Knowledge Graph

By

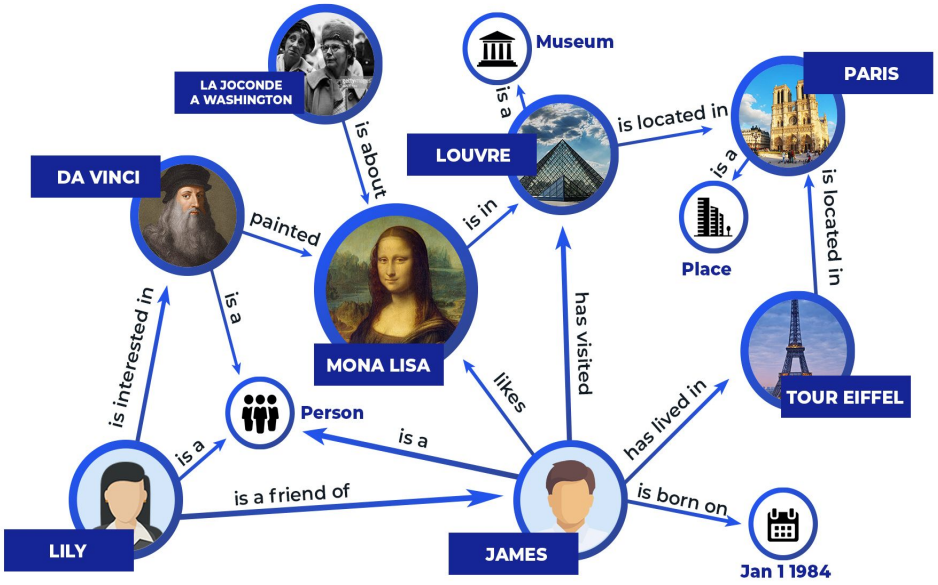
Vibhav Nirmal (160280107061)
Stuti Sharma (160280107108)
Vedant Vachharajani (160280107118)

Outline

- What is knowledge graph
- What we tried
- BiDAF
- Knowledge Graph Representation
- Examples and graph
- The major differences observed
- Technology used
- Limitations
- Future Scope

What is Knowledge Graph and Why is it used?

- A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge
- Here, you can easily insert the relations and traverse the links to discover the relations between data
- A graph is one of the the most flexible formal data structures, so you can easily map other data formats to graphs using generic tools



What we have tried to achieve

Our **objectives** of the project can be outlined as:

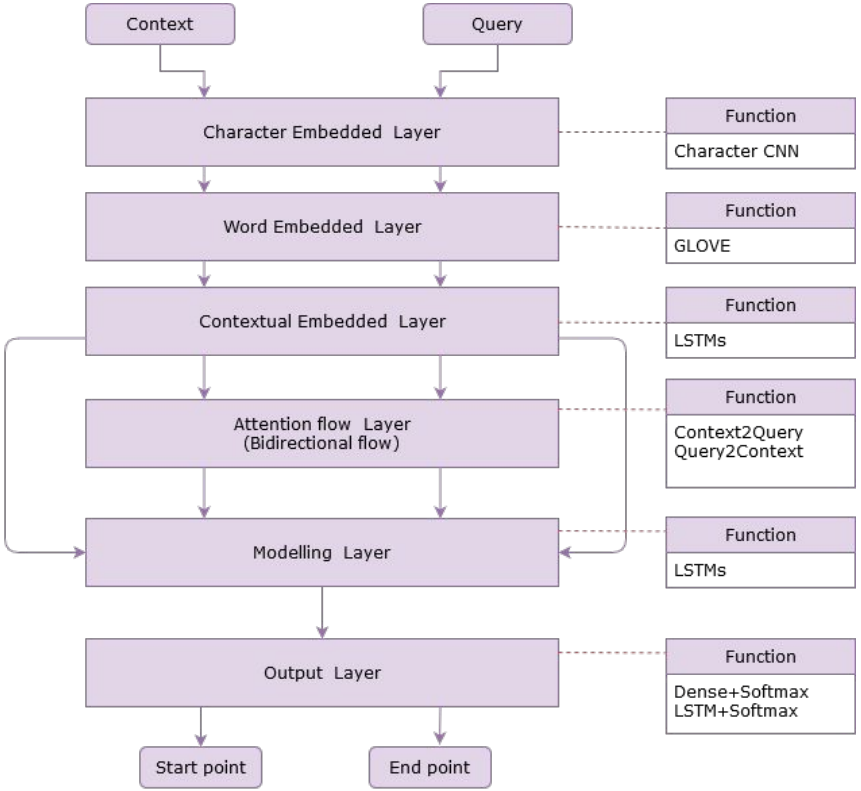
1. To implement different, novel approaches to intelligent question answering system
2. To evaluate advantages and disadvantages of each studied method
3. To study each approach in order to generate the best cases under which one method outperforms another
4. To solve queries and answer questions efficiently from a given context

The first approach is the **Bidirectional Attention Flow model (BiDAF)** developed by AllenAI

The other approach is **Knowledge Graph Representation.**

Bidirectional Attention Flow

- Bidirectional Attention Flow or BiDAF is a Machine Comprehension model
- Structured over a 6-layer architecture
- Modelling complex interactions between context and query was its main concept
- The model is trained over Stanford Question Answering Dataset (SQuAD)



Bi-directional Attention Flow Demo

for Stanford Question Answering Dataset (SQuAD)

Direction : Select a paragraph and write your own question. The answer is always a subphrase of the paragraph - remember it when you ask a question!

Select Paragraph

[05] Teacher

Paragraph

The role of teacher is often formal and ongoing, carried out at a school or other place of formal education. In many countries, a person who wishes to become a teacher must first obtain specified professional qualifications or credentials from a university or college. These professional qualifications may include the study of pedagogy, the science of teaching. Teachers, like other professionals, may have to continue their education after they qualify, a process known as continuing professional development. Teachers may use a lesson plan to facilitate student learning, providing a course of study which is called the curriculum.

Question

Where is a teacher most likely to be teaching at?

new question!

Answer

a school or other place of formal education

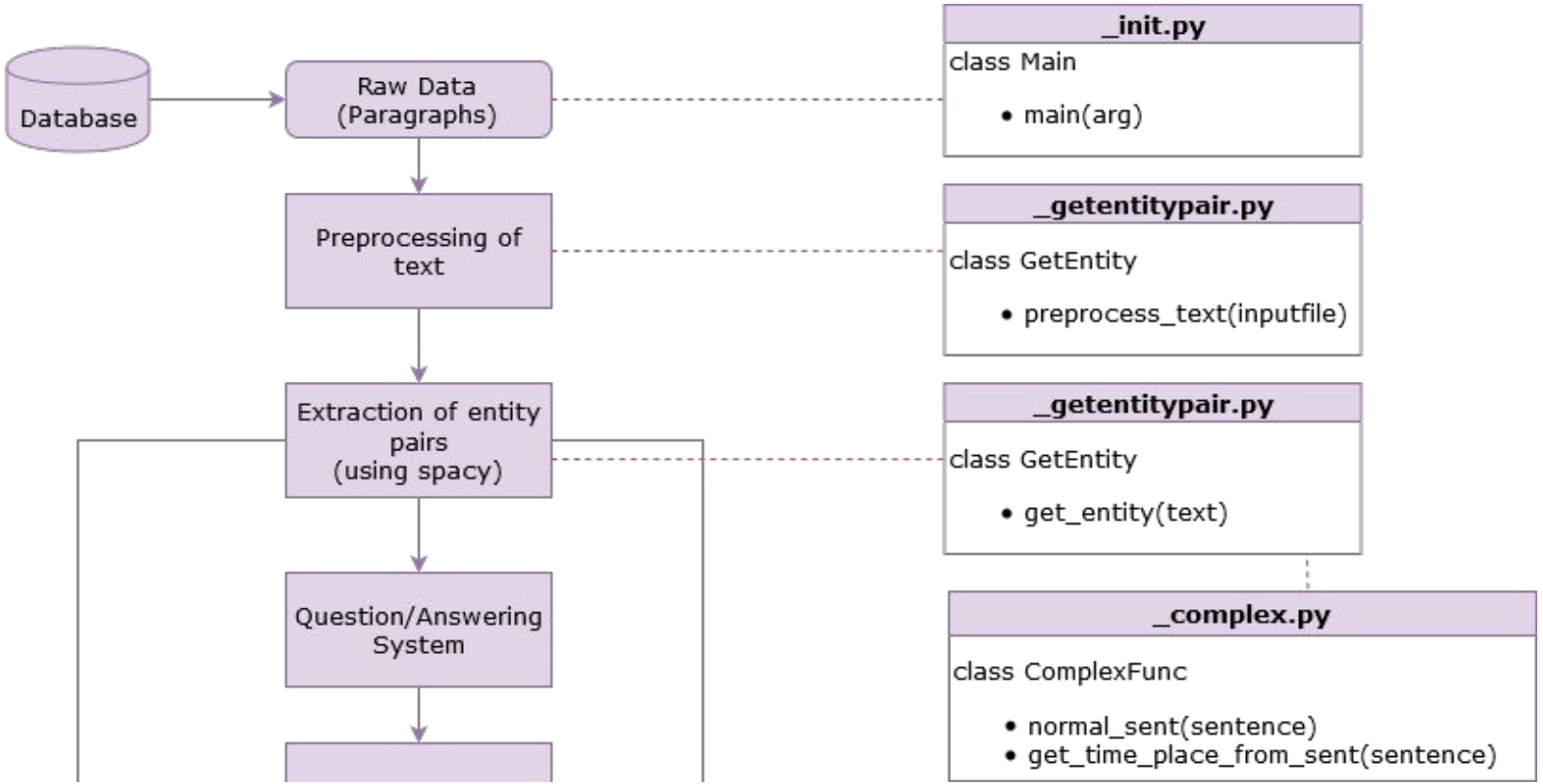
Reference : Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, Hannaneh Hajishirzi. "Bidirectional Attention Flow for Machine Comprehension" [link]

Demo by : Sewon Min

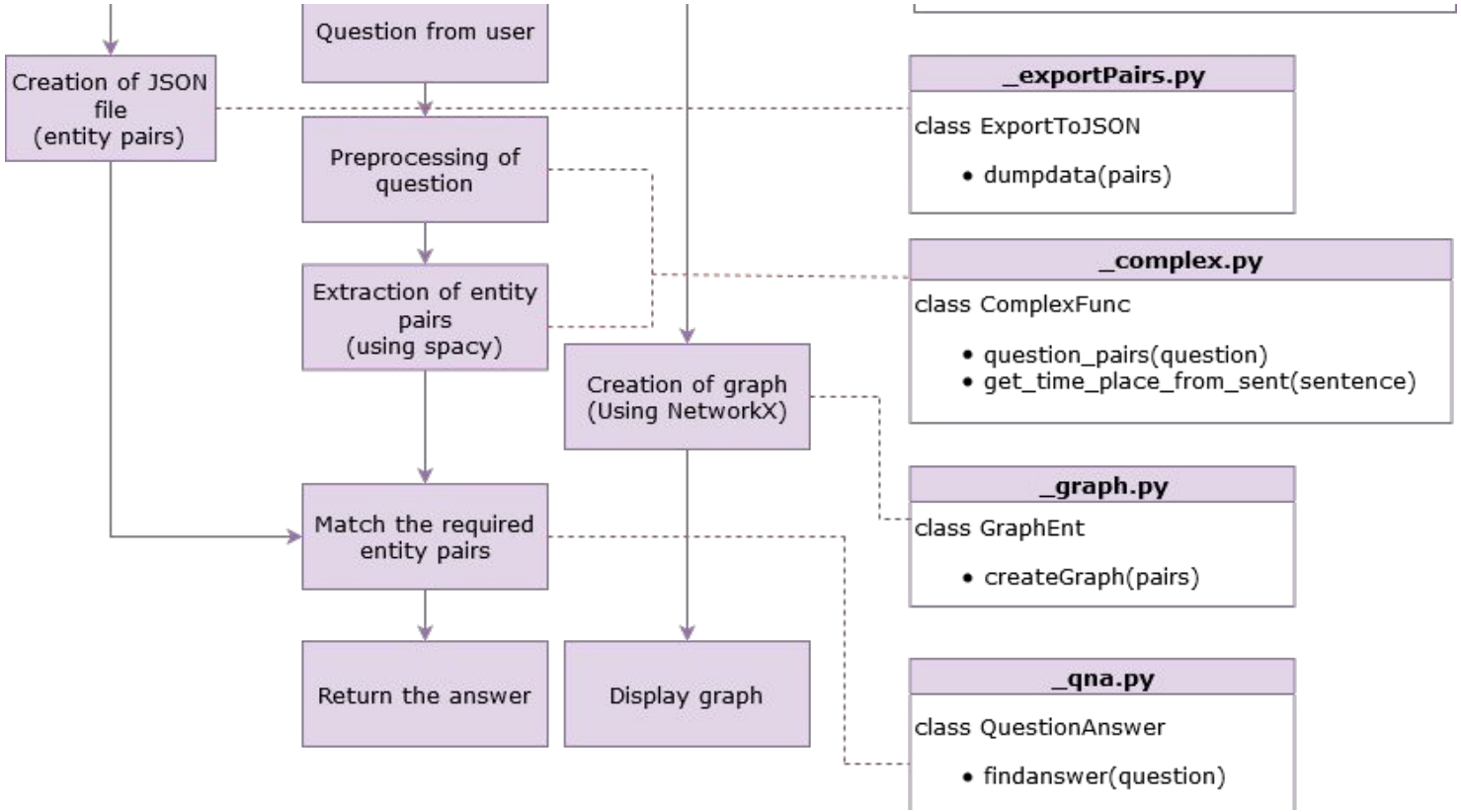
Knowledge Graph Representation

- Involves extraction of **entities** in the form of subject, predicate and object
- Uses **dependency parsing** tools and methods
- Entities are used to **extract answers** from the questions asked by user
- The approach works with **questions** in the form of ‘who’, ‘what’, ‘when’ and ‘where’
- To extract answers, we **match common entities** from questions and the context
- As of now, the answers obtained are **simple** and in limited number of words

Workflow

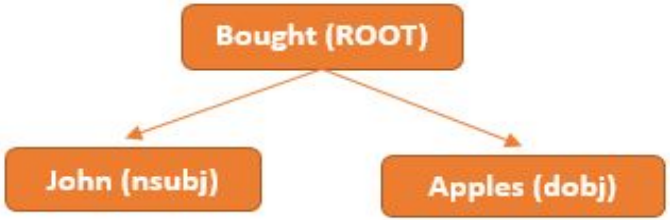

Database
Raw Data (Paragraphs)
Preprocessing of text
Extraction of entity pairs (using spacy)
Question/Answering System
_init.py
class Main
• main(arg)
_getentitypair.py
class GetEntity
• preprocess_text(inputfile)
_getentitypair.py
class GetEntity
• get_entity(text)
_complex.py
class ComplexFunc
• normal_sent(sentence)
• get_time_place_from_sent(sentence)

Workflow Cont.

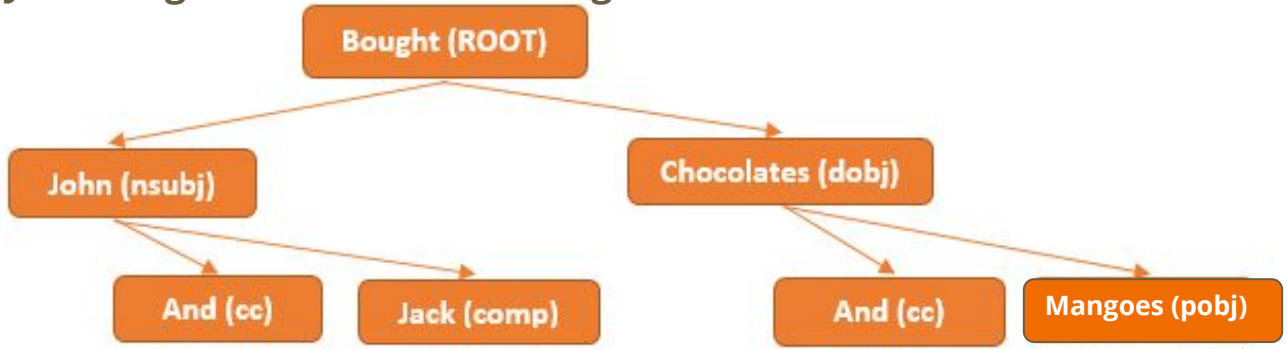


Dependency Parsing of Context

John bought apples.

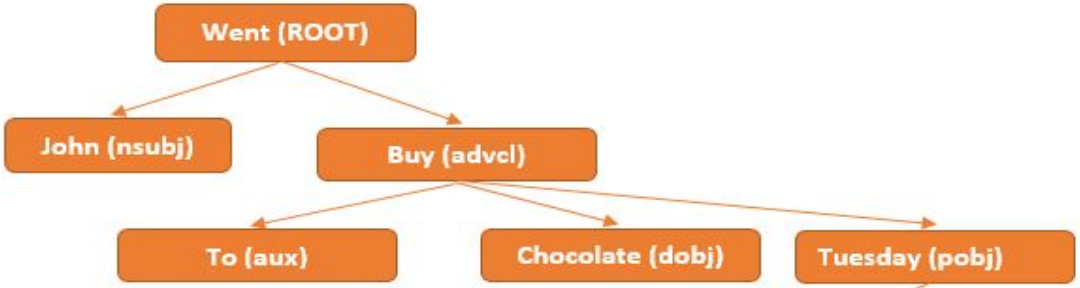


John and Jack bought chocolate and mangoes.

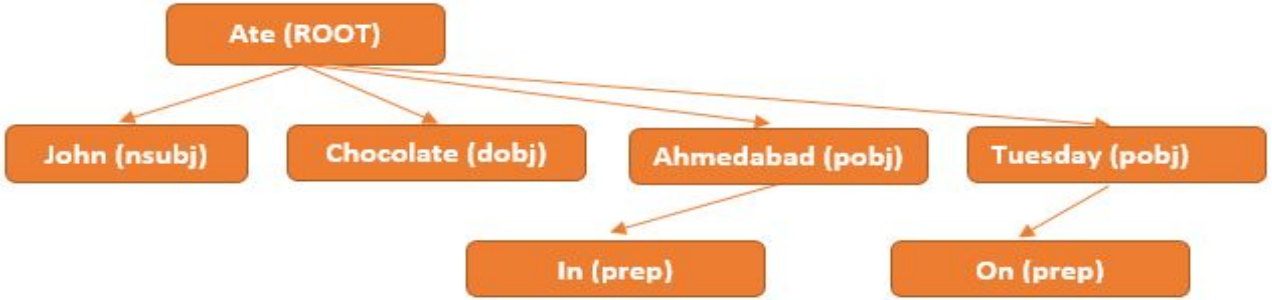


Dependency Parsing of Context

John went to buy chocolate on Tuesday.



John ate chocolate in Ahmedabad on Tuesday.



Example for extraction of entities

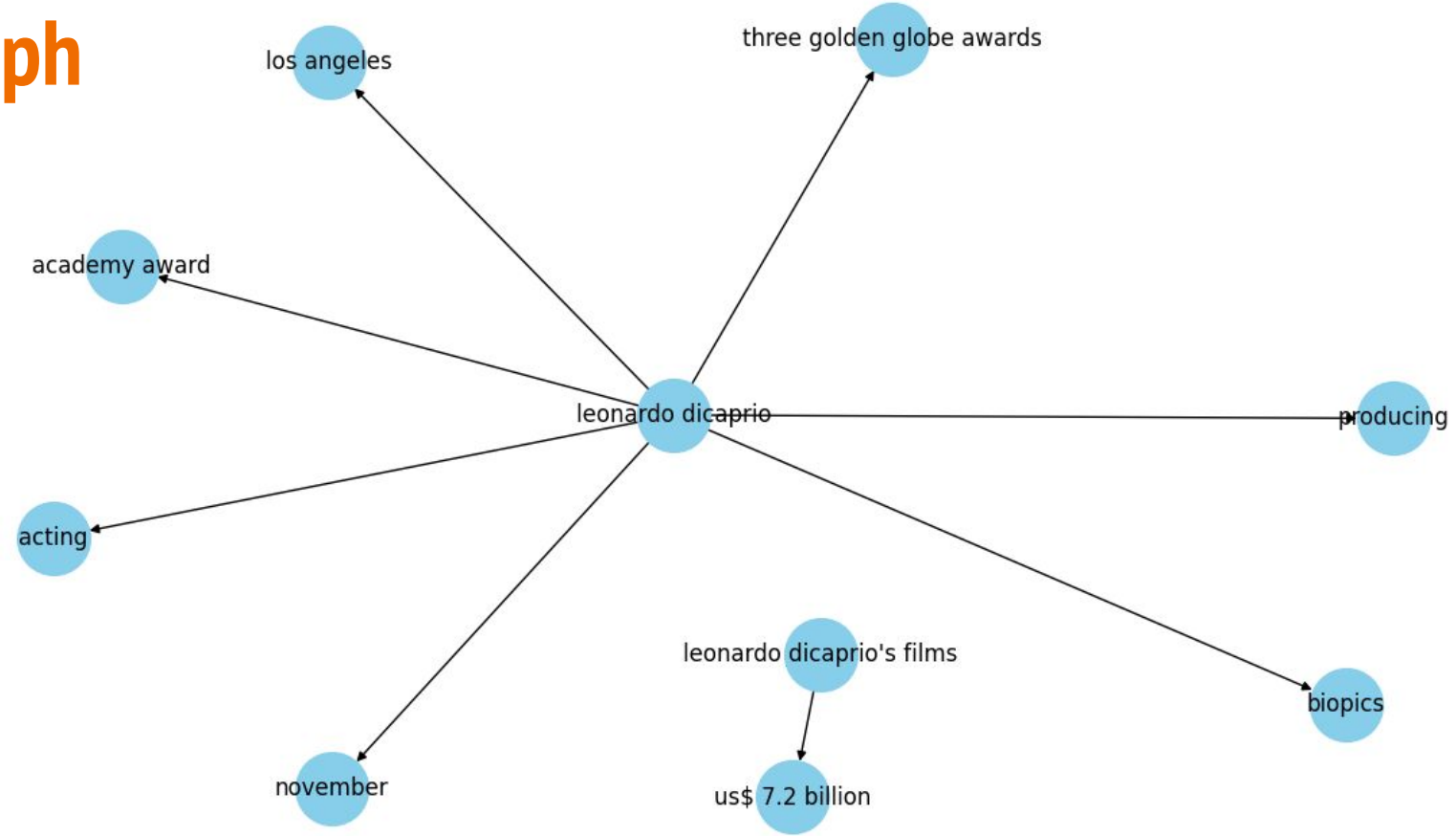
- The Context:

  "Leonardo DiCaprio was born on November 11, 1974. He can do acting and producing. He has acted in biopics. His films have earned US$7.2 billion. He has won an Academy Award and three Golden Globe Awards. He was born in Los Angeles."

- We extract following entities in pandas **dataframe** and save it in **JSON** database

	source	relation	aux_relation	target	time	place
0	leonardo dicaprio	born		november	November 11, 1974	
1	leonardo dicaprio	do		acting		
2	leonardo dicaprio	do		producing		
3	leonardo dicaprio	acted		biopics		
4	leonardo dicaprio's films	earned		us$ 7.2 billion		
5	leonardo dicaprio	won		academy award		
6	leonardo dicaprio	won		three golden globe awards		
7	leonardo dicaprio	born		los angeles		Los Angeles

Graph


los angeles
three golden globe awards
academy award
leonardo dicaprio
producing
acting
leonardo dicaprio's films
biopics
november
us$ 7.2 billion

Pseudo Code for entity extraction from context

- In the given paragraph, **preprocess** the text to remove/change the unwanted characters/symbols
- Now, in each word of a sentence:

    Check for **word_dep**:

        If Subject : extract and save it as variable subject

        If Root(Verb) : extract and save it as variable relation

        If Object : extract and save it as variable object

- Now save the obtained variables into a dataframe and dump the data into JSON file

This **dataframe** contains: (“SUBJECT”, ”RELATION”, ”AUX_RELATION”, ”OBJECT”, ”TIME”, ”PLACE”)

Pseudo Code for matching entity pairs as per ques

- In each question asked, **preprocess** the question to remove/change the unwanted characters/ symbols

- **If question contains "WHO"**
  Check for all entity pairs:
  If Relation(Question)==Relation(EntPair)
  and
  Object(Question)==Object(EntPair)
  Return (Subject from EntPair)

If question contains "WHERE"
Check for all entity pairs:
If Relation(Question)==Relation(EntPair) &
Subject(Question)==Subject(EntPair) &
Object(Question)==Object(EntPair)
Return (PLACE from EntPair)

- **If question contains "WHAT"**
  Check for the entity pairs:
  If Relation(Question)==Relation(EntPair)
  and
  Subject(Question)==Subject(EntPair)
  Return (Object from EntPair)

If question contains "WHEN"
Check for all entity pairs:
If Relation(Question)==Relation(EntPair) &
Subject(Question)==Subject(EntPair) &
Object(Question)==Object(EntPair)
Return (TIME from EntPair)

- Now save the obtained variables into a dataframe and dump the data into JSON file along with data_entities

Knowledge Graph representation with Question Answering

Project by Vibhav Nirmal, Vedant Vachharajani, Stuti Sharma

Select a paragraph and write your own question.
The answer is always a subphrase of the paragraph.

Paragraph Write Paragraph

Select or write your own paragraph

Question Write Question

Select or ask your own question

submit Clear

Answer

Our Model's Answer

Our Projected Answer

BIDAF Answer

BIDAF Model Answer

This project's main objective is to implement different and novel approaches to autonomous question answering systems. Here we are evaluating the advantages and disadvantages of the two different methods. We are studying each approach methodically in order to generate the best case results under which each scenario, that can be used to solve queries and answer the questions efficiently.

Reference : Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, Hannaneh Hajishirzi. "Bidirectional Attention Flow for Machine Comprehension" [link]

The Differences Observed

Aspect	BiDAF	Our Model
Scope	Paragraph with complex sentences with 1000 word limit	Considers relatively simpler sentences
Performance	Performs better on larger paragraphs with much complex sentences but fails to extract answers from simpler sentences	Performs perfectly for a limited type of sentences but fails to extract answers from complex paragraphs
Paragraph	Cannot work properly on multiple paragraphs with related context	Can work properly on multiple paragraphs with related or different context
Principle	Works on the principle of Bidirectional Attention mechanism	Works on the principle of dependency checking and entity pairs extraction
Answers	Extracts a portion from paragraph	Extracts answers from entities and hence gives multiple words which may be from different sentences and paragraphs
Scope of Questions	Scope of questions is limited to the information directly stated in the context without any inferences.	Scope of questions is limited to the amount of information confided in the sentences. Such as what, who, when and where

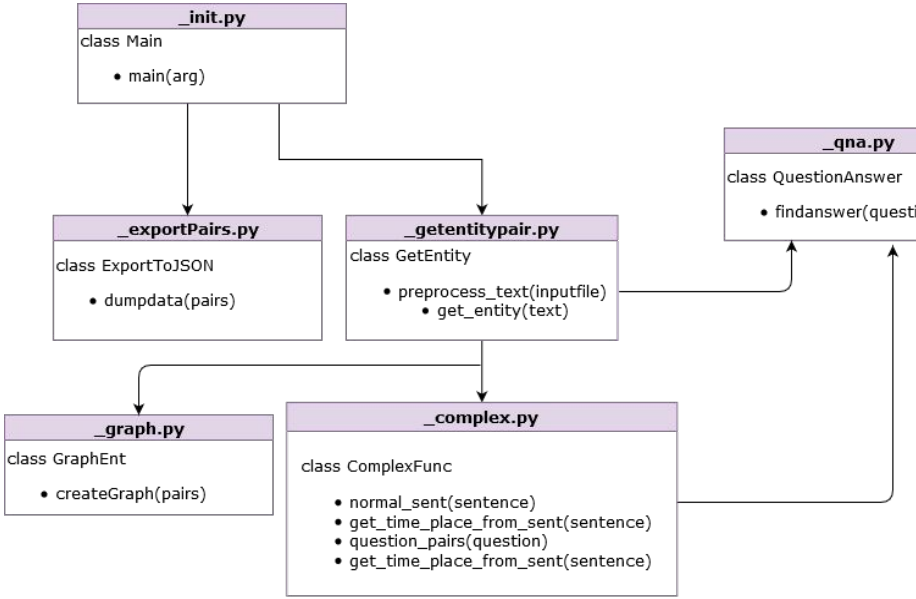
Technology Stack

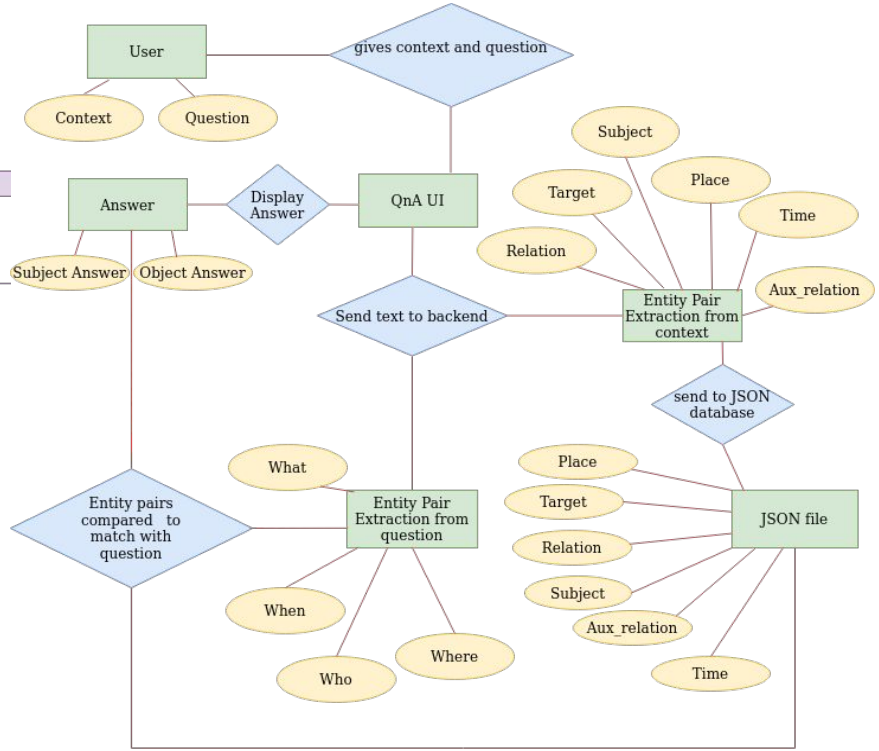
Python3.6 or Above
Tensorflow==1.1.0

Libraries Used:

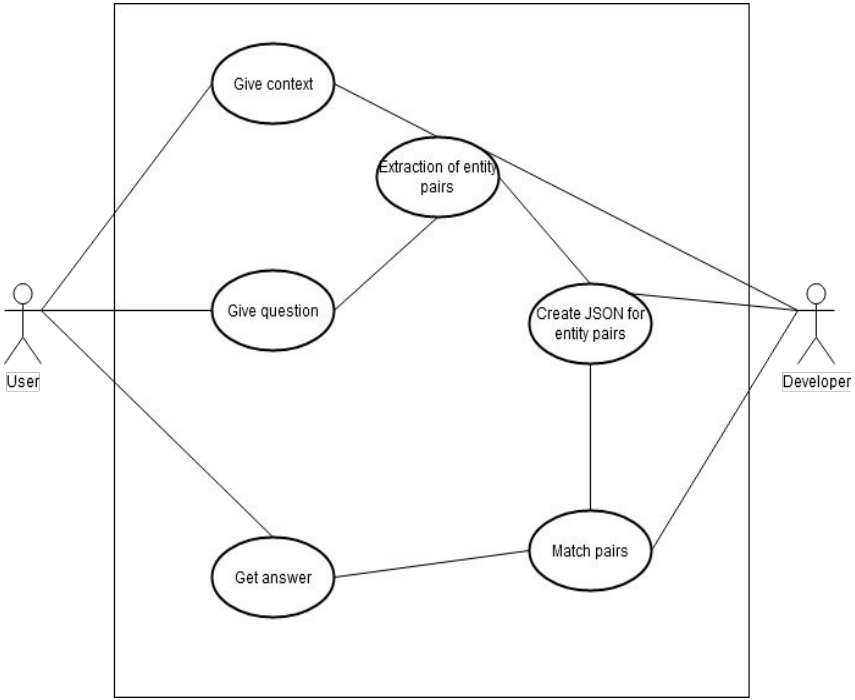
- Spacy
- Json
- Pandas
- Networkx
- Tqdm
- Jinja2
- Flask
- Inflect

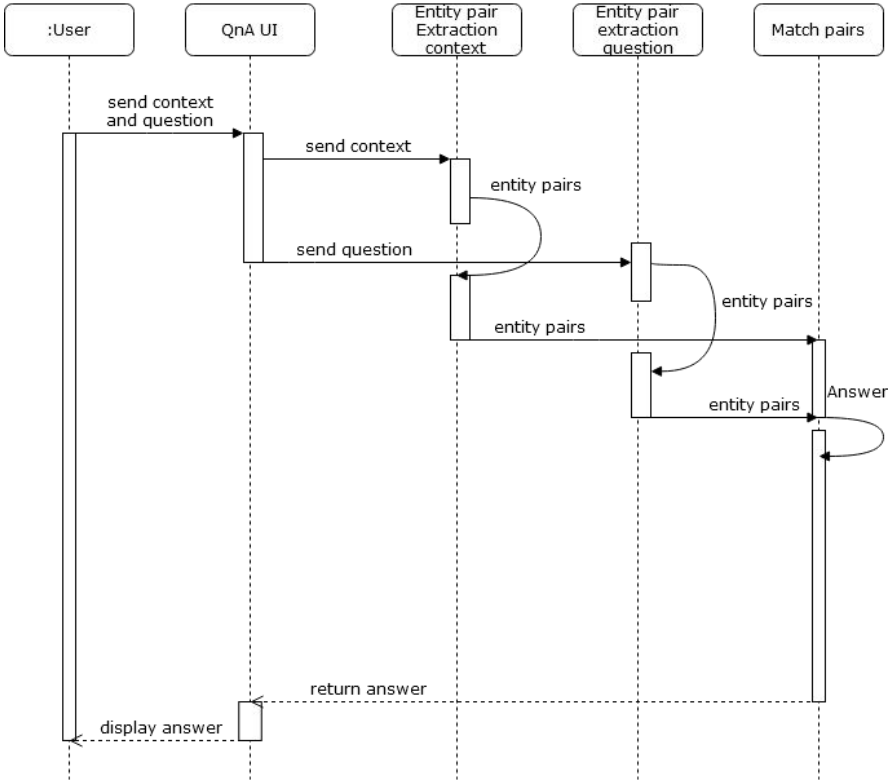
ER and Class Diagrams





Use Case and Sequential Diagrams





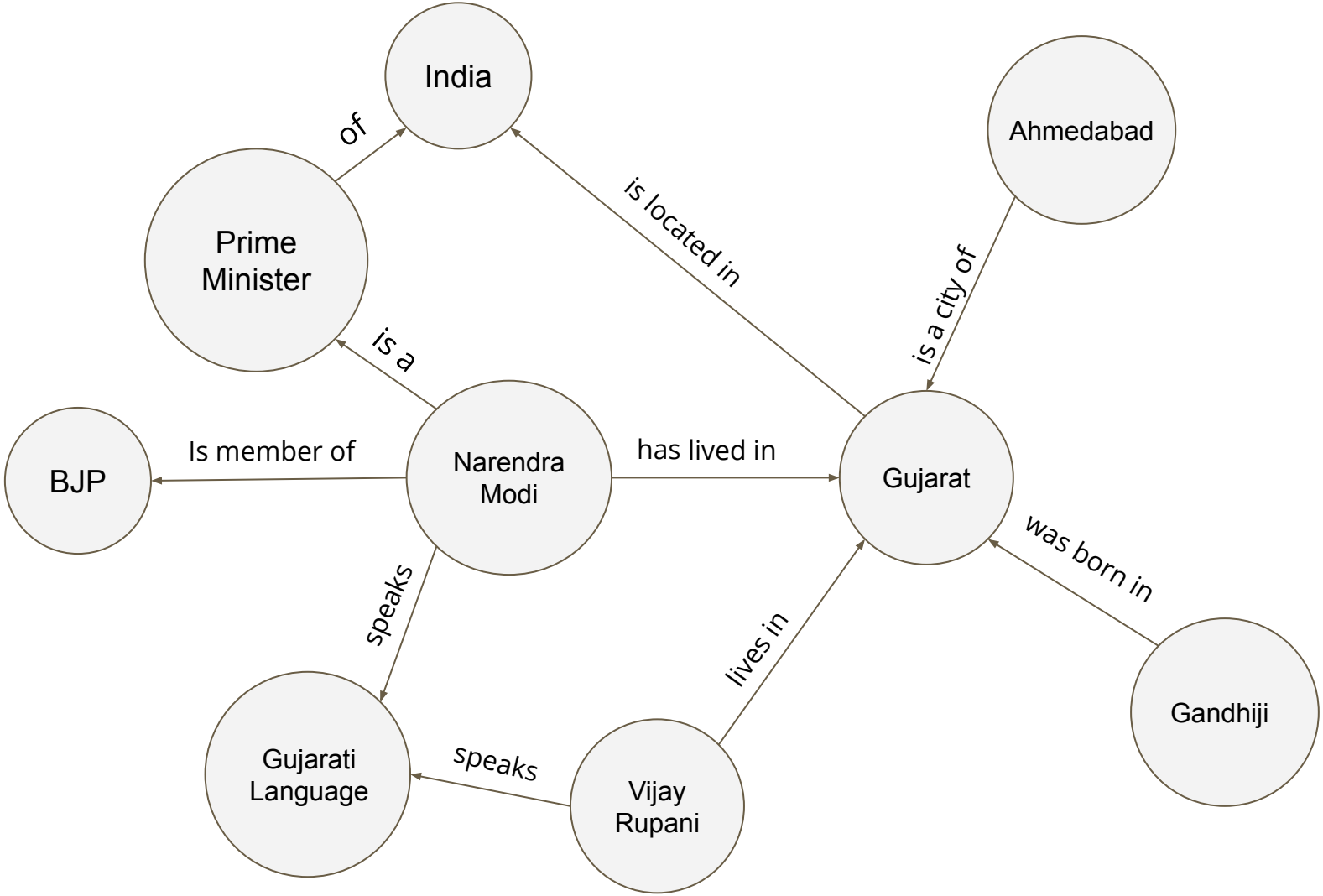
Limitations

- **Rules** for only **limited sentences** have been defined
- Can **answer** only **four types** of questions
- Sentences with **complex structure** have not been implemented yet
- Sentences with the similar meaning but different structure will generate **redundant entities**
- Extraction of entities takes into account the dependency tree of the sentence as the concept is highly **dependent on parsing of dependency trees**

Future Scope

- The **complex sentences** can be solved by creating their individual rules
- **More types of questions** can be answered by extracting maximum data out of the context in form of entities
- Better **searching algorithms** can be used for quicker and more accurate answers
- **User feedback** for correct answers can be taken to improve the accuracy
- The data accumulated can be trained to create **neural network** for improving the QnA system.
- After certain level of improvements, the system can be **deployed** in digital libraries, used in a classroom to solve queries of students or can be used to retrieve certain answers from product manuals.

THANK YOU


India
Ahmedabad
of
Prime Minister
is located in
is a city of
is a
BJP
Is member of
Narendra Modi
has lived in
Gujarat
speaks
was born in
lives in
Gandhiji
Gujarati Language
speaks
Vijay Rupani