BOOTCAMP ASSIGNMENT 1 | MAKRAND BHONDE

Problem: Grammar Based triplet extraction from text & knowledge graph creation with visualization. Build a flask-based web application to create a knowledge graph with unstructured data.

Approach:

1. Reading input text with flask web app based UI.
2. Divide the text into tokens and get part of speech (POS) tags for each token.
3. Based on the POS tags, write regular expression-based rules to extract only Nouns.
4. Based on POS tags write regular expression-based rules to extract relation between two nouns if present in a sentence.
5. Create triplets with above rules – (Noun, Relation, Noun)
6. Exploring the Knowledge Graph libraries and build a KG with the above triplets.

Requirements:

flask
re
pandas
bs4
requests
spacy
networkx
matplotlib

Above all are the requirements for the project to get executed. Installing python is the first need to get started. We can save above requirements in a text document named 'requirements.txt' and then just we have to hit
"pip install requirements.txt"

Pandas: Performs data manipulation and analysis. Data frames and tables can be created.
Spacy: It is used to perform NLP (Natural Language Processing)
Networkx: It is used to study graphs and build networks.
Matplotlib: Plotting and visualizing graphs.
Flask: Lightweight python web-framework

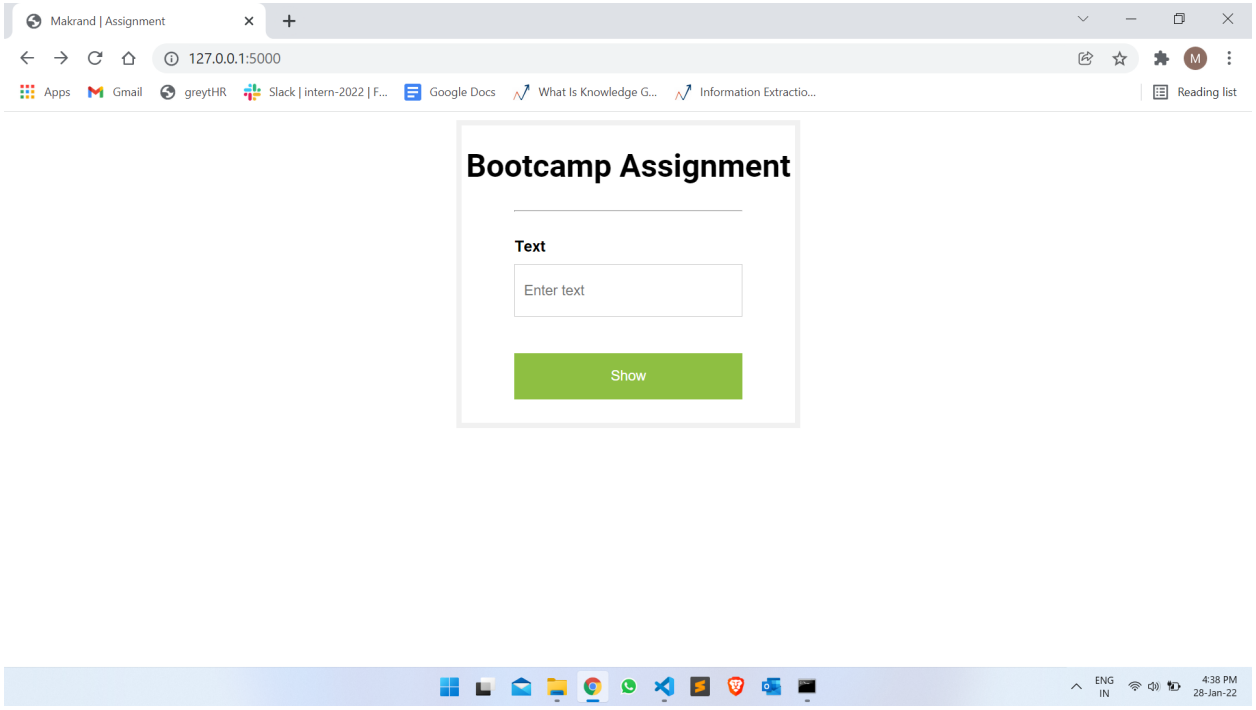
My Methodology and steps involved:

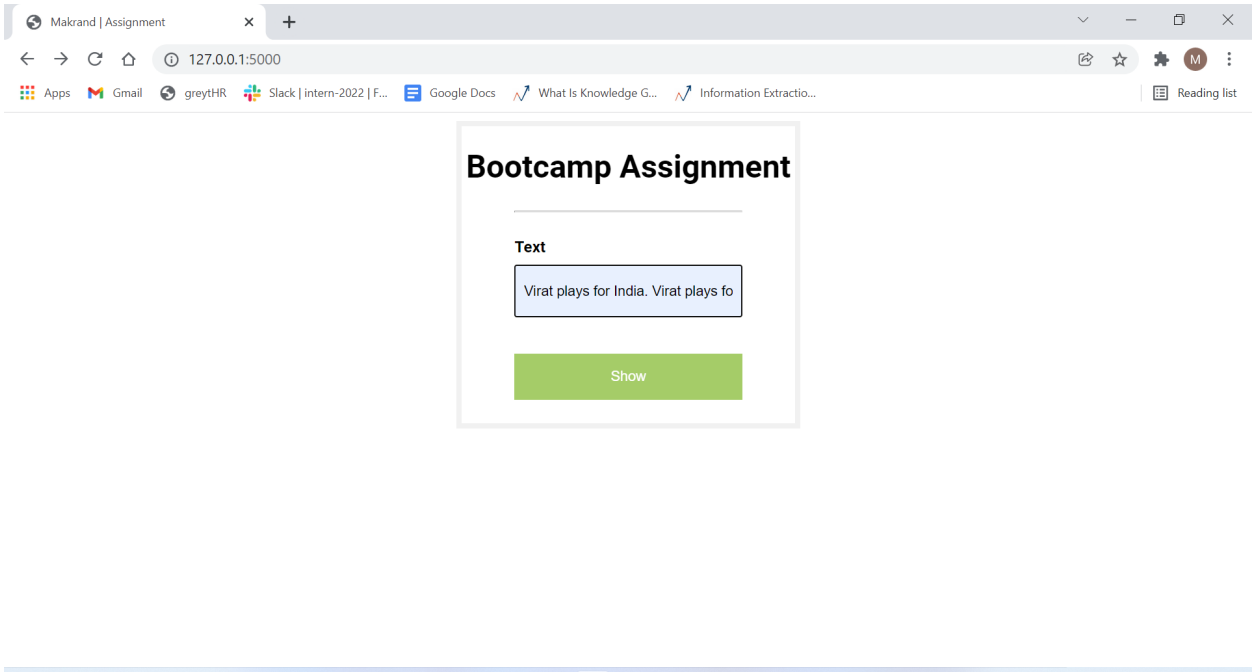
- Using flask, first I created a simple form which contains a text field and a show button.
- This text field takes input and after clicking the show button it passes this string input to a function in flask.
- This function first splits the sentences on the 'full stop'(.) and makes a list containing all the sentences.
- These sentences are appended to a csv containing a column sentence and out of which we make data frames.
- Using NLP techniques with help of spacy we define patterns on which we divide sentences into subject, object and relation between them

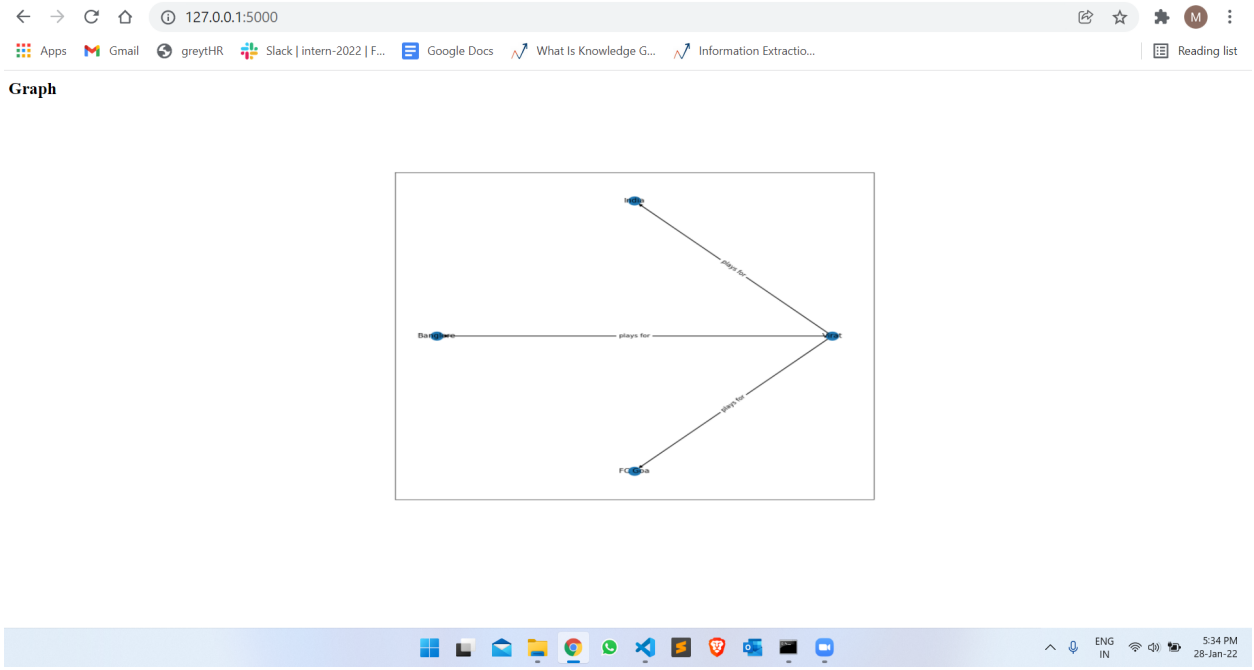
- Three of them above are stored separately treating them as separated entities and then using networkx library we create a network out of it.
- Using matplotlib we draw this network creating a directed graph which has subject and object as nodes and edge, labeled as relation between them if they exist.
- Thus, the output graph is shown by rendering a html page containing a graph figure.

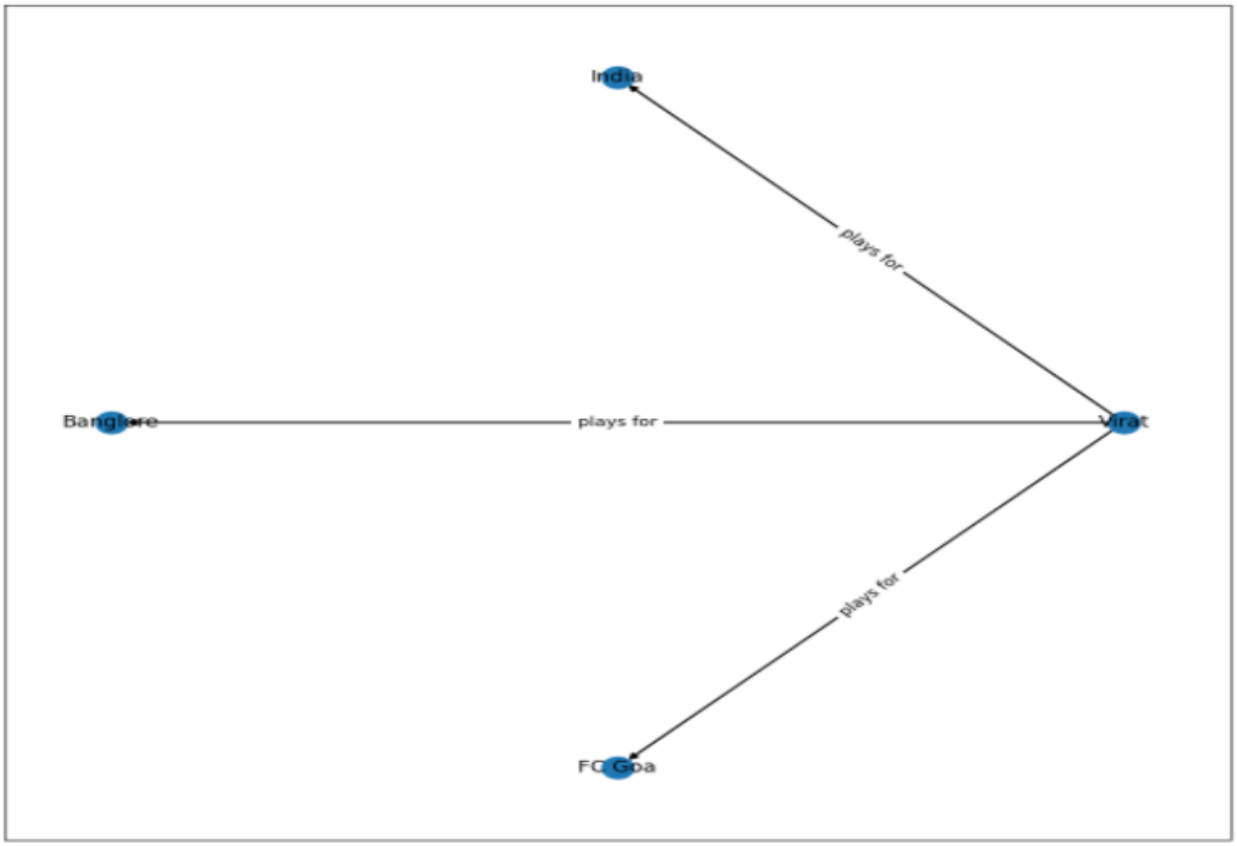
Input string:
“Virat plays for India. Virat plays for Bangalore. Virat plays for FC Goa.”

Output snaps:




Makrand | Assignment
127.0.0.1:5000
Apps
Gmail
greytHR
Slack | intern-2022 | F...
Google Docs
What Is Knowledge G...
Information Extractio...
Reading list
Bootcamp Assignment
Text
Virat plays for India. Virat plays fo
Show
ENG
IN
4:39 PM
28-Jan-22


127.0.0.1:5000
Apps
Gmail
greytHR
Slack | intern-2022 | F...
Google Docs
What Is Knowledge G...
Information Extractio...
Reading list
Graph
plays for
plays for
plays for
ENG
IN
5:34 PM
28-Jan-22



References:

https://www.analyticsvidhya.com/blog/2019/09/introduction-information-extraction-python-spacy/
https://networkx.org/documentation/stable/tutorial.html
https://flask.palletsprojects.com/en/2.0.x/quickstart/
https://docs.python.org/3/library/csv.html
https://stackoverflow.com/