Data Collection

1.

A dataset: https://drive.google.com/drive/folders/11sglwm6-cY7gjeqlZaMxL_MDKDMLdhym (https://drive.google.com/drive/folders/11sglwm6-cY7gjeqlZaMxL_MDKDMLdhym). The dataset includes about 600 conversation records. Each record include a description of symptom and a single round conversation between patient and doctor. In the future, the dataset needs to be enlarged. For now, we can use this for our template-based chatbot.

(Thanks for Kavi found a paper which provides this conversation dataset) The group has reviewed this dataset together and we all agree that it can be adopted for now.

2.

Scrawling from website: https://www.justanswer.co.uk/medical/ (https://www.justanswer.co.uk/medical/) JustAnswer is an online QA website contains QA data on different topics including healthcare QA data. I tried to scrawl all medical QA data from this website and get 1338 QA data. Each QA data contains:

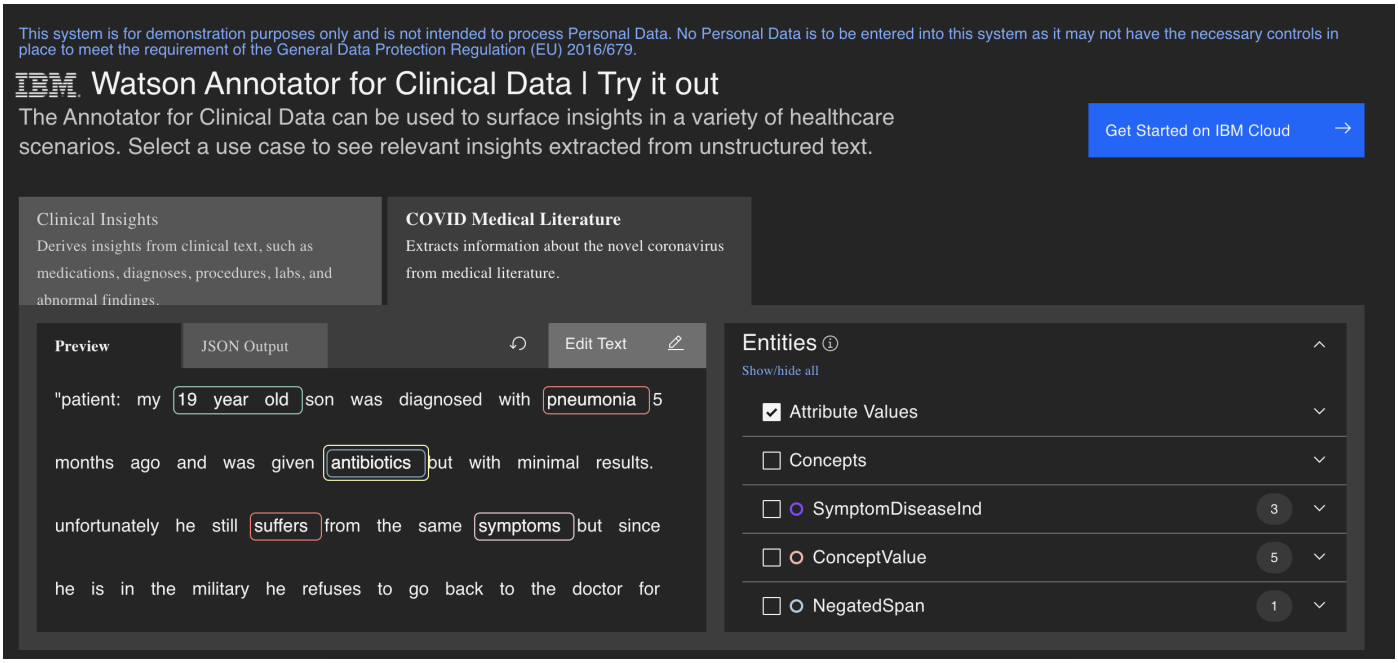
1. the URL to the QA page,
2. a query from a patient,
3. response from JA (the chatbot of the website),
4. response from an online doctor(human).

Existing Work

1. NER

https://cloud.ibm.com/catalog?category=ai#services (https://cloud.ibm.com/catalog?category=ai#services)

The IBM provide a cloud service called "Annotator for Clinical Data" which can help us to do NER. Below is a result of entities recognition using one of our conversation data



2. Healthcare chatbot demo

https://github.com/RasaHQ/medicare_locator (https://github.com/RasaHQ/medicare_locator)

Medicare Locator is an open source starter pack based on Rasa framework. It is a rule-based chatbot framework.

The dataset used in this project comes from: https://data.cms.gov/provider-data/ (https://data.cms.gov/provider-data/) which unfortunately not include doctor-patient conversation.

The work helps people search for a hospital, nursing home or home health agency in a US city. It also hands basic chitchat like greeting and saying goodbye.

Following is a demo of hospital-locator chatbot.

```
Bot loaded. Type a message and press enter (use '/stop' to exit):
Your input ->  Hey!
I'm good, thanks!
127.0.0.1 - - [2021-04-11 15:28:24] "POST /webhooks/rest/webhook?stream=true&token= HTTP/1.1" 200 185 0.155521
Your input ->  I'm in Sitka, I need a hospital
What is your current city?
127.0.0.1 - - [2021-04-11 15:28:36] "POST /webhooks/rest/webhook?stream=true&token= HTTP/1.1" 200 194 0.027149
Your input ->  Sitka
Here is a hospital near you:
Buttons:
1: Mt Edgecumbe Hospital (/inform{"facility_id":"021314"})
127.0.0.1 - - [2021-04-11 15:28:44] "POST /webhooks/rest/webhook?stream=true&token= HTTP/1.1" 200 295 0.699193
Your input ->  Bye
Have a good day.
127.0.0.1 - - [2021-04-11 15:28:51] "POST /webhooks/rest/webhook?stream=true&token= HTTP/1.1" 200 184 0.019094
Your input ->  /stop
```

Conclusion

Next week I will replace some diagnosis data on this framework and see the results, continue to study deep learning basic knowledge and do literature review. Here's my plan for the first version of the chatbot:

https://lucid.app/lucidchart/4b278d37-0f16-4f71-9d91-3fed085af4d7/edit?shared=true&page=uDe-dIt-NWfS#
(https://lucid.app/lucidchart/4b278d37-0f16-4f71-9d91-3fed085af4d7/edit?shared=true&page=uDe-dIt-NWfS)