**How to run WikiScrap.py to scrap data from Wikipedia.**

* A movie metadata was downloaded from imdb which is present in the data folder. This data set was used to get the name of the movies to extract the information of the movies from Wikipedia.
* Place “movies\_metadata.csv” in the working directory if you want to extract the data again.
* Install scrapy in python: run command ‘pip install Scrapy’
* Run the python program WikiScrap.py
* This will dump all the collected data in the form of JSON in movies.json
* The collected data is available in dataset folder.

**Creation of Knowledge Graph (KG.ipnyb):**

* Place the dataset movies.json and movie.json in the working directory. Movie.json has just a single movie for data analysis.
* Run command ‘pip install spacy’ to install spacy
* Run command ‘pip install displacy’ to install displacy .
* Install Neo4j database and create a new graph database. Start the database server
* Change the password “adbi” to your password in the KG.inpyb file
* Install the libraries required to connect neo4j from python
* Run command : ‘pip install py2neo’
* Execute all cells of KG.ipnyb to create the knowledge graph in the neo4j database.
* Run the following command in the neo4j db to check the knowledge graph

Match (n) return n limit 100;

The limit will limit the number of records as the KG will be huge. The query returns any 100 nodes from the database.

**Executing API.py:**

* Install flask by running command: ‘pip install Flask’
* Install APOC and Graph algorithms in neo4j to enable random walk.
* Run api.py

Query the database using the URL in the following format:

<http://localhost:5005/search?person=keyword>

<http://localhost:5005/search?movie=keyword>

<http://localhost:5005/search?date=keyword>

Replace the keyword with your own keywords to get the results on the application.