

CS5560 Knowledge Discovery  
and Management  
Project Team: 8  
Increment-1 Report

Team Members:

Megha Nagabhushan

Rohithkumar Nagulapati

## Motivation

Geospatial data is all around us. geospatial data is any data with a spatial identifier referring to a position on the earth: a house, building, road, lake, mountain, or countless others. Geospatial data is also highly influential in today's business market, and businesses that incorporate geospatial data into their analysis, reporting, and forecasting have the potential to outpace competitors through smarter use of their data.

We can use this geospatial data to train our model and develop a question answering system.

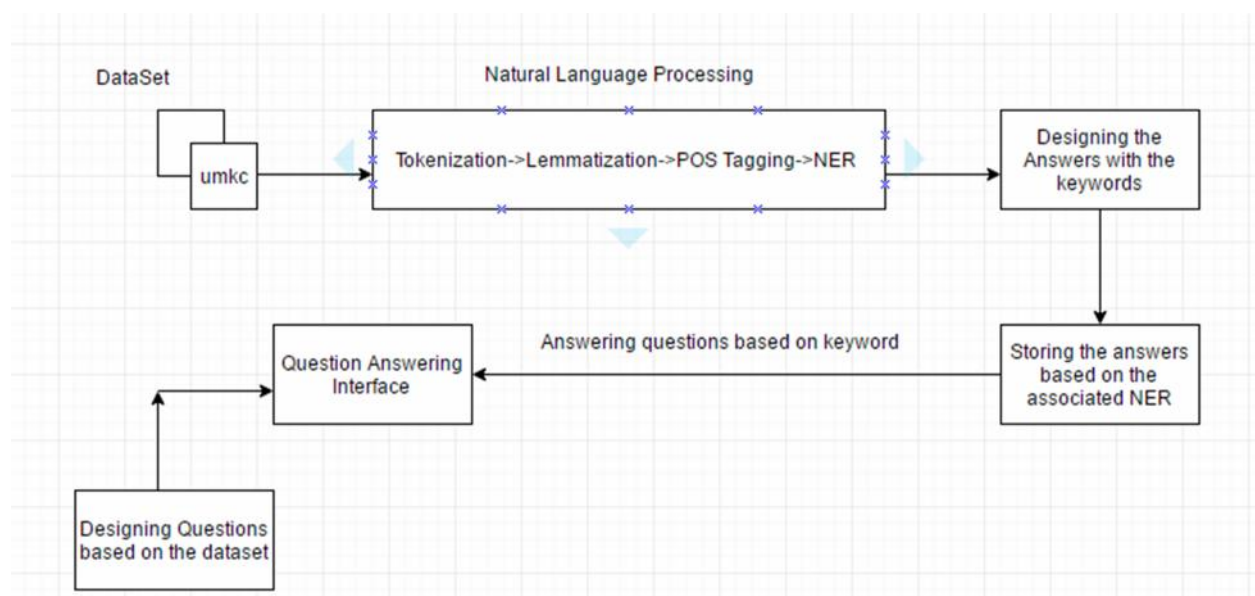
## Objectives

The objective of this project is to develop an intelligent question answering system for the geospatial data collected in the UMKC campus.

## Datasets

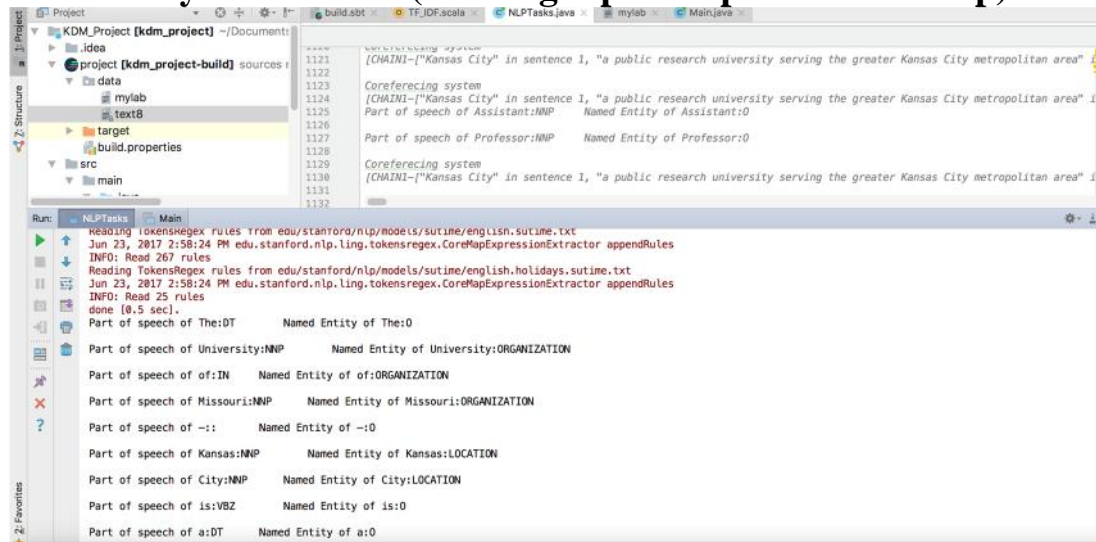
The dataset we will be using for the future increments will have 3D images taken inside UMKC campus. For this increment, we are using our own text dataset which contains information about UMKC faculty.

## Workflow



# Implementation

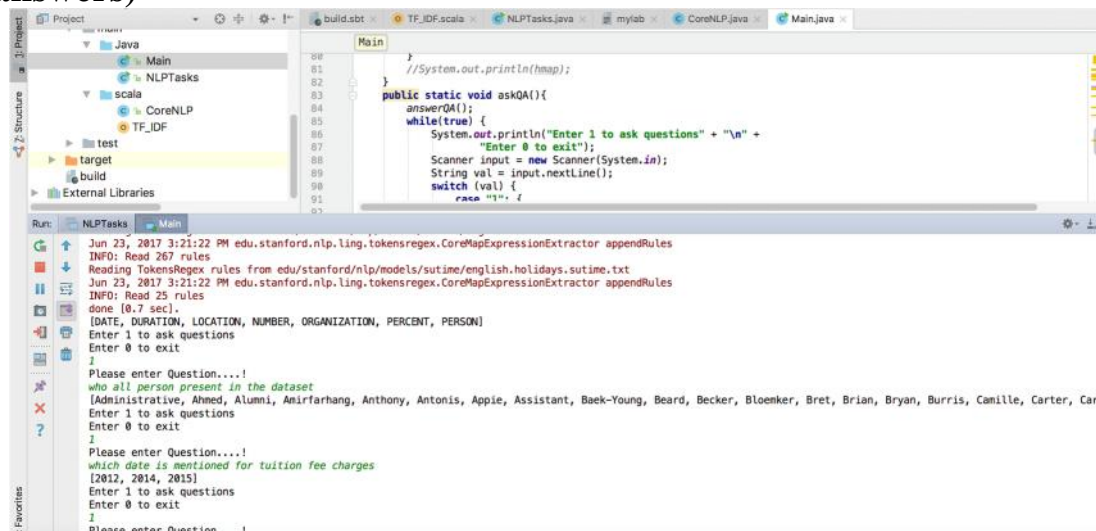
## a. NLP with your datasets (showing input/output for each step)



```
Project [kdm_project] ~\Documents\...
├── .idea
├── project [kdm_project-build] sources
│   ├── data
│   │   ├── mylab
│   │   ├── text8
│   │   ├── target
│   │   ├── build.properties
│   │   ├── src
│   │   └── main
│   └── ...
└── ...

Run: NLPTasks Main
Reading tokensRegex rules from edu.stanford.nlp.models.sutime/english.sutime.txt
Jun 23, 2017 2:58:24 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 23, 2017 2:58:24 PM edu.stanford.nlp.ling.tokensRegex.CoreMapExpressionExtractor appendRules
INFO: Read 25 rules
done (0.5 sec).
Part of speech of The:DT Named Entity of The:0
Part of speech of University:NNP Named Entity of University:ORGANIZATION
Part of speech of of:IN Named Entity of of:ORGANIZATION
Part of speech of Missouri:NNP Named Entity of Missouri:ORGANIZATION
Part of speech of -: Named Entity of -:0
Part of speech of Kansas:NNP Named Entity of Kansas:LOCATION
Part of speech of City:NNP Named Entity of City:LOCATION
Part of speech of is:VBZ Named Entity of is:0
Part of speech of a:DT Named Entity of a:0
```

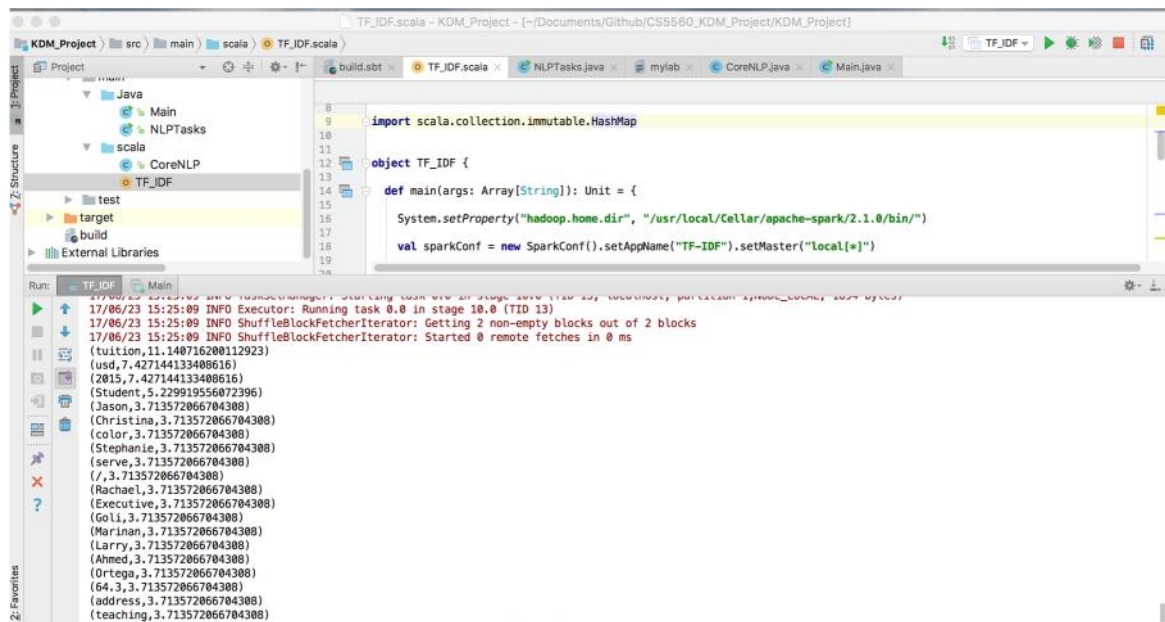
## b. Question Answering with your datasets (showing the list of questions/answers)



```
Project [kdm_project] ~\Documents\...
├── Java
│   ├── Main
│   ├── NLPTasks
│   └── scala
│       ├── CoreNLP
│       └── TF_IDF
├── test
├── target
├── External Libraries
├── build
└── ...

Run: NLPTasks Main
Jun 23, 2017 3:21:22 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 23, 2017 3:21:22 PM edu.stanford.nlp.ling.tokensRegex.CoreMapExpressionExtractor appendRules
INFO: Read 25 rules
done (0.7 sec).
[DATE, DURATION, LOCATION, NUMBER, ORGANIZATION, PERCENT, PERSON]
Enter 1 to ask questions
Enter 0 to exit
1
Please enter Question....!
who all person present in the dataset
[Administrative, Ahmed, Alumni, Amirfarhang, Anthony, Antonis, Appie, Assistant, Baek-Young, Beard, Becker, Bloenker, Bret, Brian, Bryan, Burris, Camille, Carter, Caru]
Enter 1 to ask questions
Enter 0 to exit
1
Please enter Question....!
which date is mentioned for tuition fee charges
[2012, 2014, 2015]
Enter 1 to ask questions
Enter 0 to exit
1
Please enter Question....!
```

## b. TFIDF with your datasets (showing input/output)



## Project Management

### a. Contribution of each member

Megha Nagabhushan – 50%

RohitKumar Nagulapati – 50%

### b. Include ZenHub and GitHub URL/statistics/screens

[https://github.com/ROHITHKUMARN/CS5560\\_KDM\\_Project](https://github.com/ROHITHKUMARN/CS5560_KDM_Project)

### c. Future Work

For the future work, we will be using rest services to predict the image data and then use the predictions to develop knowledge graph. We will use Google's knowledge graph API for generating the knowledge graph.