Part 1: Creation of a knowledge base

1. Firstly, I have read the json file and extract all the text from that file.

2. Then using NLTK, I have done the sentence tokenization of the paragraph and then word tokenization of each sentence.

3. Then I tagged the words using POS tagging.

4. Then I performed the NER for relation extraction.

5. Then using NLTK, relation extractor, I have found the relation between each pair of named entities that I have specified using a regular expression between them.

6. Then I list all the Entity pairs, relation between them and the related paragraph from where they are extracted.

Result:

```
[ORG: 'Artistic/NNP'] 'Director/NNP of/IN the/DT' [ORG: 'National/NNP Youth/NNP']
816
Edward Wilson (actor)

Edward William "Ed" Wilson, FRSA (13 July 1947 - 2 February 2008) was an English actor and the Artistic Director of the National Youth Theatre from 1987-2003; he later moved to Los Angeles.

[ORG: 'English/NNP High/NNP Court/NNP'] 'of/IN' [ORG: 'Justice/NNP']
817
Gavin Lightman

Sir Gavin Anthony Lightman (born 20 December 1939), styled The Hon. Mr Justice Lightman, has been a judge of the English High Court of Justice, Chancery Division, since 1994. He retired from

[ORG: 'Academy/NNP'] 'was/VBD founded/VBN in/IN 1774/CD by/IN' [PER: 'Alexander/NNP Macwhorter/NNP']
826
Newark Academy

Newark Academy is the second oldest day school in the state of New Jersey, and one of 11 pre-Revolutionary schools still operating in the United States. The Academy was founded in 1774 by Al
 Mission Statement : Newark Academy will contribute to society thoughtful and compassionate citizens who embrace their responsibilities as ethical, intellectual and civic stewards in the gl
```

7. Using this I have created the knowledge base, which is used to fetch data while chatting with chat bot.

```json
{
    "company": [
        {
            "id": 0,
            "name": "Acumen Resources Ltd",
            "ceo": "Paul Walsh",
            "subsidiaries": ["Actuarial Community UK", "Actuarial Community Ireland", "Actuarial Community Europe"]
        },
        {
            "id": 1,
            "name": "BAE Systems",
            "ceo": "Roger Carr"
        },
        {
            "id": 2,
            "name": "Mind Your Business Inc",
            "foundation-date": "1996",
            "ceo": "Karen Caruso"
        },
        {
            "id": 3,
            "name": "The Walt Disney Co.",
            "subsidiaries": ["Walt Disney Music Company"],
            "foundation-date": "October 1949",
            "location": "Burbank, California"
        },
        {
            "id": 4,
            "name": "EndoTex Interventional Systems, Inc",
            "ceo": "John Maroney",
            "foundation-date": "2007",
            "founders": ["John Maroney"]
        },
```

Part 2: Creation of a chat bot

1. I have used RASA for the creation of the chatbot.

2. It have inbuilt models to learn from the conversion and give the results dynamically.

```yaml
pipeline:
  - name: "WhitespaceTokenizer"
  - name: "RegexFeaturizer"
  - name: "LexicalSyntacticFeaturizer"
  - name: "CountVectorsFeaturizer"
  - name: "CountVectorsFeaturizer"
    analyzer: "char_wb"
    min_ngram: 1
    max_ngram: 4
  - name: "DIETClassifier"
    epochs: 100
  - name: "EntitySynonymMapper"

policies:
  - name: RulePolicy
```

RASA uses these models to train the overall chatbot.

3. I have mentioned the intents, entities, actions and the responses in the domain.yml file.

```yaml
actions:
- action_query_knowledge_base

responses:
  utter_greet:
  - text: "Hey!"
  - text: "Hello! How can I help you?"

  utter_goodbye:
  - text: "Bye"
  - text: "Goodbye. See you soon."

  utter_ask_rephrase:
  - text: "Sorry, I'm not sure I understand. Can you rephrase?"
  - text: "Can you please rephrase? I did not got that."

  utter_iamabot:
  - text: "I am a bot, powered by Rasa."
```

```yaml
intents:
  - greet
  - goodbye
  - query_knowledge_base
  - bot_challenge

entities:
  - object_type
  - mention
  - attribute
  - company
```

4. Rules are mentioned in the rules.yml file.

```yaml
version: "3.1"
rules:
  - rule: greet
    steps:
      - intent: greet
      - action: utter_greet
  - rule: goodbye
    steps:
      - intent: goodbye
      - action: utter_goodbye
  - rule: query knowledge base
    steps:
      - intent: query_knowledge_base
      - action: action_query_knowledge_base
  - rule: bot challenge
    steps:
      - intent: bot_challenge
      - action: utter_iamabot
```

5. Stories are mentioned in the stories.yml file.

```yaml
stories:
  - story: greet and then query knowledge base
    steps:
    - intent: greet
    - action: utter_greet
    - intent: query_knowledge_base
    - action: action_query_knowledge_base
```

6. NLU intent and respective examples are mentioned in the nlu.yml file.

```yaml
nlu:
  - intent: bot_challenge
    examples: |
      - are you a bot?
      - are you a human?
      - am I talking to a bot?
      - am I talking to a human?

  - intent: greet
    examples: |
      - hey
      - hello
      - hi
      - good morning
      - good evening
      - hey there

  - intent: goodbye
    examples: |
      - bye
      - goodbye
      - see you around
      - see you later

- intent: query_knowledge_base
  examples: |
    - what [companies]{"entity": "object_type", "value": "company"} can you recommend?
    - list some [companies]{"entity": "object_type", "value": "company"}.
    - can you name some [companies]{"entity": "object_type", "value": "company"} please?
    - can you show me some [companies]{"entity": "object_type", "value": "company"} options?
    - are there any [subsidiaries]{"entity": "attribute", "value": "subsidiaries"} of the [Acumen Resources Ltd]{"entity": "com
    - are there any [subsidiaries]{"entity": "attribute", "value": "subsidiaries"} of the [first]{"entity": "mention", "value":
    - are there any [subsidiaries]{"entity": "attribute", "value": "subsidiaries"} of [that one]{"entity": "mention"}?
    - are there any [subsidiaries]{"entity": "attribute", "value": "subsidiaries"} of [this company]{"entity": "mention"}?
    - list all the [subsidiaries]{"entity": "attribute", "value": "subsidiaries"} of the [Acumen Resources Ltd]{"entity": "comp
    - list all the [subsidiaries]{"entity": "attribute", "value": "subsidiaries"} of the [fifth]{"entity": "mention", "value":
    - list all the [subsidiaries]{"entity": "attribute", "value": "subsidiaries"} of [that one]{"entity": "mention"}
    - list all the [subsidiaries]{"entity": "attribute", "value": "subsidiaries"} of [this company]{"entity": "mention"}
    - who is the [ceo]{"entity": "attribute", "value": "ceo"} of the [third]{"entity": "mention", "value": "3"} one?
    - who is the [ceo]{"entity": "attribute", "value": "ceo"} of [that one]{"entity": "mention"}?
    - who is the [ceo]{"entity": "attribute", "value": "ceo"} of [this company]{"entity": "mention"}?
    - who is the [ceo]{"entity": "attribute", "value": "ceo"} of the [BAE Systems]{"entity":"company"}?
    - what is the [foundation date]{"entity": "attribute", "value": "foundation-date"} of [Mind Your Business Inc]{"entity": "c
    - what is the [foundation date]{"entity": "attribute", "value": "foundation-date"} of [last]{"entity": "mention", "value":
    - what is the [foundation date]{"entity": "attribute", "value": "foundation-date"} of [that one]{"entity": "mention"}?
    - what is the [foundation date]{"entity": "attribute", "value": "foundation-date"} of [this company]{"entity": "mention"}?
    - who are the [founders]{"entity": "attribute", "value": "founders"} of [fourth]{"entity": "mention", "value": "4"} one?
```

7. Action for quering knowledge base in written in actions.py file.

```python
from rasa_sdk.knowledge_base.storage import InMemoryKnowledgeBase
from rasa_sdk.knowledge_base.actions import ActionQueryKnowledgeBase


class ActionMyKB(ActionQueryKnowledgeBase):
    def __init__(self):
        # load knowledge base with data from the given file
        knowledge_base = InMemoryKnowledgeBase("knowledge_base_data.json")

        # overwrite the representation function of the company object
        # by default the representation function is just the name of the object
        knowledge_base.set_representation_function_of_object(
            "company", lambda obj: obj["name"]
        )

        super().__init__(knowledge_base)
```

Results:

1.

2.

```
Your input ->  hi
Hey!
Your input ->  what companies can you recommend?
Found the following objects of type 'company':
1: EurActiv media network
2: BAE Systems
3: Acumen Resources Ltd
4: Bertelsmann Foundation
5: Bettys & Taylors Group Limited
Your input ->  what is the foundation date of the last one?
'Bettys & Taylors Group Limited' has the value '1886' for attribute 'foundation-date'.
Your input ->  what is the location of this company?
'Bettys & Taylors Group Limited' has the value 'Yorkshire, England' for attribute 'location'.
Your input ->  are there any subsidiaries of that one?
'Bettys & Taylors Group Limited' has the value '['Bettys', 'Taylors of Harrogate', 'Yorkshire Tea']' for attribute 'subsidiaries'.
Your input ->
```

3.

```
Hello! How can I help you?
Your input ->  Hi
Hey!
Your input ->  list all the companies
Found the following objects of type 'company':
1: Acumen Resources Ltd
2: Central Glass Co.
3: The Walt Disney Co.
4: Mind Your Business Inc
5: Mountainside Oral Care
Your input ->  are there any subsidiaries of the Acumen Resources Ltd?
'Acumen Resources Ltd' has the value '['Actuarial Community UK', 'Actuarial Community Ireland', 'Actuarial Community Europe']' for attribute 'subsidiaries'.
Your input ->  who is the ceo of the Acumen Resources Ltd?
'Acumen Resources Ltd' has the value 'Paul Walsh' for attribute 'ceo'.
Your input ->  what is the foundation date of the third one?
'The Walt Disney Co.' has the value 'October 1949' for attribute 'foundation-date'.
Your input ->  what is the location of this company?
'The Walt Disney Co.' has the value 'Burbank, California' for attribute 'location'.
Your input ->
```

4.

```
Your input ->  Hi
Hello! How can I help you?
Your input ->  list all the companies
Found the following objects of type 'company':
1: EurActiv media network
2: The Walt Disney Co.
3: BAE Systems
4: EndoTex Interventional Systems, Inc
5: Tata Chemicals Limited
Your input ->  who is the ceo of the BAE Systems?
'BAE Systems' has the value 'Roger Carr' for attribute 'ceo'.
Your input ->  who is the founder of the first one?
Can you please rephrase? I did not got that.
Your input ->  who are the founders of the first one?
'EurActiv media network' has the value '['Christophe Leclercq']' for attribute 'founders'.
Your input ->  Are there any subsidiaries of the Tata Chemicals Limited?
'Tata Chemicals Limited' has the value '['Tata Chemicals Europe']' for attribute 'subsidiaries'.
Your input ->
```

5.

```
Your input ->  Hi
Hello! How can I help you?
Your input ->  can you show me some companies?
Found the following objects of type 'company':
1: Acumen Resources Ltd
2: EurActiv media network
3: Central Glass Co.
4: The Walt Disney Co.
5: Tata Chemicals Limited
Your input ->  what is the foundation date of the central glass co.?
Sorry, I'm not sure I understand. Can you rephrase?
Your input ->  what is the foundation date of the third one?
'Central Glass Co.' has the value '1946' for attribute 'foundation-date'.
Your input ->  who are the founders of that one?
'Central Glass Co.' has the value '['John Dale', 'Stanley Gordon']' for attribute 'founders'.
Your input ->  what is the location of this company?
'Central Glass Co.' has the value 'Brockton, Massachusetts' for attribute 'location'.
Your input ->
```

6.

```
Your input ->  hi
Hey!
Your input ->  list some companies
Found the following objects of type 'company':
1: The Walt Disney Co.
2: BAE Systems
3: Acumen Resources Ltd
4: EndoTex Interventional Systems, Inc
5: Bettys & Taylors Group Limited
Your input ->  who is the ceo of the fourth one?
'EndoTex Interventional Systems, Inc' has the value 'John Maroney' for attribute 'ceo'.
Your input ->  what is the foundation date this company?
'EndoTex Interventional Systems, Inc' has the value '2007' for attribute 'foundation-date'.
Your input ->  who are the founders of that company?
'EndoTex Interventional Systems, Inc' has the value '['John Maroney']' for attribute 'founders'.
Your input ->
```

7.

```
Your input ->  hi
Hey!
Your input ->  list some companies
Found the following objects of type 'company':
1: The Walt Disney Co.
2: BAE Systems
3: Mountainside Oral Care
4: EndoTex Interventional Systems, Inc
5: Tata Chemicals Limited
Your input ->  what is the foundation date of the third one?
'Mountainside Oral Care' has the value '2010' for attribute 'foundation-date'.
Your input ->  who are the founders of that one?
'Mountainside Oral Care' has the value '['Dr. Olaru']' for attribute 'founders'.
Your input ->  are there any subsidiaries of the last one?
'Tata Chemicals Limited' has the value '['Tata Chemicals Europe']' for attribute 'subsidiaries'.
Your input ->
```

8.

```
Your input ->  hi
Hey!
Your input ->  what companies can you recommend?
Found the following objects of type 'company':
1: Mountainside Oral Care
2: Acumen Resources Ltd
3: Tata Chemicals Limited
4: Mind Your Business Inc
5: Bettys & Taylors Group Limited
Your input ->  at which place the headquarters of the last one located?
Sorry, I'm not sure I understand. Can you rephrase?
Your input ->  what is the location of the last one?
'Bettys & Taylors Group Limited' has the value 'Yorkshire, England' for attribute 'location'.
Your input ->  at which place the headquaters of the last one located?
'Bettys & Taylors Group Limited' has the value 'Yorkshire, England' for attribute 'location'.
Your input ->  are there any subsidiaries of that one?
'Bettys & Taylors Group Limited' has the value '['Bettys', 'Taylors of Harrogate', 'Yorkshire Tea']' for attribute 'subsidiaries'.
Your input ->
```

9.

```
Your input ->  hi
Hello! How can I help you?
Your input ->  list some companies
Found the following objects of type 'company':
1: Bertelsmann Foundation
2: EurActiv media network
3: Tata Chemicals Limited
4: Central Glass Co.
5: EndoTex Interventional Systems, Inc
Your input ->  who is the ceo of the first one?
'Bertelsmann Foundation' has the value 'Gunter Thielen' for attribute 'ceo'.
Your input ->  at which place the headquaters of the fourth one located?
'Central Glass Co.' has the value 'Brockton, Massachusetts' for attribute 'location'.
Your input ->  what is the foundation date of that one?
'Central Glass Co.' has the value '1946' for attribute 'foundation-date'.
Your input ->
```

10.

```
Your input ->  hi
Hey!
Your input ->  list some companies
Found the following objects of type 'company':
1: Mind Your Business Inc
2: BAE Systems
3: Central Glass Co.
4: Bertelsmann Foundation
5: Acumen Resources Ltd
Your input ->  who is the ceo of the first one?
'Mind Your Business Inc' has the value 'Karen Caruso' for attribute 'ceo'.
Your input ->  who is the ceo of the mind your business inc?
'Mind Your Business Inc' has the value 'Karen Caruso' for attribute 'ceo'.
Your input ->  what is the foundation date of this company?
'Mind Your Business Inc' has the value '1996' for attribute 'foundation-date'.
Your input ->
```

11.

```
Your input ->  hi
Hey!
Your input ->  list some companies
Found the following objects of type 'company':
1: Mind Your Business Inc
2: Bertelsmann Foundation
3: Central Glass Co.
4: Bettys & Taylors Group Limited
5: The Walt Disney Co.
Your input ->  list all the subsidiaries of the last one
'The Walt Disney Co.' has the value '['Walt Disney Music Company']' for attribute 'subsidiaries'.
Your input ->  what is the foundation date of this company?
'The Walt Disney Co.' has the value 'October 1949' for attribute 'foundation-date'.
Your input ->  what is the location of this company?
'The Walt Disney Co.' has the value 'Burbank, California' for attribute 'location'.
Your input ->
```

12.

```
Your input ->  hi
Hey!
Your input ->  list some companies
Found the following objects of type 'company':
1: Mountainside Oral Care
2: EurActiv media network
3: Central Glass Co.
4: Bettys & Taylors Group Limited
5: The Walt Disney Co.
Your input ->  who are the founders of the fourth one?
'Bettys & Taylors Group Limited' has the value '['Frederick Belmont']' for attribute 'founders'.
Your input ->  what is the foundation date of the first one?
'Mountainside Oral Care' has the value '2010' for attribute 'foundation-date'.
Your input ->  who are the founders of the third one?
'Central Glass Co.' has the value '['John Dale', 'Stanley Gordon']' for attribute 'founders'.
Your input ->
```

13.

```
Your input ->  hi
Hello! How can I help you?
Your input ->  list some companies
Found the following objects of type 'company':
1: Mountainside Oral Care
2: Tata Chemicals Limited
3: The Walt Disney Co.
4: Central Glass Co.
5: Acumen Resources Ltd
Your input ->  who is the ceo of the acumen resources ltd?
'Acumen Resources Ltd' has the value 'Paul Walsh' for attribute 'ceo'.
Your input ->  list all the subsidiaries of the second one
'Acumen Resources Ltd' has the value '['Actuarial Community UK', 'Actuarial Community Ireland', 'Actuarial Community Europe']' for attribute 'subsidiaries'.
```

14.

```
Your input ->  hi
Hey!
Your input ->  can you name some companies please?
Found the following objects of type 'company':
1: Tata Chemicals Limited
2: EndoTex Interventional Systems, Inc
3: Central Glass Co.
4: Mountainside Oral Care
5: The Walt Disney Co.
Your input ->  who is the ceo of the second one?
Can you please rephrase? I did not got that.
Your input ->  what is the foundation date of the third one?
'Central Glass Co.' has the value '1946' for attribute 'foundation-date'.
Your input ->  what is the location of the last one?
'The Walt Disney Co.' has the value 'Burbank, California' for attribute 'location'.
Your input ->  list all the subsidiaries of this company
'The Walt Disney Co.' has the value '['Walt Disney Music Company']' for attribute 'subsidiaries'.
Your input ->
```

15.

```
Your input -> hi
Hello! How can I help you?
Your input -> list some companies
Found the following objects of type 'company':
1: The Walt Disney Co.
2: Acumen Resources Ltd
3: Bettys & Taylors Group Limited
4: EurActiv media network
5: BAE Systems
Your input -> what is the location of the third one?
'Bettys & Taylors Group Limited' has the value 'Yorkshire, England' for attribute 'location'.
Your input -> what is the foundation date of this company?
'Bettys & Taylors Group Limited' has the value '1886' for attribute 'foundation-date'.
Your input -> what is the foundation date of that one?
'Bettys & Taylors Group Limited' has the value '1886' for attribute 'foundation-date'.
Your input -> who is the ceo of the last one?
'BAE Systems' has the value 'Roger Carr' for attribute 'ceo'.
Your input ->
```

Limitations:

1. This chat bot remembers last few interactions and started to forget distant conversations. This can be improved so that it can be made for reliable.

2. For the entries not present in the knowledge base, It asks to repeat the query. It can be improved, so that it replies that the data in not present in the knowledge base instead for repeatedly ask for repeating the query.

Contribution:

1. Chirag (B19CSE026)

Colab File link:

https://colab.research.google.com/drive/1osCuBS2Ddl7XOy5kWR2acz-S6IA1YSfo?usp=sharing

Drive link:

https://drive.google.com/drive/folders/1NlsYGO48_eqFDC0pcbDWg5V8hDmPZiM_?usp=sharing