelligence. QA systems are required in all the fields such as Health & Medical care, Education, Customer service as its always good to have assistance from computers.

# **1.1. Objective**

Our main objective is to design and implement a question answering system which can answer the posed questions accurately. We would start off with basic approach where our system can list the people, locations and organizations in the dataset, the we would refine our architecture by implementing word2 vec,TF-IDF,ngram techniques where the system can answer way better than the previous approach, we will look at the drawbacks of this system and refine our architecture to yield the best results.

# **1.2. Significance**

Question and Answering systems are widely used now and they are trending in the field of Artificial intelligence and Machine learning. Few such examples would be Siri, IBM Watson, Amazon Echo that have done tremendously well and beat the national champions in many Atari games. Their significance lies in understanding the user queries and giving the best relevant results in short span of time. There has been a tremendous improvement that has taken place and still we are trying to minimize the gap between completeness and correctness.

# **2.Question Answering Application**

We are implementing a question and answering system on Sports where the user can ask about the current matches held on various games, live scores ,locations of the games, ticket availability, weather conditions. We will start with the basic questionnaire system and incrementally build an enriched QA system.

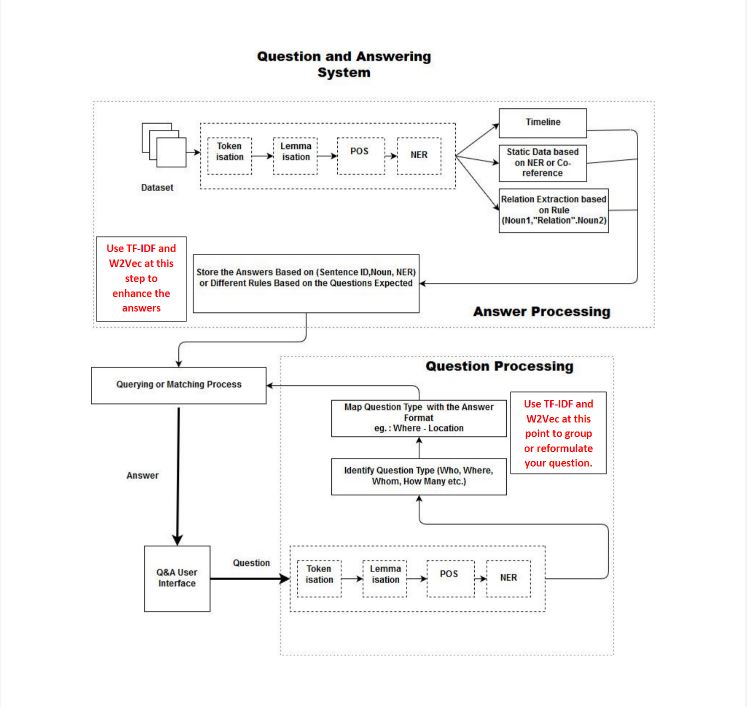
# **3.Dataset**

We have chosen **wikiref150 and BBC Sports** datasets for our project.

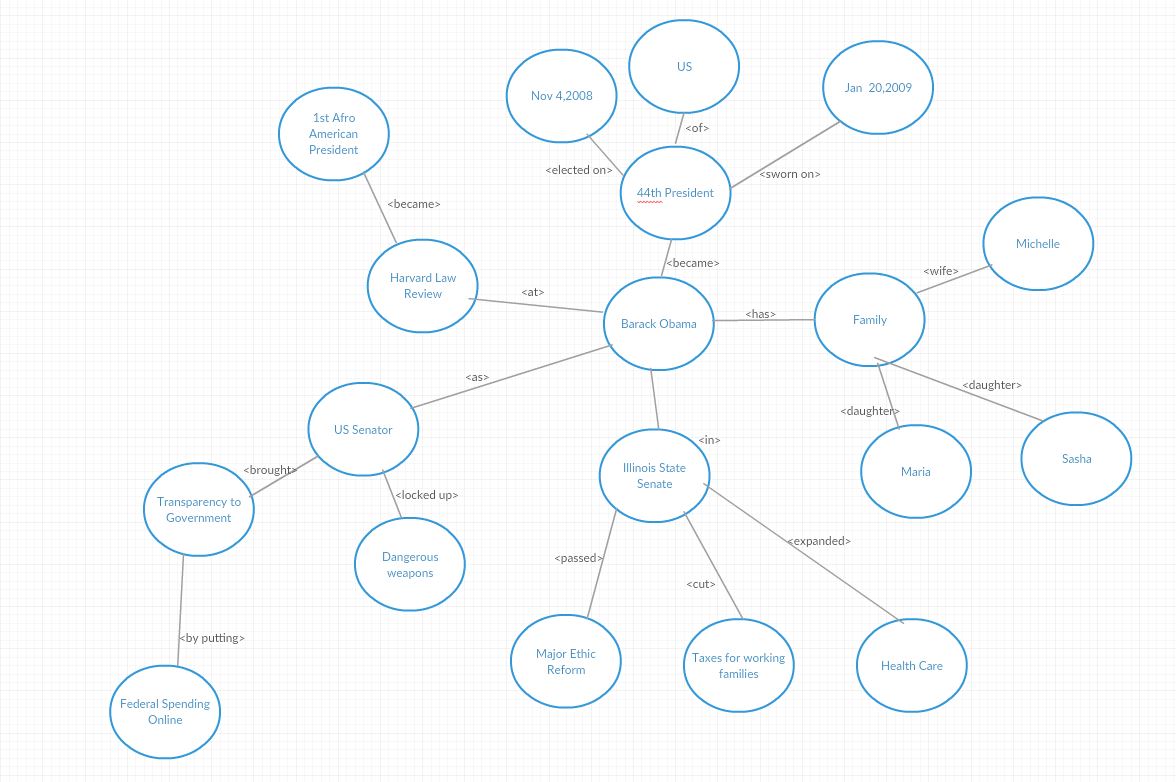
**BBC Sports Dataset Link:** <http://mlg.ucd.ie/datasets/bbc.html>

**Wikiref150 Dataset Link:** http://mklab.iti.gr/project/web-news-article-dataset

# **4.Workflow**



# **4.1 Knowledge Graph**



# **4.2 Question Answering**

The questions asked would be as follows.

1.Which location is the India vs Pakistan match held?

2. List the people that are playing the match?

3. What is the pitch condition there?

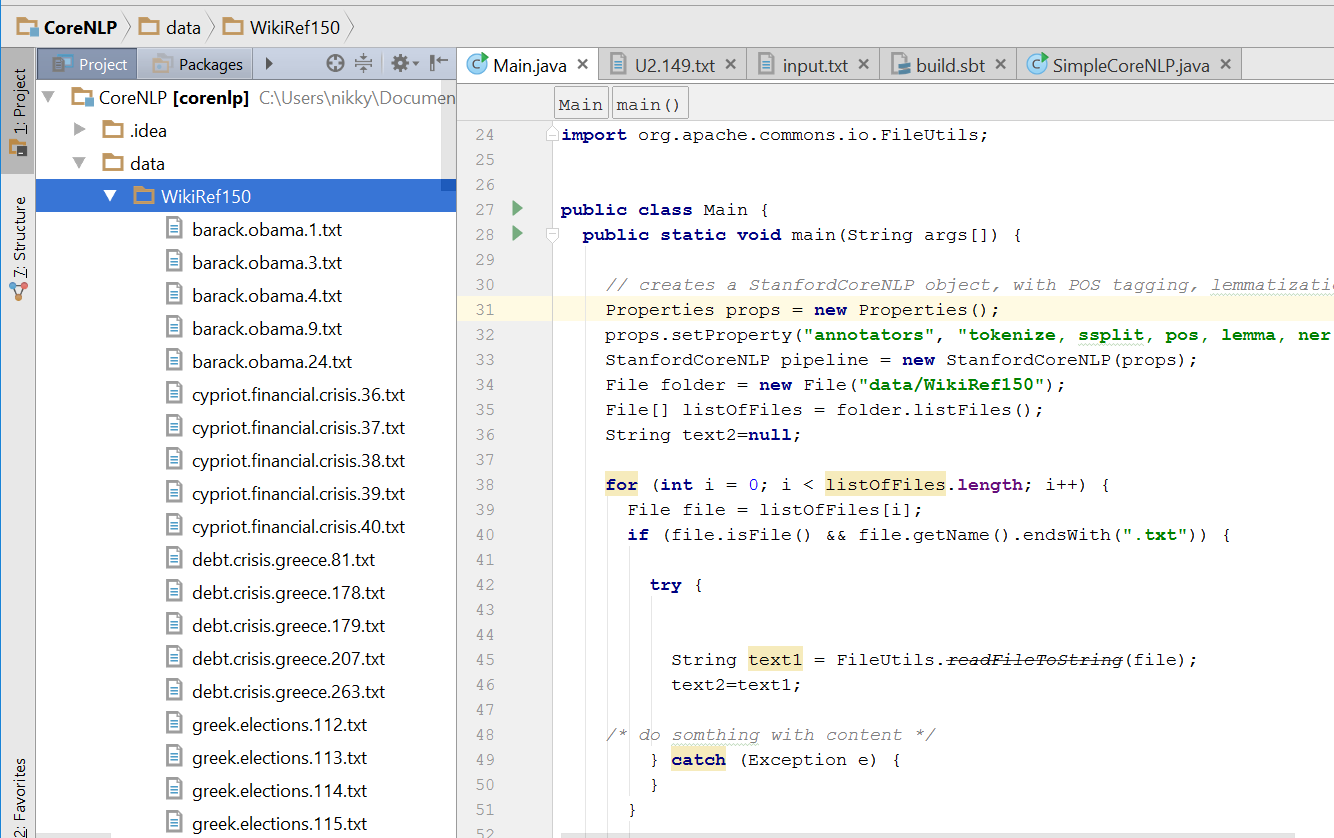
4.Who won the toss?

# **5.Implementation**

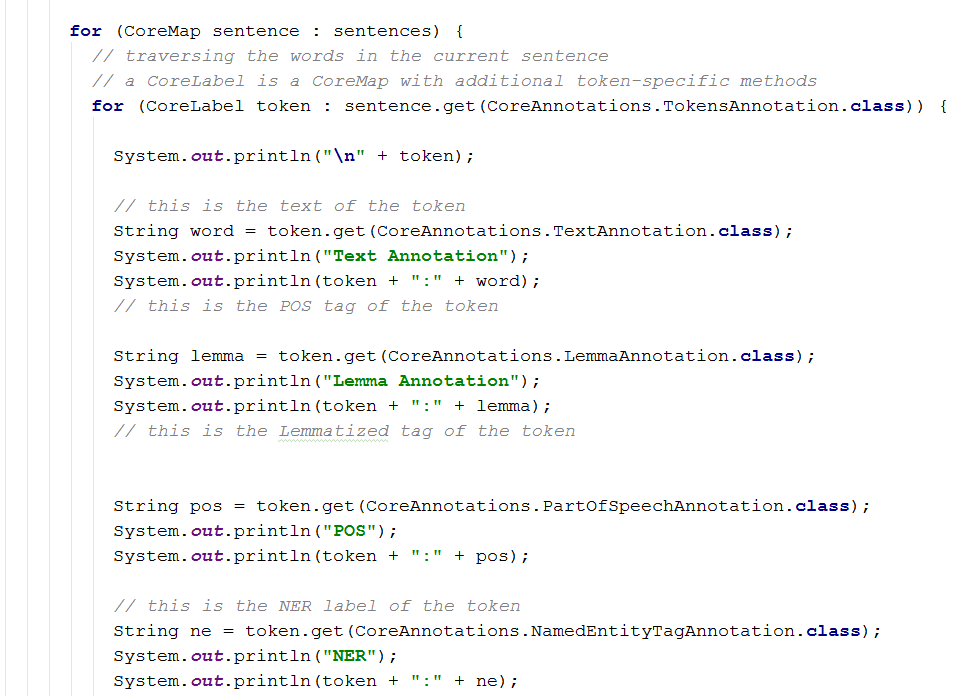
# **5.1 NLP with Dataset**

## **5.1.1Reading the Entire Dataset**

We have taken **wikiref150** dataset. We read the entire dataset and performed the all the NLP operations.

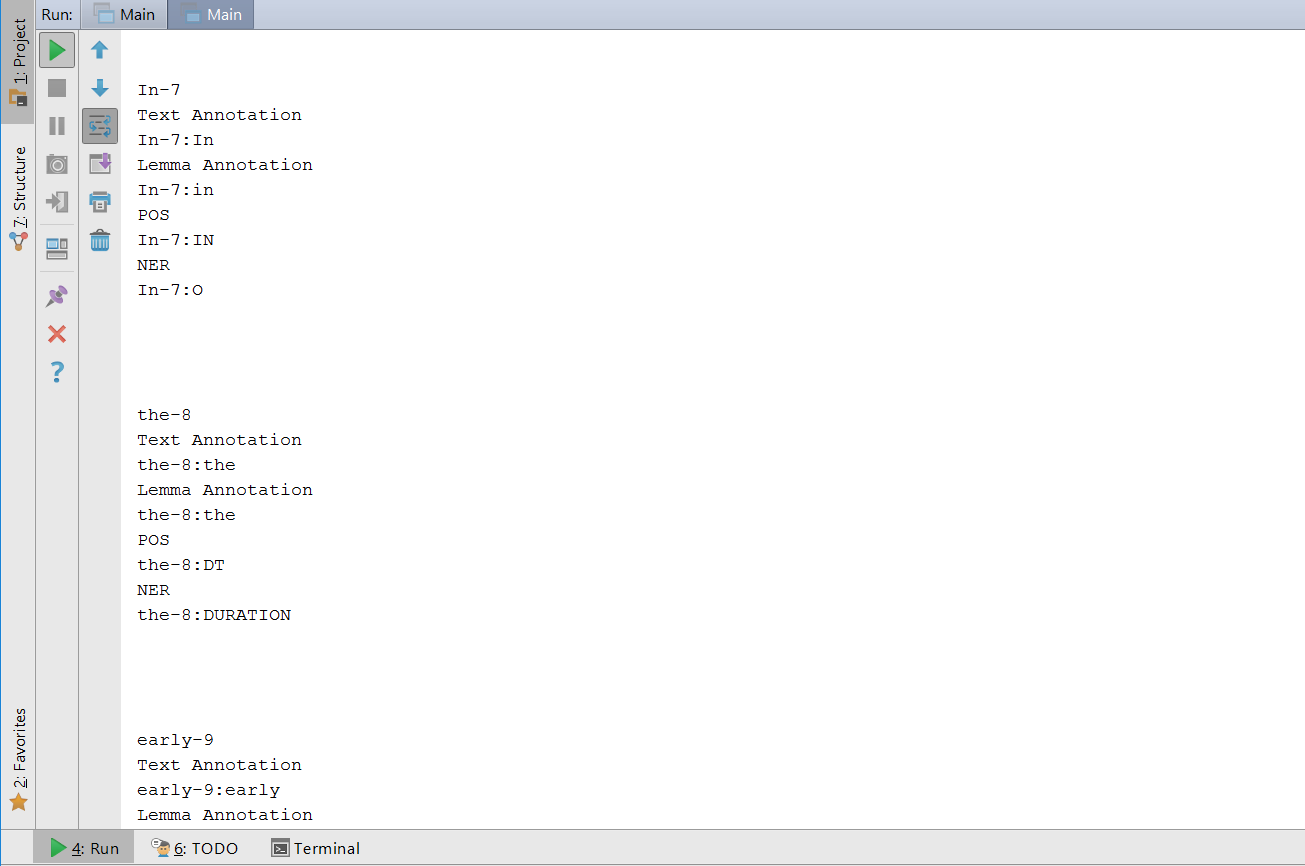


## **5.1.2 Program to Perform NLP Operations**

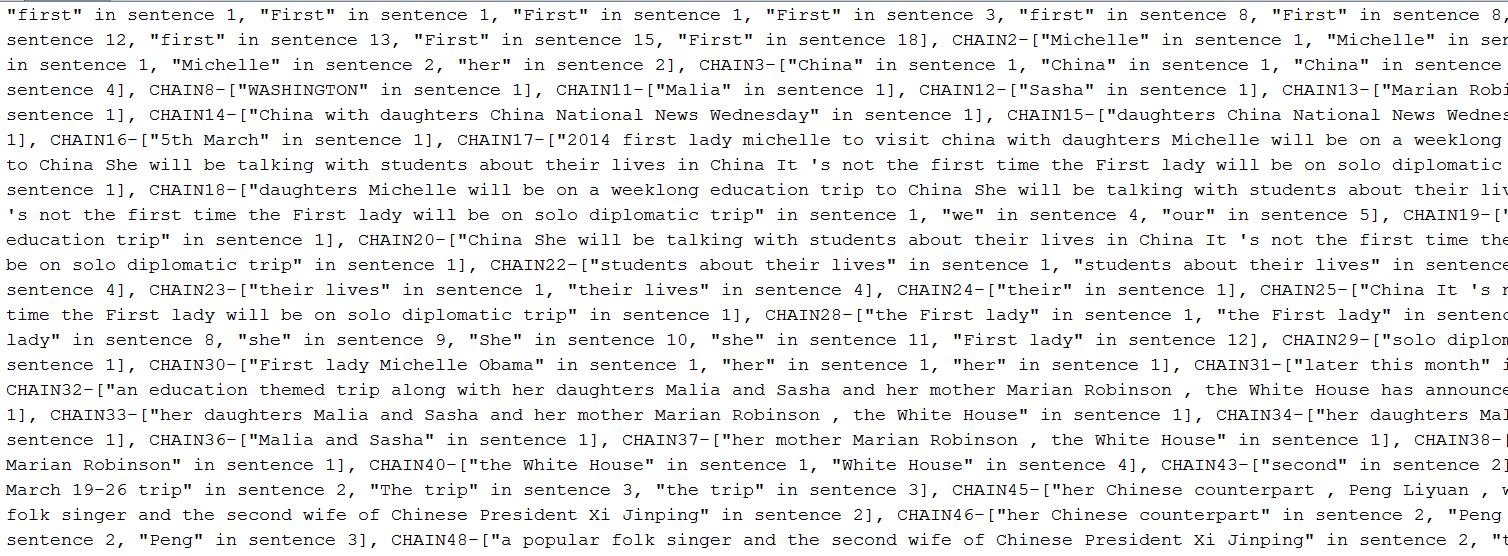


## **5.1.3 Output Screenshots**

### **5.1.3.1 Text Annotation,Lemma Annotation,POS,NER**



### **5.1.3.2 Co-referencing**

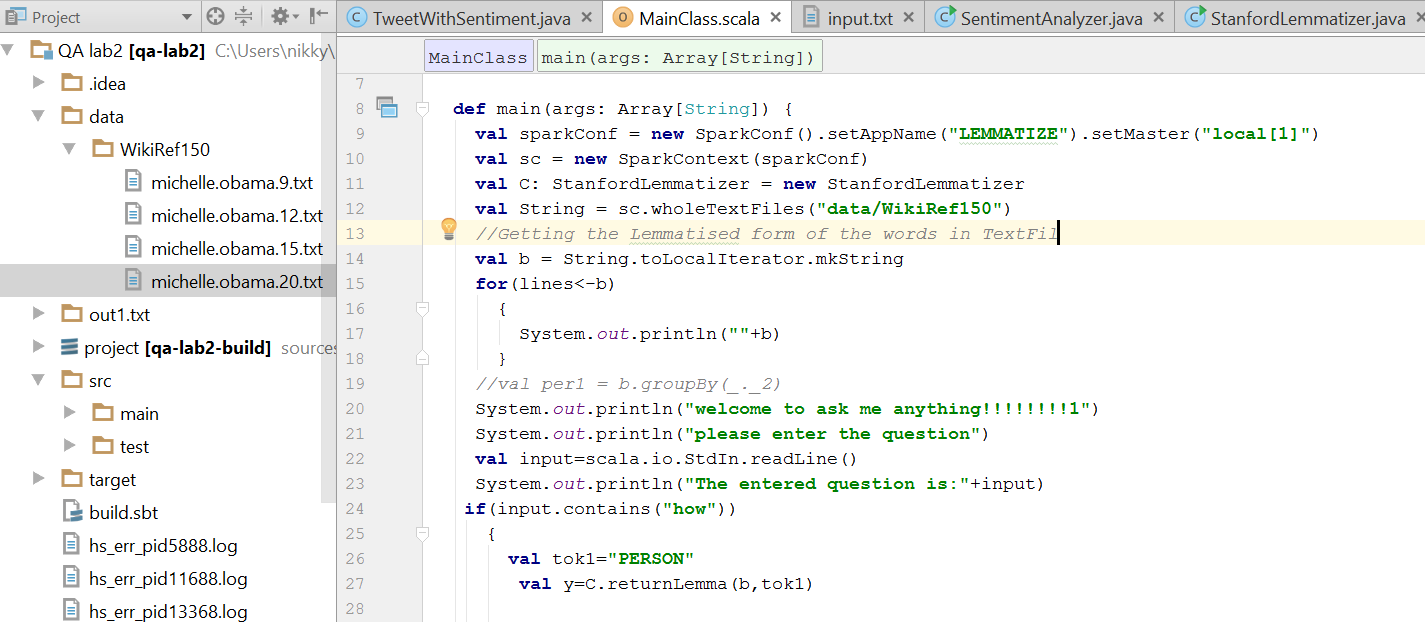


## **5.1.2 Question Answering**

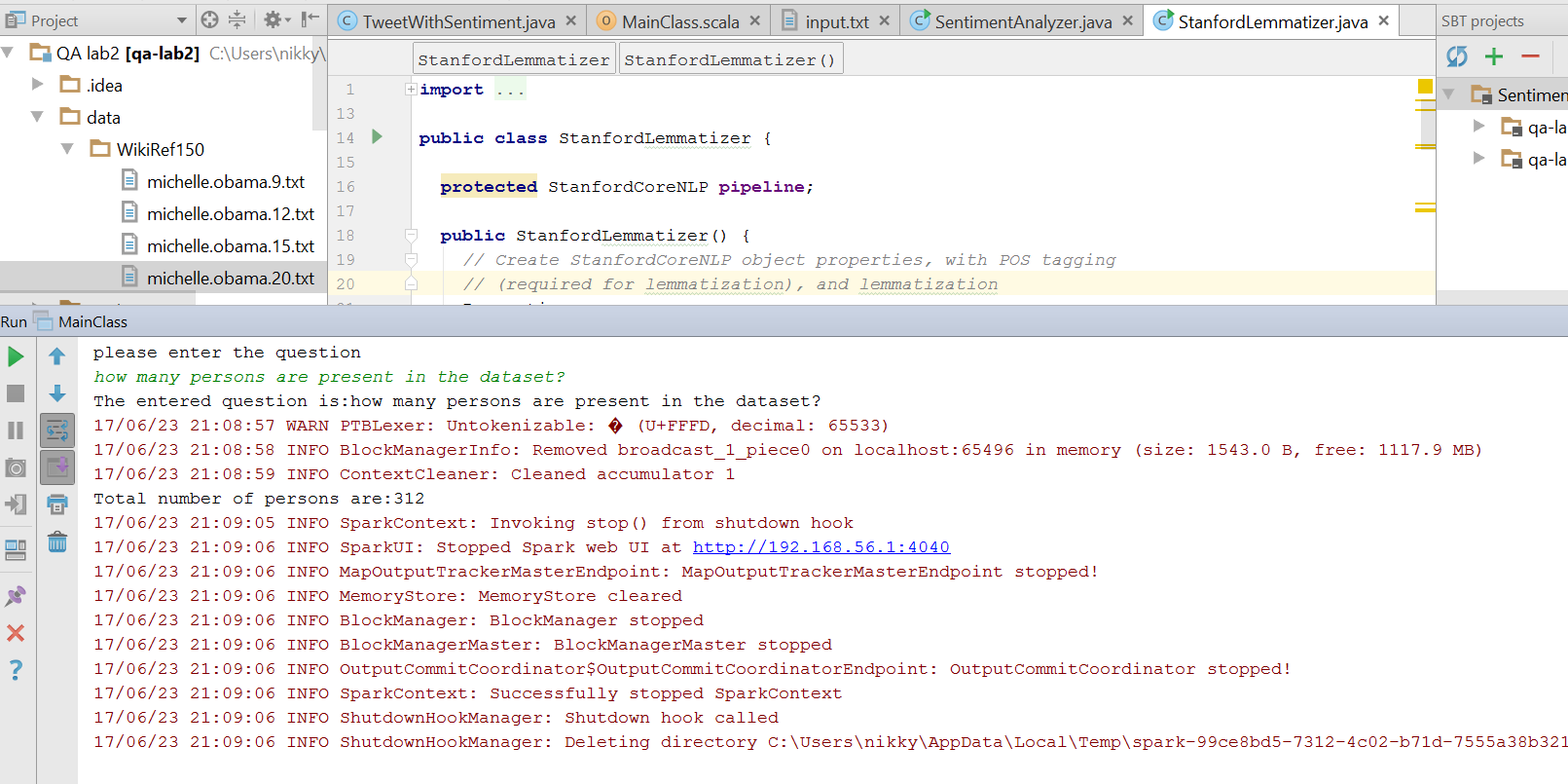
We have implemented a simple question answering system that answers basic questionaries’ asked to the system.

### **5.1.2.1 Scala Program Reading the Entire Dataset**

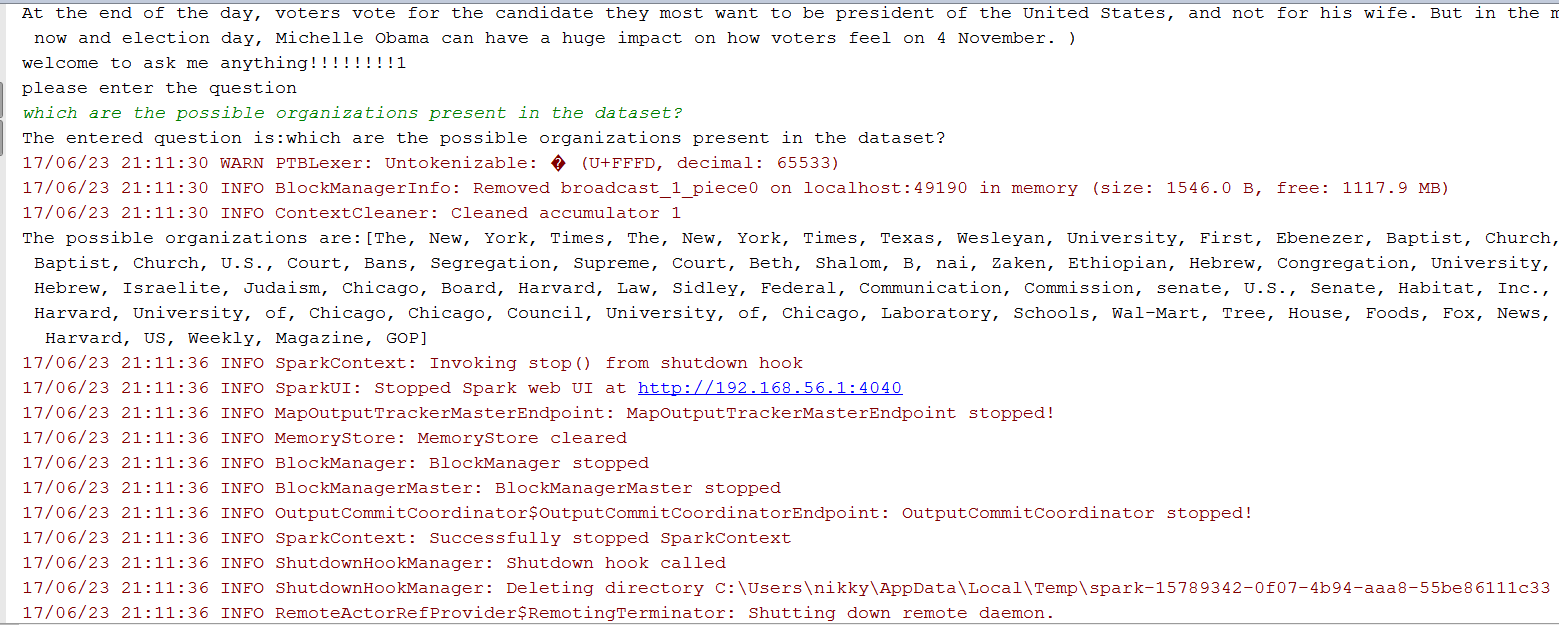
The below scala program takes the entire dataset and performs a basic question and answering.



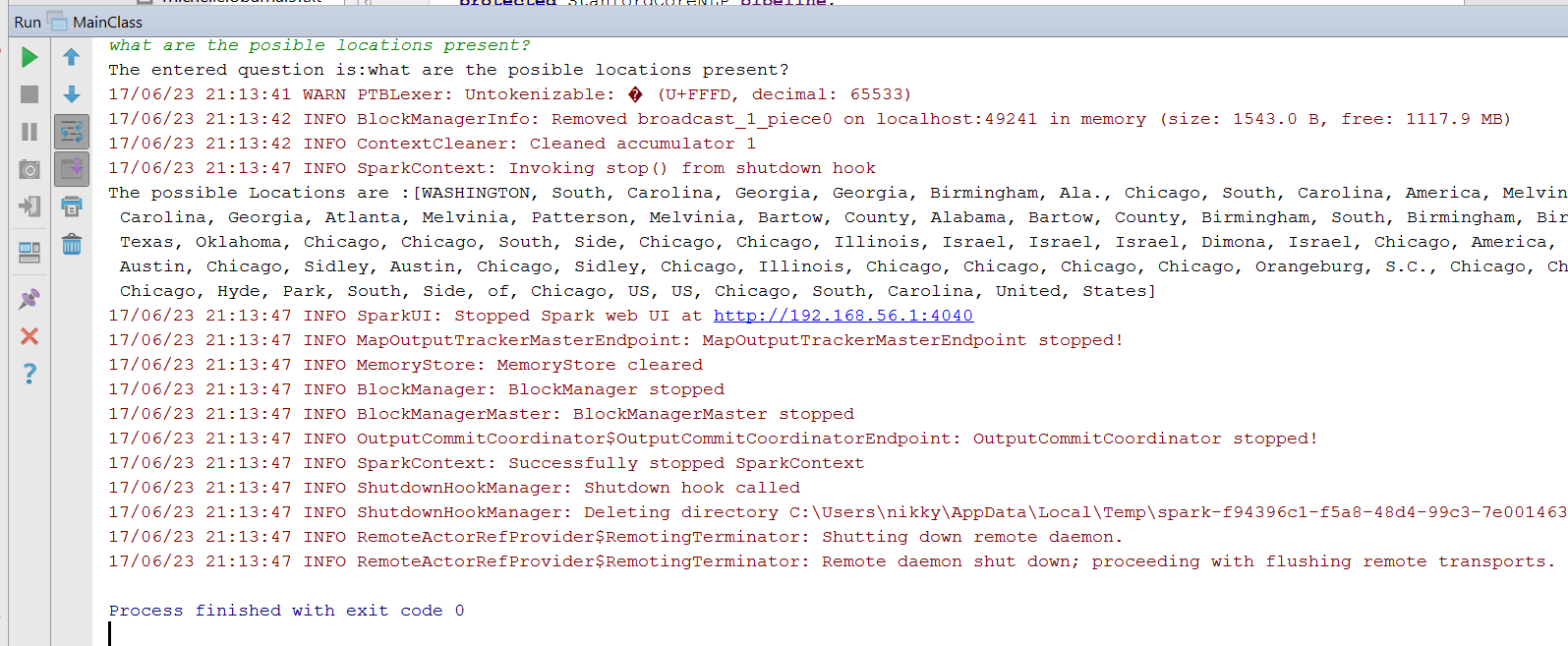
### **5.1.2.2Question1-How**



### **5.1.2.3 Question-Which**



### **5.1.2.4 Question-What**

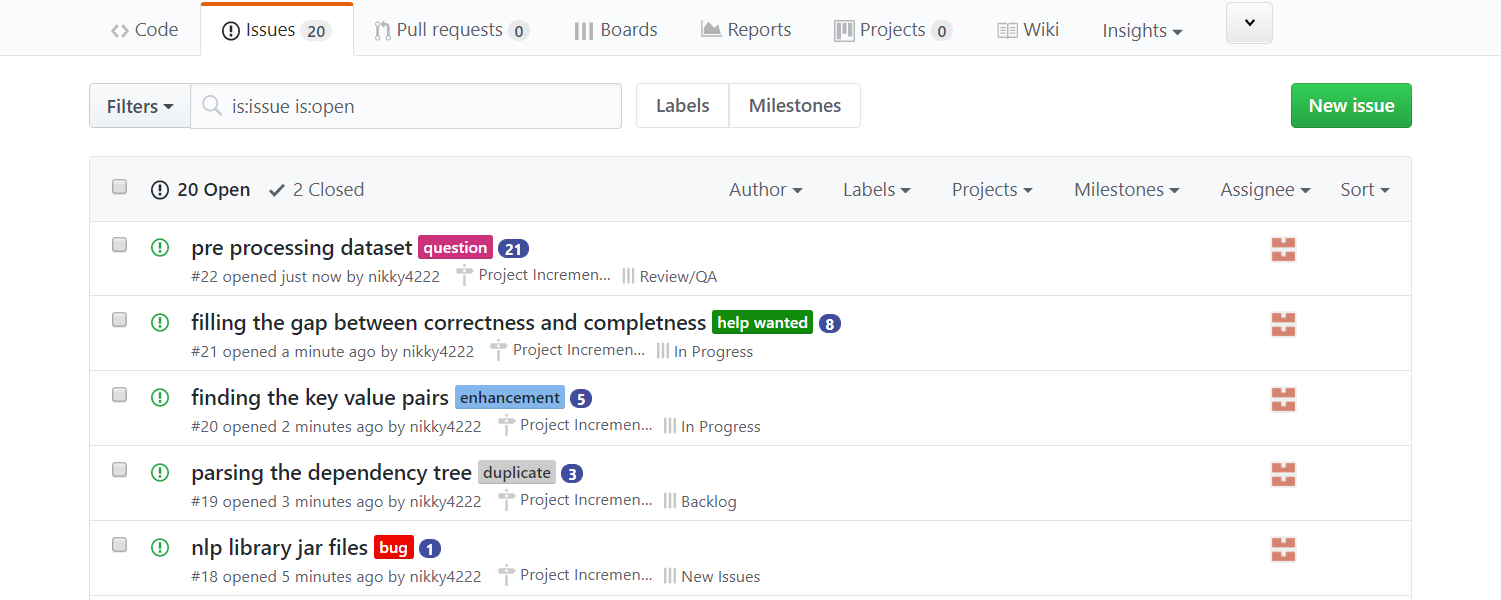


# **6.Project Management**

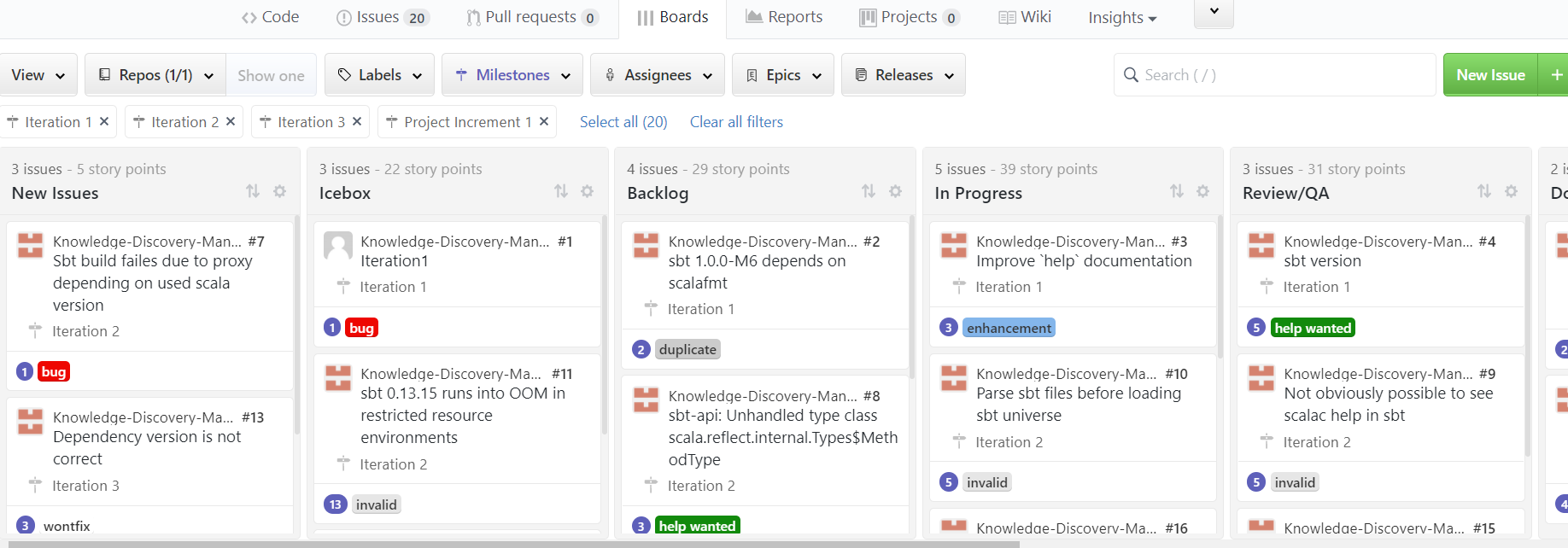
## **6.1 Contribution**

Prudhvi:25%||Sadanand:25%||Nikitha:25%||Harsha:25%

## **6.2 Issues**



## **6.3 Board**



## **6.3 Burndown Chart**



## **6.4 Future Works**

We would like to implement our existing question answering system to various sectors like medicine and healthcare, education, entertainment, also we would like to use the machine learning to enrich the ontology so that our QA system becomes rich.