Natural Language Processing

Project Review 2

Analysis of Knowledge Graphs and WordNET for Wikipedia based Question Answering System

Team Members

Kriti Gupta 17BCE1327

Amrit Gupta 17BCE1082

Introduction

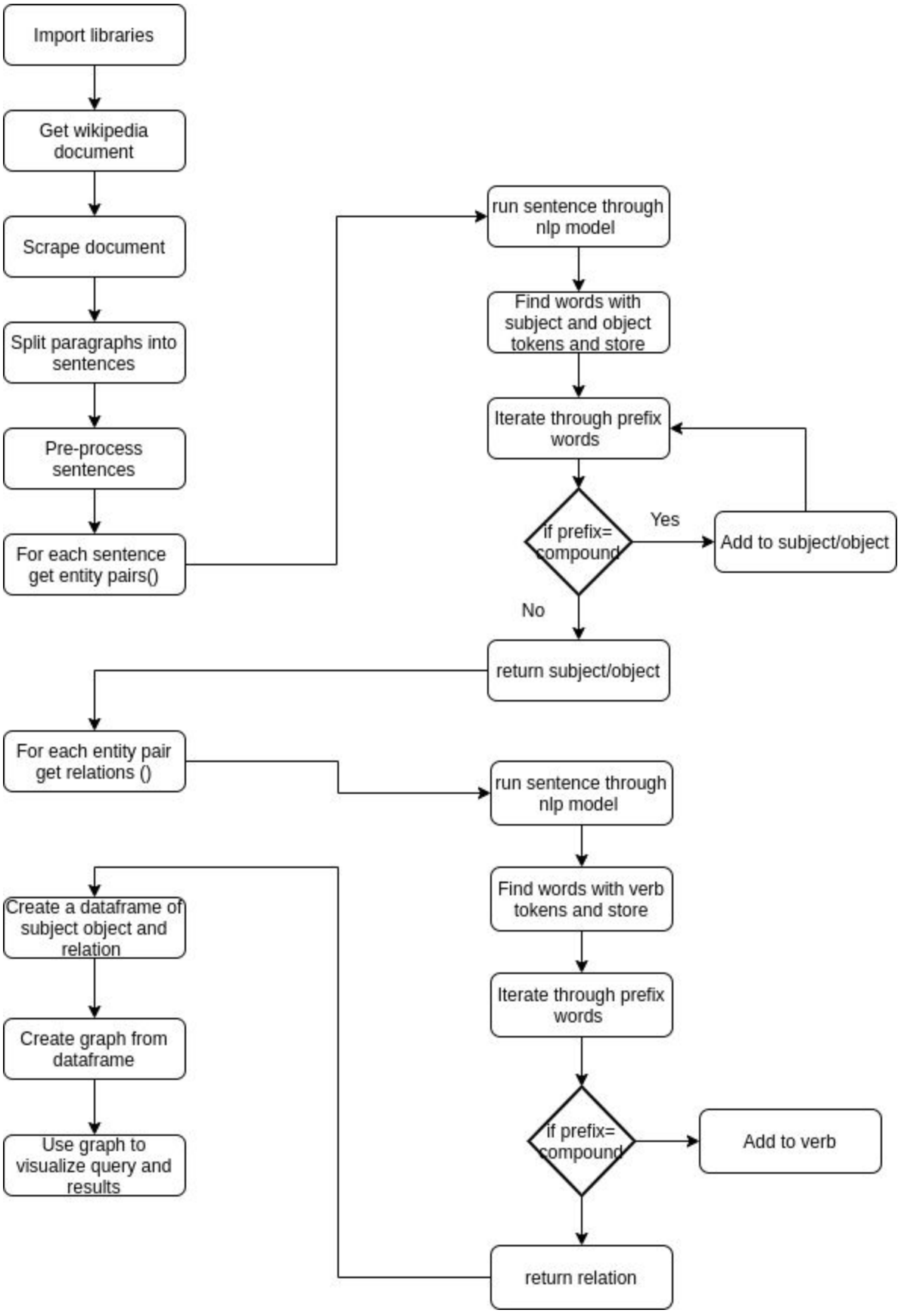
A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge.In other words, a knowledge graph is a programmatic way to model a knowledge domain with the help of subject-matter experts, data interlinking, and machine learning algorithms.

In this project we aim to build a system that scrapes wikipedia documents depending on a search query and builds a knowledge graph of its contents. It then uses the structure of the sentence of the query to find the entity pair(subject,object) and relation from the query sentence and match it to the knowledge graph. The final result would be a visualization of the relations found in the document that match the given query, thus creating an effective question answering system.

This system is very useful as in a lot of situations small queries such as “How did Micheal Jackson die?” take a long time to solve manually as wikipedia offers a hist of knowledge that may or may not be relevant. This system will help solve this problem.

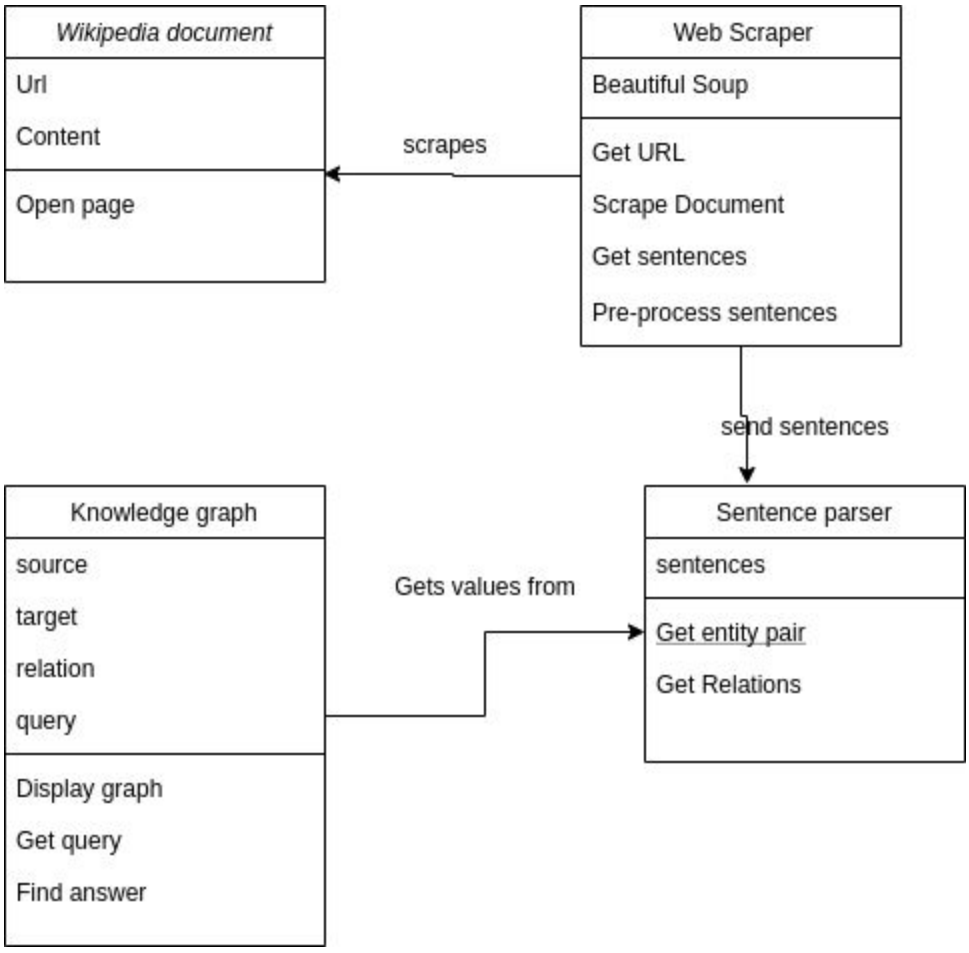
Low Level UML Diagram

Flowchart

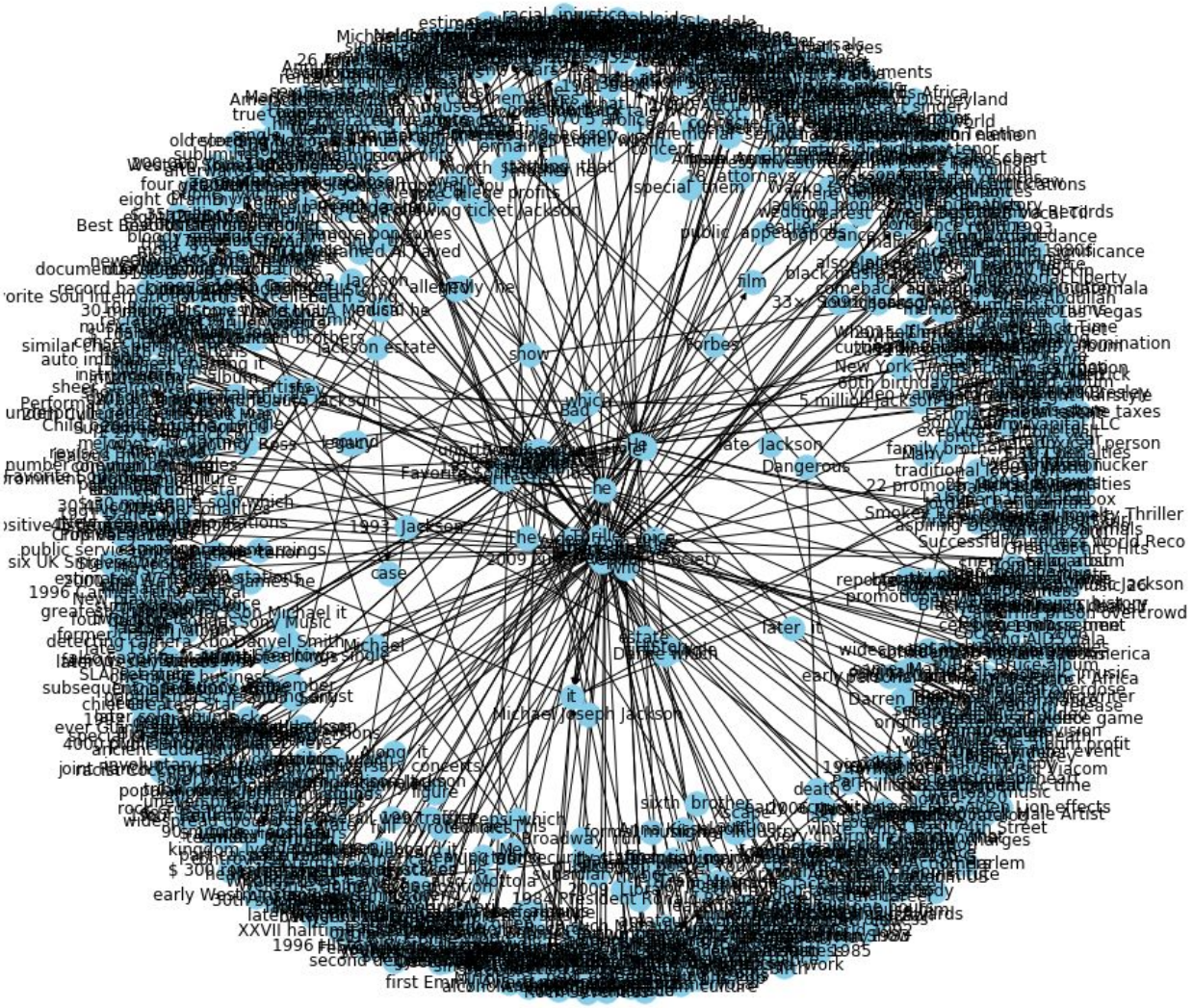

Import libraries
Get wikipedia document
Scrape document
Split paragraphs into sentences
Pre-process sentences
For each sentence get entity pairs()
run sentence through nlp model
Find words with subject and object tokens and store
Iterate through prefix words
if prefix= compound
Yes
Add to subject/object
No
return subject/object
For each entity pair get relations ()
run sentence through nlp model
Find words with verb tokens and store
Iterate through prefix words
if prefix= compound
Add to verb
return relation
Create a dataframe of subject object and relation
Create graph from dataframe
Use graph to visualize query and results

High Level UML Diagram

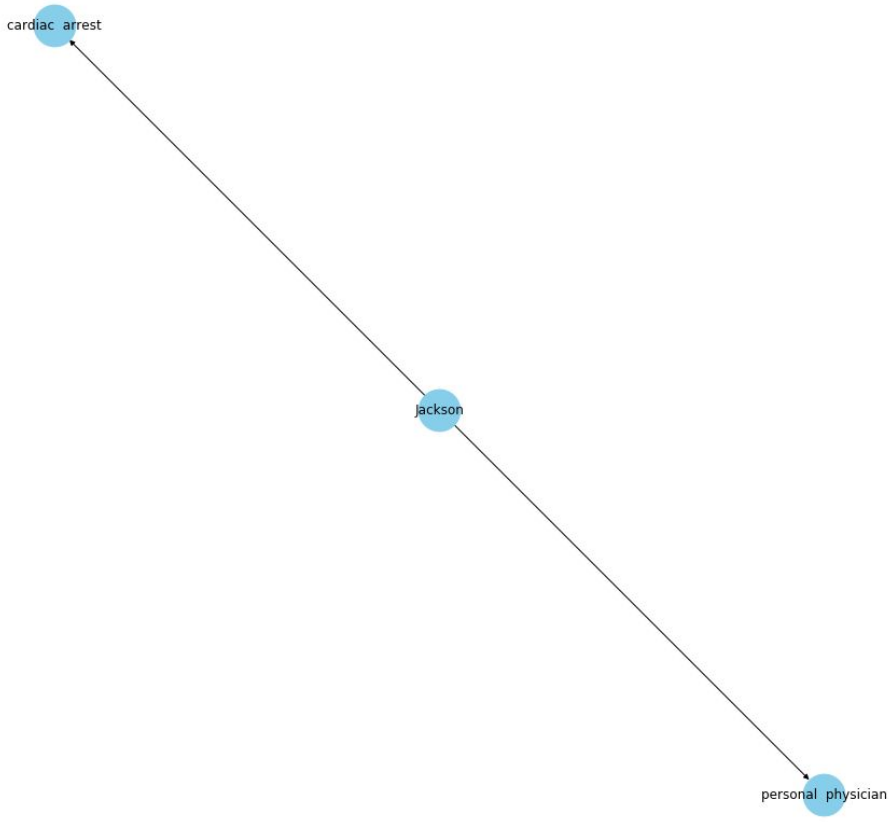
Class Diagram



Output Graph



For query: “died from” for Micheal Jackson page



Future Work

- Adding WordNet to identify synonymous or hyponymous questions and relations
- Adding automatically search functionality to make the application end-to-end
- Expand the graph from wikipedia to other infotainment webpages like: ask.com, and other article sites such as TOI, Buzzfeed etc.