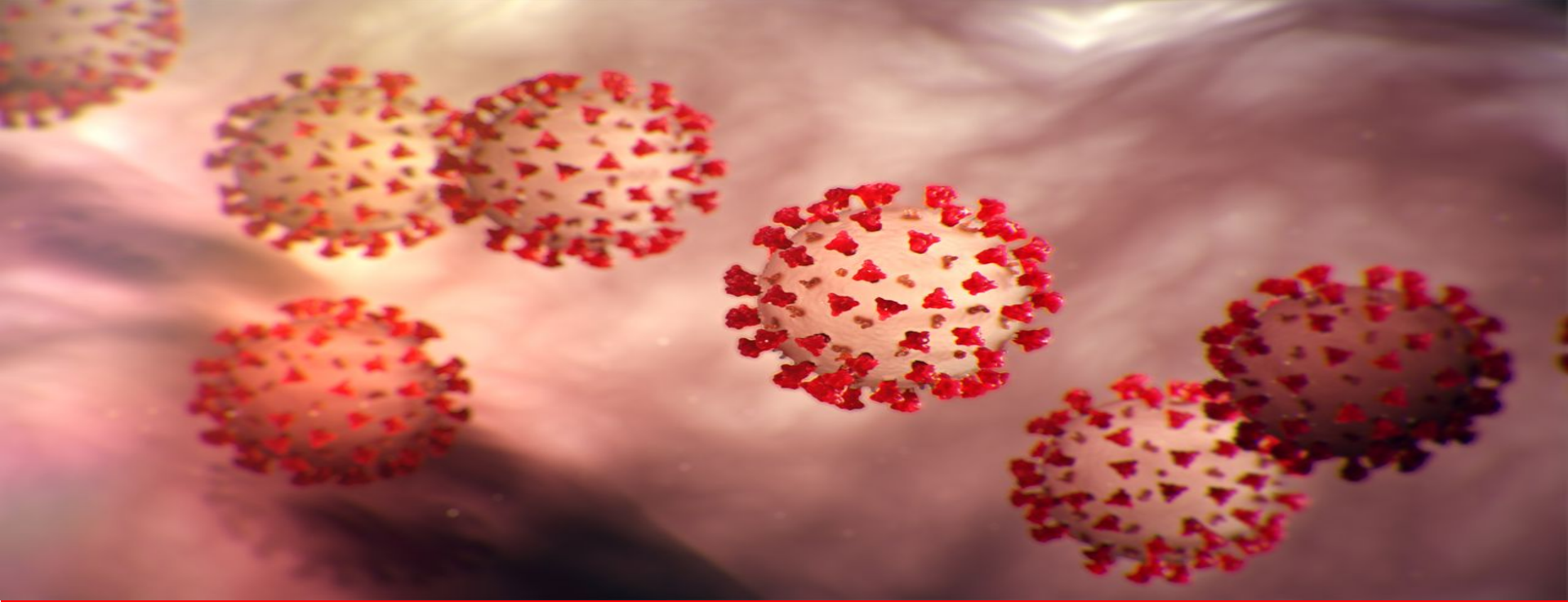


COVID-19 Chatbot Challenge

Anand S Menon

TABLE OF CONTENTS

PREFACE

We know we are in a very uncertain and hostile situation amid the COVID outbreak and as a human being it's our duty to do something about it and as a human being i should contribute my fare share towards this cause.So this is conversation assistant which can provide its users with important information regarding corona outbreak like it's global statistics, preventive measures ...etc. In the current situation a chatbot like this can be very useful to people around the world to have an up to date information about this deadly disease. So i will take you through the steps i have followed to collect data, build model, tools used, deployment strategies ..etc.

The intention of this bot is educational and my concern towards this pandemic.Thanks to iNeuron for setting up this challenge.

INTRODUCTION

COVID-19 is spreading and spreading rapidly, taking severe measure to stop this is of utmost importance to the human race. Government is doing everything within their power to stop this catastrophe but that alone won't be enough. So as a citizen we should also need to take necessary steps to keep yourself as well as the people around us healthy and safe from this deadly virus.

In this project i will be showcasing steps by step process of building a conversational assistant which can be used for the following.

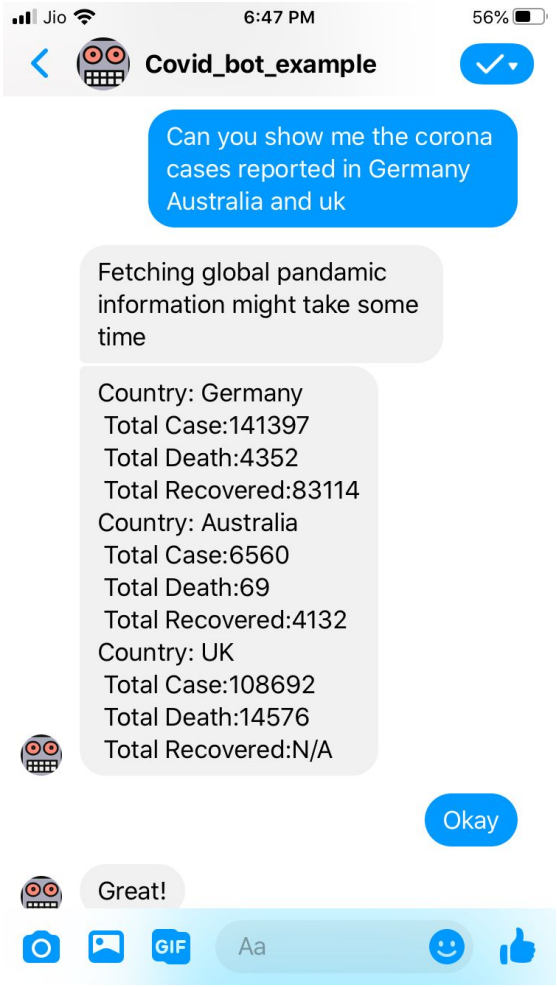
- Up to date COVID-19 **statistics** of each country
- COVID-19 prevention **notification** to the user
- Latest global pandemic report using **visual maps**
- Accessibility through famous platforms like **facebook** and **telegram**

Before getting into the details lets checkout some cool features of the application:

As we can see the user can ask about any country in the world and bot would fetch those information from a global API and provides useful informations like

1. **Number of people Affected**
2. **Number of people dead**
3. **Number of people recovered**

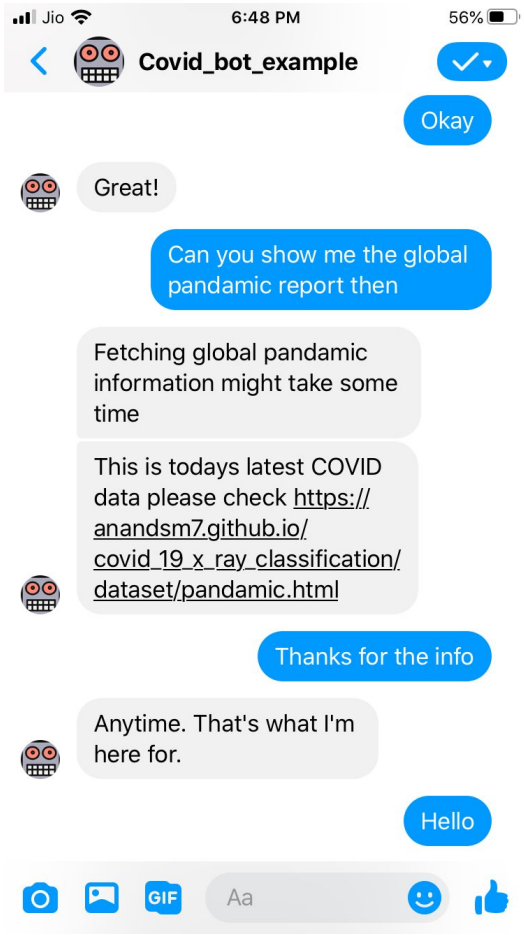
Made use the rapidAPI to fetch the updated COVID info
https://coronavirus-info.p.rapidapi.com/country



There will be cases in which the user never asks for a specific country.So in this situation fetching the informations of all covid affected countries is not feasible and the global data extraction can be very time consuming.

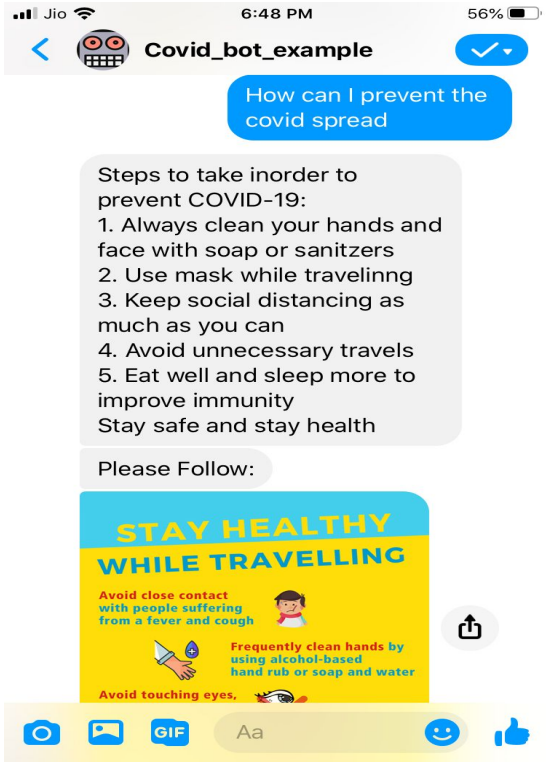
So in order to tackle such a situation i will be hosting a flask base website in heroku which will be showing covid statistics of every affected countries in the world using a world map.

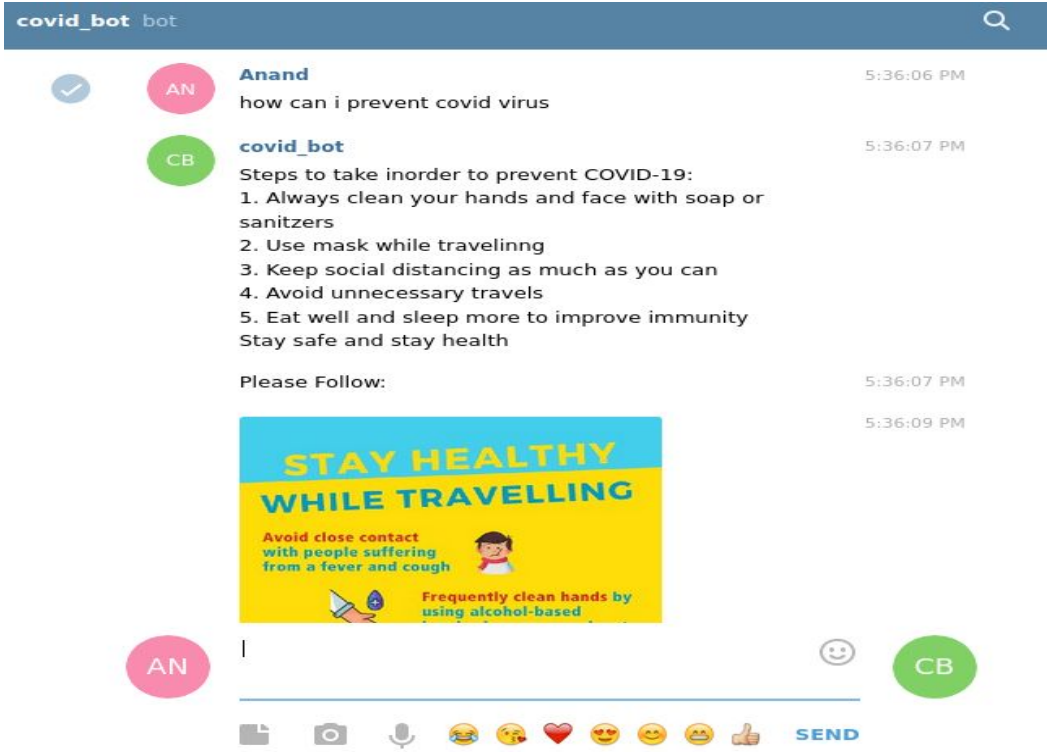
You can directly access this website as well
https://covidworldwidemap.herokuapp.com/insight



Facebook/Telegram Bot Deployment

The bot will be deployed in both facebook messenger as well as telegram as shown below

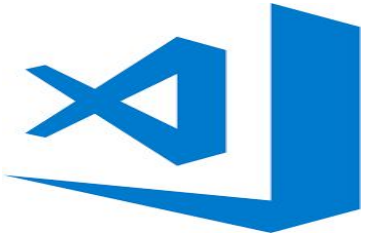




PREREQUISITES

In order to build the conversation assistant the prerequisites are as follows:

- **RASA** - Open source conversational AI
  - pip install rasa==1.6.0
- **Flask -** Microservice framework for web app/REST apis
  - pip install flask==1.1.1
- **MongoDB** - NoSQL database
  - https://docs.mongodb.com/manual/installation/
- **IDE** - VS Code/PyCharm/Spyder
  - https://github.com/microsoft/vscode/releases
  - https://www.anaconda.com/
  - https://www.jetbrains.com/pycharm/



RASA

When it comes to chatbot development framework there are several to choose from like luis,dialogue flow,Lex,RASA ..etc. In this project i will be going with the RASA chatbot framework because of several reasons like

- **Leveraging NLP for building scalable data driven models**
  - Data driven solutions are always great when it comes to building chatbot since there is surge in data growth and scalability is very important.RASA uses latest NLP techniques when it comes to intent identification and entity extraction.The SoTA RASA framework leverages BERT/ConvERT/DIET models for building the model pipeline.

- **Open Source**
  - One reason which is very critical is the open source nature of RASA. We can do any hack/customize to this framework in any way you want, like adding custom NLP components like custom tokenizer/extractor/classifier pipelines to our existing model pipeline which is very much useful for building custom bot solutions.

- **Deployment Environments**
  - Deployment environment unless local machines can be very costly at times. As a newbie and fascinated developer money always comes as a hurdle in between our imaginations.This is where RASA comes handy ie it can be deployed anywhere you want let it be your local machines, private cloud, public clouds, famous chatbot platforms like facebook, telegram, slack you name it. Where as the other framework kind of struggle in these task.

The other mentioned framework has a lot of cool features and strength but for this project ii will be sticking with RASA..

For more info on RASA please visit
https://rasa.com/docs/getting-started/

Now let's dive deep into RASA and its component. As mentioned before RASA is OSS conversational AI which helps developers in building SToA chatbot leveraging ML.

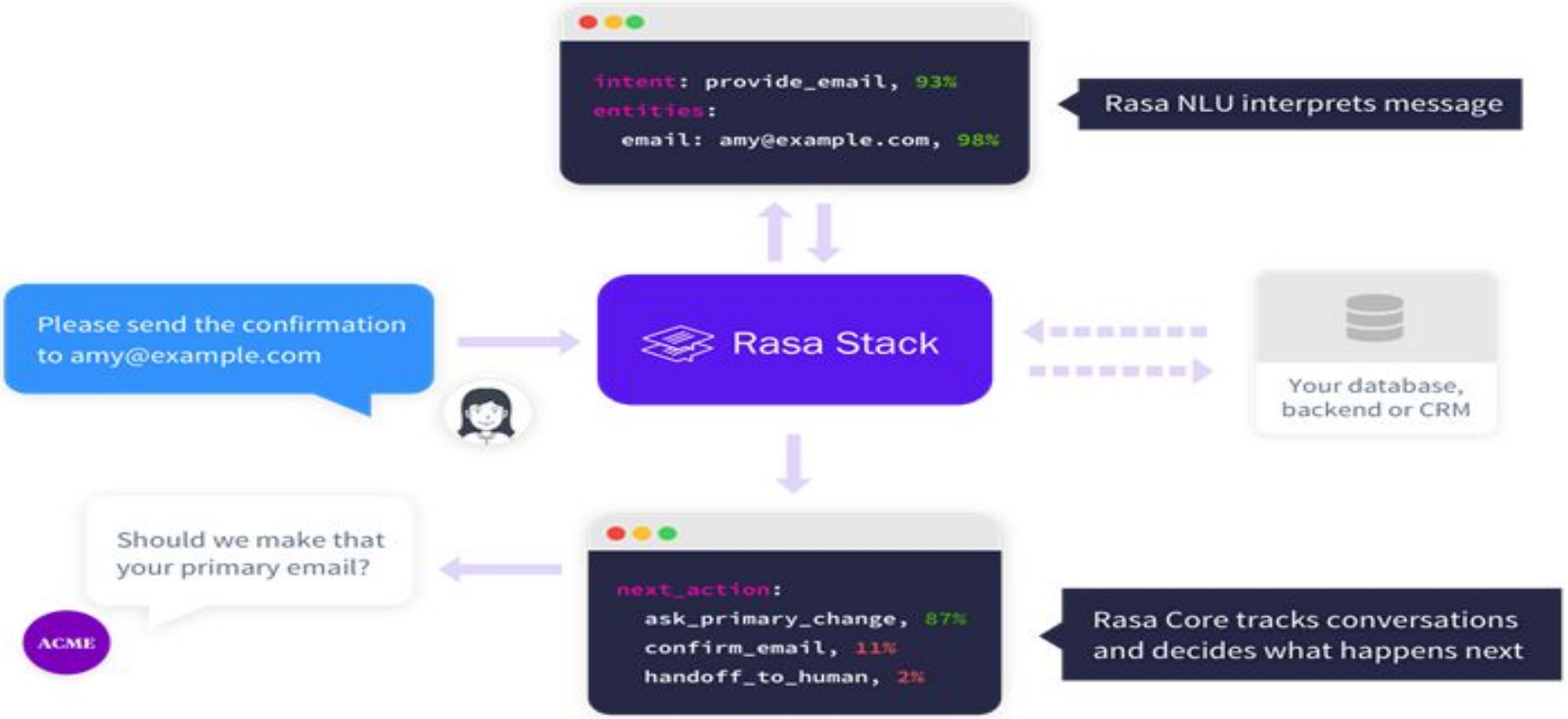
The RASA stack has two parts

- RASA NLU
- RASA Core

NLU deals with intent identification and entity extraction which is very useful in chatbot architecture.
eg: I would like to know the corona virus spread report in Germany and Italy ?
Here the intent is covid details and entity extracted or of countries

Core on the other hands decides what actions are to be taken based on intent and entity combinations. It can be calling a method, fetching something from a database, hitting an api ..etc.
In the above example it might hit an api with country details to fetch the current covid reports of that country

RASA ARCHITECTURE



RASA NLU components are as follows

- nlu.md/nlu.json
  - This is the training file in markdown/JSON format used for training the NLU model
- config.yml
  - This decides what kind of pipeline we are using like:
    - Type of work tokenizer
    - Type of entity extractor **(CRF/Spacy/Duckling)**
    - Type of classifier **(Embedding/conVERT/DIET)**

```
- i would like to know the covid cases reported
- show me the current cases in india
- number of corona cases in germany
- show me the patient history in usa cases
- how many are affected in australia
- i would like to know the number of covid cases in usa
- may i know the number of people affected by corona in japan
- what is the current corona situation in sweden
- how is many people are affected in china
- in china how many are affected by corona
- how many people are affected with covid in austria
- i would like to see the covid details of korea
- may i have the complete info on germany on covid cases
- show me covid cases in india and china
- i want to know total corona cases in germany and italy
- what is the present situation in italy
- how is spain doing with corona cases
- may i get the total corona cases in norway and sweden
- i would like to get the covid details of germany and china and italy and india and usa
- i would like to get the covid details of germany china italy india usa
- show me detailed report of covid virus spread in india italy and germany
- show my the current covid cases reported in [dubai](GPE:UAE)
- i would like to see corona details of [UAE](GPE)
- how is the current situation is [USA](GPE)
- show me a detailed report of covid cases in [United States](GPE:USA)
- how are the covid reports in [states](GPE:USA)
- fetch me the complete report of covid cases in [united kingdom](GPE:UK)
- show covid 19 reports of [england](GPE:UK)
- may i have a look at reported corona cases in [U.k](GPE:UK)
- how to see the corona spread accross germany
- i want to see worldwide covid report
- show me the world wide report of corona
- may i have look at the covid cases around the world
- show me worldwide corona report
- i would like to see the corona reports of italy
- what is the current situation in italy
```

```
# Configuration for Rasa NLU.
# https://rasa.com/docs/rasa/nlu/components/
language: "en"
pipeline:
  - name: WhitespaceTokenizer
  - name: RegexFeaturizer
  - name: CRFEntityExtractor
  - name: SpacyNLP
  - name: SpacyEntityExtractor
  - name: EntitySynonymMapper
  - name: CountVectorsFeaturizer
  - name: EmbeddingIntentClassifier
```

RASA Core components are as follows

- Domain.yml
  - This is master file where we store all intent/entities/slots/templates/actions/formaction
- Stories.md
  - Stories are basically conversation samples given markdown. The RASA uses a keras based model and train on the stories to decide on which action to call next
- Config.yml
  - Just like NLU core also needs to specify some policies like keras policy, memotization policy ..etc
- Actions.py
  - This is a python scripts where all application logics are written. This is used to many purpose like calling api,db ...etc

```yaml
templates:
   utter_greet:
    - text: "Hi there 😀, Welcome to COVID-19 helpline 🏥."
   utter_ask_person:
    - text: "May i know your full name please?"
    - text: "Could you please tell me your full name"
   utter_ask_phone:
    - text: "Hi 🙍‍♂️.Could you please provide your phone number 📱"
    - text: "Hi 🙍‍♂️.Can you provide your contact number 📱"
   utter_ask_pin:
    - text: "May i know your pin for upto date notification on COVID spread"
    - text: "I will be needing your pin for upto date notification on COVID spread"
   utter_ask_email:
    - text: "Almost done could you please provide your Email ID"
   utter_cheer_up:
    - text: "Please Follow:"
      image: "https://i.imgur.com/3H1hkNJ.png"


entities:
   - country
   - state
   - phone
   - email
   - pin
   - GPE
   - PERSON

actions:
- action_faq
- action_covid
- fetch_info

forms:
   - detail_fetch
```

```yaml
# Configuration for Rasa Core.
# https://rasa.com/docs/rasa/core/policies/
policies:
  - name: MemoizationPolicy
  - name: KerasPolicy
  - name: MappingPolicy
  - name: FormPolicy
```

Model Building

Once the data collection is completed we need to train our model. We can train the model using the command

```
rasa train
```

Now we need to up the rasa model server using the command

```
rasa run --model [model_name] --endpoints [endpoint_yml] --credentials [credentials_yml]
```

We also need to run the rasa action end point using the command

```
rasa run actions
```

There are several other features which are available but not needed in the scope of or project

Sources:

https://rasa.com/docs/rasa/user-guide/rasa-tutorial/

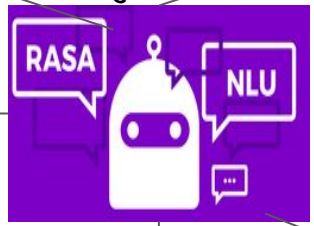
https://rasa.com/docs/rasa/api/http-api/

APPLICATION ARCHITECTURE

Send email notification on covid to the registered users

Send corona preventive measure

Show the corona report of Italy?



Calls covid api in case user provide the country name

Returns info like death/recovered count

In case user doesn't provide country name user is redirected to website showing global covid spread

Intent:
Covid_search
Entity:
Italy

PROJECT IMPLEMENTATION

- NLU model building
  - The first part of the project is the natural language understanding model building. **Data collection** is the first step. I wanted user data on **covid search, covid prevention queries, user names, pin codes, phone numbers, email, country names** to train the model.
  - I have synthetically created **pin code,phone numbers** and made use of the RASA **regexfeaturizer** to extract those data
  - Made use of the **spacy pipeline** extract the **user name** and **country** informations. The issue here was that **indian names are not in the spacy corpus.** So i have collected indian names from **kaggle dataset** to train the entity extractor pipeline.
  - Finally questions regarding **covid search** and **preventive measures** are manually created then augmented, some relevant data is collected from external sites
  - **All these data comes base for the rasa NLU model**

Sources:

- https://www.kaggle.com/chaitanyapatil7/indian-names/version/1
- https://github.com/benoitalvarez/Covid-19-QBox-ChatbotModel/blob/master/Models/Rasa/COVID-19%20FAQ%20RASA.json

Conversation Flow

Small talk:

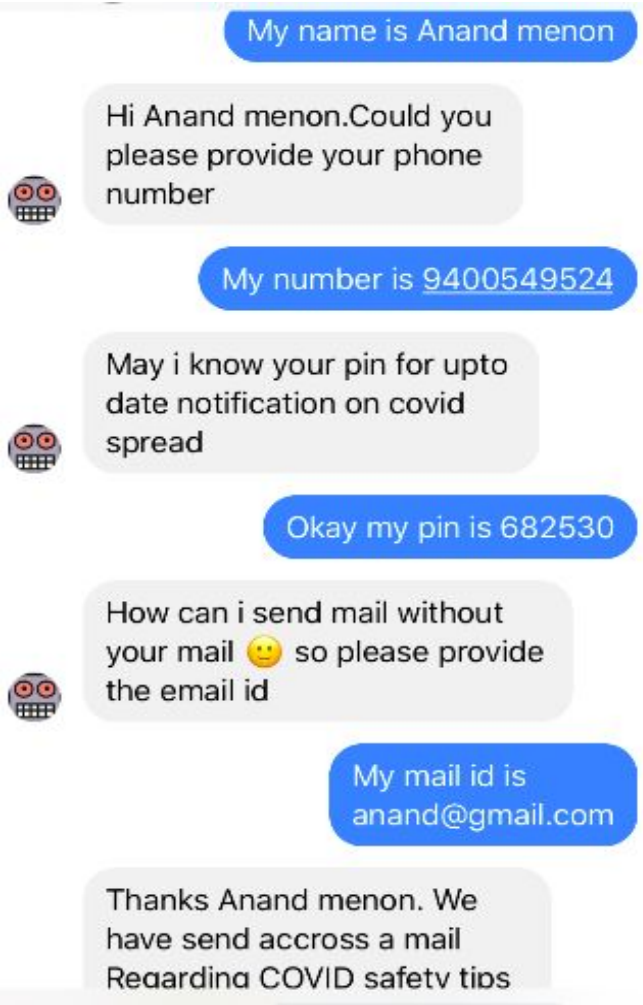
The conversation flow of the chatbot is pretty straight forward

- A user can have any causal conversation with the bot and the **small talk** responses are stored in a dataframe for response generation
- The model is currently trained to answer several other small talk queries like good morning,night,thanks,yes,no,bye...etc
    - Eg:

The model is trained to respond to 80 different small talk intents. So it’s pretty great in having a casual conversation as well

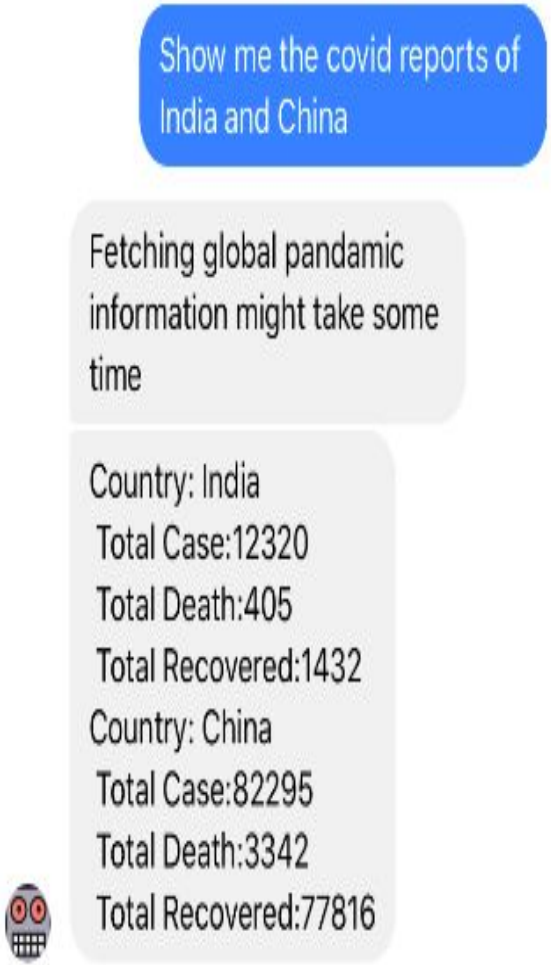
Collecting user information:

- When the user starts with a **“hi”** the bot asks all of his information like n**ame, contact, pin code,email ..etc**. and if everything is correct and validated bot would send a **mail** to the user specifying **preventive measure** to his respective email address
- The bot can understand and extract **indian names as well since indian names are as well added to the training data**
- I have used **regex based validation** to capture to **pin code,contact number and email id**
- Once everything is extracted the action will trigger to send an email to the respective email id using **smtp**
- The mail includes all relevant **details on how to prevent covid spread** with attachments like **world wide covid map links**, **covid preventive guide posters** ..etc.



COVID statistics of affected countries:

- User can ask about covid details of any country and the bot would calls any action called **covid_search** which in turn call an api which respond back with number of people **affected, dead, recovered**
- The custom action will take the extracted entities and make necessary preprocessing on the extracted data and send it across to REST API, which in turn provide COVID statistics
    - https://coronavirus-info.p.rapidapi.com/country
- User can pass any number of country name and the API will fetch these info in a loop and display all the information
- As shown in the example all the extracted countries are send across the REST API and up to date informations are captured



World Wide COVID Visualization:

- There are cases in which the user may not specify the country name. Then we need to provide worldwide covid statistical information.
- Requesting the information of each country and plotting those result for every user request will be time consuming process and user may not have the patience to wait for it to be rendered
- In order to tackle this situation i have hosted a **global website of covid map which renders the latest world wide covid cases which is hosted using heroku platform.**
- For plotting the map we need country names as well as long/lat information so I have made use of the google country dataset for **country names,longitude,latitude**
    - https://developers.google.com/public-data/docs/canonical/countries_csv
- So in case a user doesn't provide country names he will be redirected to this page
- The visualization is done using plotly scatter plot which plots the data using longitude and latitude information
- The map can be viewed from :
    - https://covidworldwidemap.herokuapp.com/insight
- The map doesn't support automatic information updation
- So i have made a endpoint to update with latest information
    - https://covidworldwidemap.herokuapp.com/reset
- The following endpoint will update our dataframe

MongoDB INTEGRATION

MongoDB

It is nosql database used to store data in a JSON document schema. In our use case we need to store information mainly on two areas

- User information
  - Since our chatbot captures user information we can store those information and can be used at point of time
- Chat logs
  - RASA by default uses redis DB to store the tracker log information but rasa has a very low in memory so storing these logs to DB is very important
  - This is very important data which is to be stored in databases. Chat bot logs are very important for log analysis, sentiment analysis, NLU performance monitoring and lot of other stuff which are very useful in the long run

So for our project will be storing user data and logs into the database and i feel like mongoDB is best candidate when it comes to storing/retrieving huge textual data

User data insertion

We will be making use of our rasa action script to insert the data once user data is validated.Since action script is written in python i have used the pymongo library to interacting with my MongoDB and insert the user data. You can install pymongo by simply using the pip install

```
pip install pymongo
```

Once pymongo is installed we can connect with our mongoDB server and connect with db and insert data to collection. As example is given below

```
client = pymongo.MongoClient(params.MONGO_URL)
db = client[params.MONGO_DB]
col = db[params.MONGO_USER]
```

We can securely store the parameters in another python file or as an environment variable to enhance security. Now i order to insert a dict/JSON data to the database is very easy as given below:

```
user_data = {‘name’:’Anand  Menon’,’contact’:’9435345436’,’pin’:’685675’,email:’anand@yahoo.com’}
rasacol.insert_one(user_dict)
```

Now the data will be inserted successfully to the MongoDB database within a collection

Inserting chat logs

Now thanks to RASA inserting rasa logs to MongoDB is simple. We just need to specify the mongo credentials in the endpoints.yml file as given below:

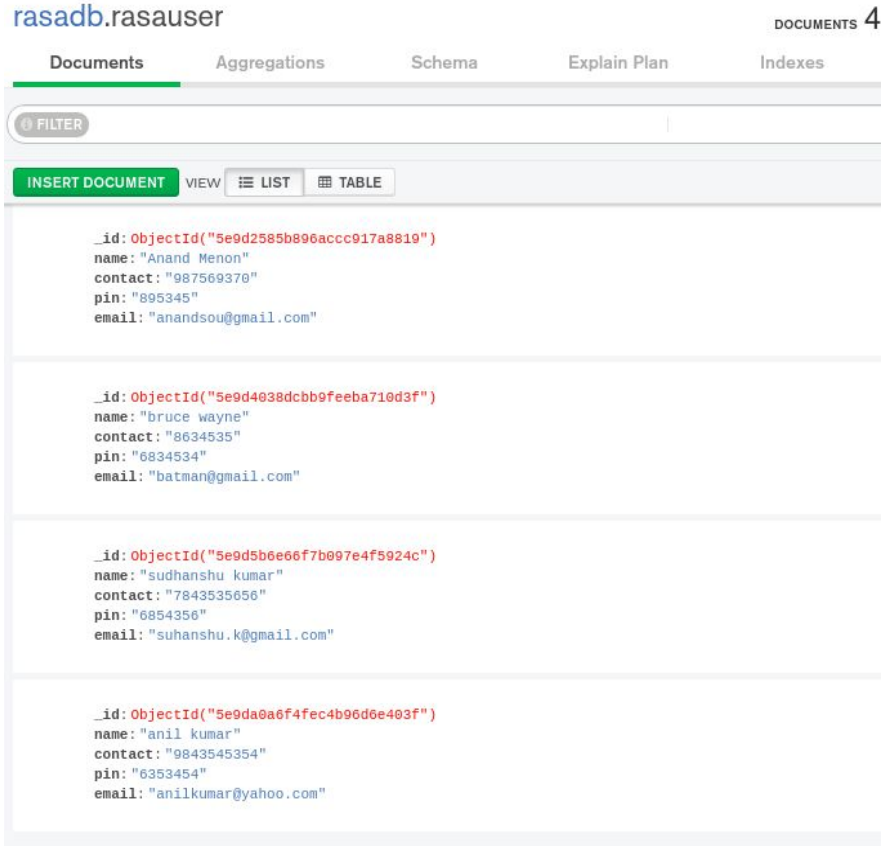
```
tracker_store:
        type: mongod
        url: mongodb://localhost:27017/    #server path
        db: rasadb                         #database name
        Username:                          #username/password is required
        password:
```

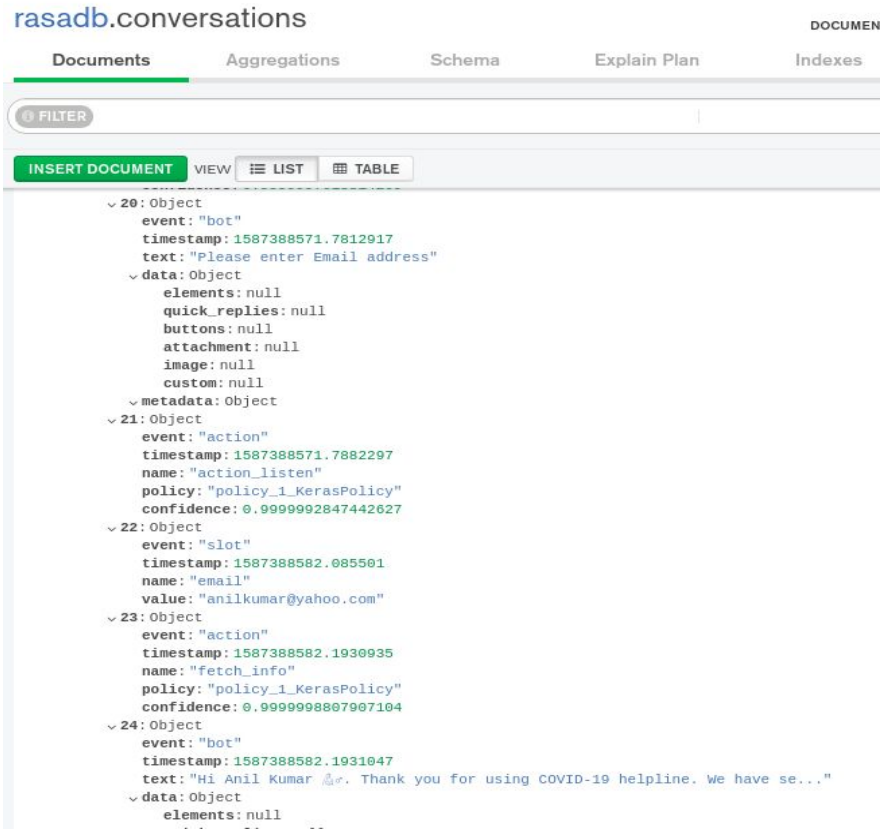
Now when starting the server add the endpoint.yml parameter to set the connection

```
rasa run --endpoints endpoint_yml
```

Now the bot will store all the conversation information like intent/entity identified slots filled action triggered user queries,bot response everything inside the mongo DB

Now both the user user data as well as logs will be inserted to the DB. To show some sample data i have made use of the mongo compass to retrieve the results and are as follows:





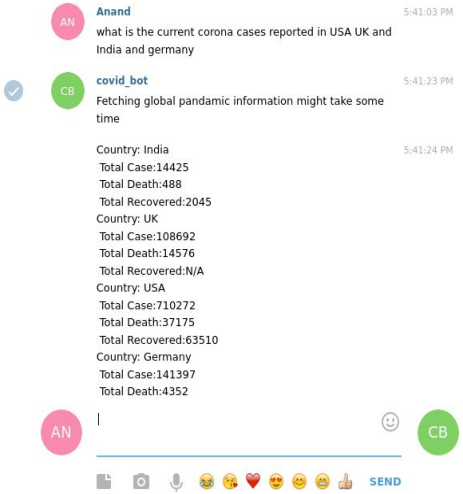
FACEBOOK/TELEGRAM INTEGRATION

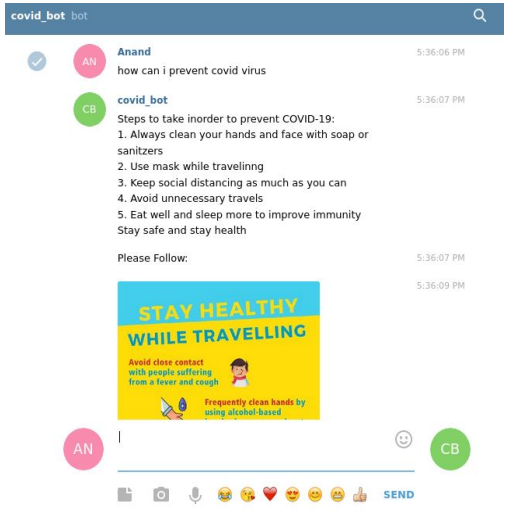
As a final step we can see how to deploy it both facebook/telegram. Now when it comes to deploying bots using rasa we need to set up credentials

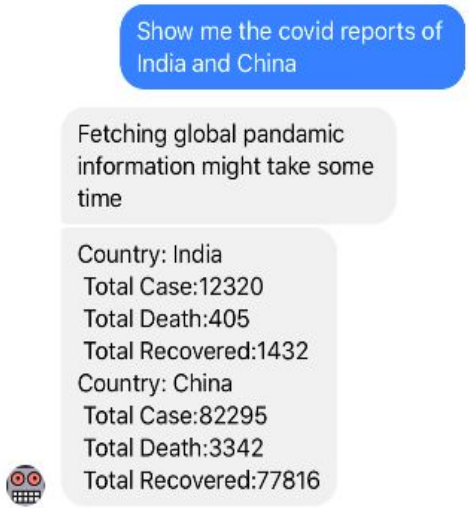
```
telegram:
access_token: <token provided by telegram>
  verify: < name of the bot in telegram>
  webhook_url: <specfiy the webhook>
facebook:
 verify:  <name of the bot>
 secret: <secret provided by facebook>
 page-access-token: <token provided by facebook messenger>
```

Now coming to creating bots in facebook/telegram please
https://developers.facebook.com/
https://web.telegram.org/#/im?p=@BotFather

Sources
https://rasa.com/docs/rasa/user-guide/connectors/telegram/
https://rasa.com/docs/rasa/user-guide/connectors/facebook-messenger/


Anand
5:41:03 PM
what is the current corona cases reported in USA UK and India and germany
covid_bot
5:41:23 PM
Fetching global pandamic information might take some time
5:41:24 PM
Country: India
Total Case:14425
Total Death:488
Total Recovered:2045
Country: UK
Total Case:108692
Total Death:14576
Total Recovered:N/A
Country: USA
Total Case:710272
Total Death:37175
Total Recovered:63510
Country: Germany
Total Case:141397
Total Death:4352
AN
CB
SEND


covid_bot bot
Anand
5:36:06 PM
how can i prevent covid virus
covid_bot
5:36:07 PM
Steps to take inorder to prevent COVID-19:
1. Always clean your hands and face with soap or sanitzers
2. Use mask while travelinng
3. Keep social distancing as much as you can
4. Avoid unnecessary travels
5. Eat well and sleep more to improve immunity
Stay safe and stay health
Please Follow:
5:36:07 PM
5:36:09 PM
STAY HEALTHY
WHILE TRAVELLING
Avoid close contact with people suffering from a fever and cough
Frequently clean hands by using alcohol-based
AN
CB
SEND


Show me the covid reports of India and China
Fetching global pandamic information might take some time
Country: India
Total Case:12320
Total Death:405
Total Recovered:1432
Country: China
Total Case:82295
Total Death:3342
Total Recovered:77816


My name is Anand menon
Hi Anand menon.Could you please provide your phone number
My number is 9400549524
May i know your pin for upto date notification on covid spread
Okay my pin is 682530
How can i send mail without your mail 😁 so please provide the email id
My mail id is anand@gmail.com
Thanks Anand menon. We have send accross a mail Regarding COVID safety tips

SOURCE CODE

The complete code including the model are check in to the github repo.

DEMO

REFERENCES

THANK YOU