

# Webcrawler

## Creation of the raw text files:

From the starting URLs, it extracts all URLs. It then randomly selects n extracted URLs and add them to the queue. There does exist a mechanism so that some manual URLs never be added. This mechanism also prevent future URLs from being from the same domain as the previous domains. It then pops the first URL from the queue and gets all information from <p> tags and stores them locally. It repeats this until there are 25 urls that was attempted to be scraped (if the scraper failed to scraped for any reason, it will be ignored).

## Creation of the clean text files:

From the clean text, re.sub is used to find patterns and clean those patterns, either by substitution or elimination.

## Creation of knowledge base:

After finding important terms from a single corpus is created with all the cleaned up text. The knowledge base is a dictionary with key:value pairs, with the terms as the keys, and a list of sentences containing that term as the value.

## Important Terms:

1. Stardew Valley
2. ChuckleFish
3. Eric Barone
4. ConcernedApe
5. Mobile
6. Music
7. Multiplayer
8. Album
9. Android
10. Version
11. iOS
12. Update

## Raw text vs Cleaned text

known professionally as ConcernedApe.[6][7][8]

*Snippet from en-wikipedia-raw-2.txt*

known professionally as ConcernedApe.

*Snippet from en-wikipedia-clean-2.txt*

This is an example what was cleaned up from the raw file.

## Knowledge Base Screenshot 1:

```
# Update infomation
for sentence in sentences:
    if check_word_in_sentence("stardew", sentence) or check_word_in_sentence("valley", sentence):
        stardew_valley_info.append(sentence)

    if check_word_in_sentence("ChuckleFish", sentence):
        ChuckleFish_info.append(sentence)

    if check_word_in_sentence("Eric", sentence) or check_word_in_sentence("Barone", sentence):
        Eric_Barone_info.append(sentence)

    if check_word_in_sentence("ConcernedApe", sentence):
        ConcernedApe_info.append(sentence)

    if check_word_in_sentence("mobile", sentence):
        mobile_info.append(sentence)

    if check_word_in_sentence("music", sentence):
        music_info.append(sentence)
```

This is a small snippet of code that looks at each sentence of the combined clean files. For each sentence, it checks if that sentence contains a certain key word(s). If it does it appends that sentence to a list of sentences. Note, this screenshot doesn't show all the terms for compactness reasons, the actual code does this for all import terms.

## Knowledge Base Screenshot 2:

```
Key: Stardew Valley
Value: ['The official Stardew Valley Piano Collections album!', 'He's best know for creating the indie farming RPG, Stardew Valley.', '"Stardew Valley came into my life when I was struggl']
Key: ChuckleFish
Value: ['That's quite astonishing, but according to publisher Chucklefish Games, it's definitely true.', 'Stardew Valley has sold over a million copies, publisher Chucklefish Games conf']
Key: Eric Barone
Value: ['Music composed by Eric Barone (ConcernedApe).', 'Eric is a solo game developer based in Seattle.', 'I also owe thanks to Eric Barone (ConcernedApe) for giving us Stardew Valley']
Key: ConcernedApe
Value: ['-Meadow Bridgham \xa0 Music composed by Eric Barone (ConcernedApe), 2016 ConcernedApe LLC.', 'Stardew Valley is a trademark or registered trademark of ConcernedApe LLC.', 'Fir']
Key: mobile
Value: ['As to mobile multiplayer, there\'s still currently no guarantee that it will happen... it may end up being possible, but I\'m not comfortable saying "yes, it will happen" at th']
Key: music
Value: ['Music composed by Eric Barone (ConcernedApe).', 'The album was envisioned by arranger Meadow Bridgham, Meadow is in the post-residential phase of their Doctorate of Musical Art']
Key: multiplayer
Value: ['In an update on Twitter, ConcernedApe clarified that the Nintendo Switch will be the first console to have cooperative multiplayer functionality.', 'It will offer 12 campaigns']
Key: album
Value: ['The album was envisioned by arranger Meadow Bridgham, Meadow is in the post-residential phase of their Doctorate of Musical Arts degree in Music Composition at Yale and lives i']
Key: android
Value: ['Things escalate pretty quickly from planting carrots... Already available on PS4, PC, Xbox One, Switch and PS Vita, Stardew Valley is now available on iOS devices, with an Androi']
Key: version
Value: ['Things escalate pretty quickly from planting carrots... Already available on PS4, PC, Xbox One, Switch and PS Vita, Stardew Valley is now available on iOS devices, with an Androi']
Key: iOS
Value: ['Things escalate pretty quickly from planting carrots... Already available on PS4, PC, Xbox One, Switch and PS Vita, Stardew Valley is now available on iOS devices, with an Androi']
Key: update
Value: ['Stardew Valley's 1.5 update released late last year and was the game's biggest yet, adding the Ginger Island location and the Beach Farm layout.', 'Since then, players have wor']
```

This shows the first couple of sentences for each Key:Value pair.

# Chatbot

## System Description:

There are two main feature that the chatbot utilizes to chat with the user.

The first feature is that the chatbot asks a question, then the user gives a response. If the chatbot cannot process that response, it asks the user for another response. The chatbot then processes that response to a question, stores corresponding information about that user.

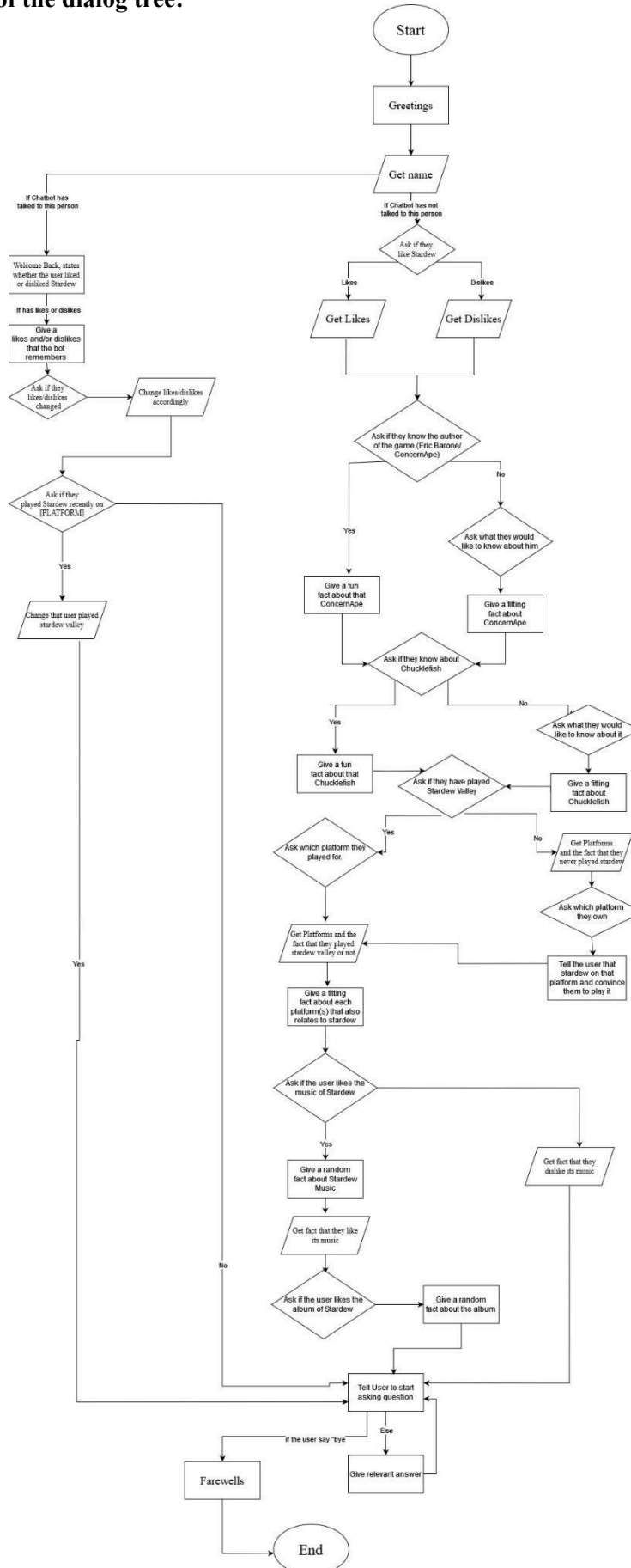
The second feature is that the chatbot asks the user for a question. Similarly, if the chatbot cannot process that response, it asks the user for another response. Then, the chatbot processes that question and using some NLP techniques, which will be elaborated on later, gives an appropriate response for that question.

Another feature is the user model feature. If the chatbot has talked to the user before, a different set of questions will be asked and the information pertaining to that user will be updated during the chat. If the chatbot has not talked to the user before, it will store relevant information during the chat.

## NLP Techniques:

- When the bot is asking for the name of the user, spaCy and NER is utilized to get the entity label to get the name of the user.
  - SpaCy/NER is also utilized to get the specific likes/dislikes of the user when the chatbot asks.
- From VaderSentiment, `SentimentIntensityAnalyzer()` is utilized to get the sentiment of the user. This is used to get the likes/dislikes of the user pertaining to a particular topic of Stardew Valley.
- Cosine similarity is used to match the best sentence from the knowledge base to each response from the user. Its also used to match the best term to a response.
- NLTK is used to get rid of stopwords, non alpha word, and various different filtering mechanisms
- While not NLP specific, Python String manipulation is also used for various things such as answering “yes/no” questions and consoles from the user.

## Diagram of the dialog tree:



## Sample Dialog 1:

Haley-Abigail: Hey there! I'm Haley-Abigail, a fan-bot of Stardew Valley. I love to talk about Stardew Valley! What is your name?  
You: My name is John Oliver  
Haley-Abigail: Hi John Oliver. Do you like Stardew Valley?  
John Oliver: Yes I love it!  
Haley-Abigail: I love Stardew Valley as well!  
Haley-Abigail: What do you like about Stardew Valley?  
John Oliver: I like the music and atmosphere  
Haley-Abigail: You like its music and atmosphere as well? Me too! Honestly, I love everything about Stardew Valley.  
Haley-Abigail: Do you know who ConcernedApe, otherwise known as Eric Barone, is?  
John Oliver: Yes  
Haley-Abigail: Yea! He's the sole creator of Stardew Valley! Here's a random fact about him:  
Haley-Abigail: After the game was shown a great deal of support from the community, Barone began working on it in full, engaging with Reddit and Twitter communities to discuss his progress and gain feedback on proposed  
Haley-Abigail: Do you know about ChuckleFish?  
John Oliver: Nope  
Haley-Abigail: It was the original publisher of Stardew Valley.  
Haley-Abigail: What would you like to know about ChuckleFish.  
John Oliver: What other games did they publish?  
Haley-Abigail: That's quite astonishing, but according to publisher Chucklefish Games, it's definitely true.  
Haley-Abigail: Have you ever played Stardew Valley?  
John Oliver: Yes. Of course!  
Haley-Abigail: Which console did you play it on?  
John Oliver: I played it on PC and the Nintendo Switch  
Haley-Abigail: You played on the Nintendo Switch and PC platform(s).  
Haley-Abigail: Have you ever played Stardew Valley on multiplayer?  
John Oliver: Yes, i dabbled in it.  
Haley-Abigail: Here's a random fact about the multiplayer:  
Haley-Abigail: To clear up concerns about multiplayer on consoles, Barone notes that mobile development "has had no impact" on the timeline "as they are being worked on by two entirely separate teams/companies."  
John Oliver: Of course!  
Haley-Abigail: Sorry about that, I couldn't get your response.  
Haley-Abigail: Do you like the music of Stardew Valley?  
John Oliver: Yes  
Haley-Abigail: Here's a random fact about its music:  
Haley-Abigail: Inspired by this game I had come to love, I began arranging the music, which helped me push through stress-induced writer's block and even begin to focus on writing original music again.  
Haley-Abigail: Did you know that Stardew Valley also has an album?  
John Oliver: Yes  
Haley-Abigail: An official sheet music book for piano and album was released in 2018 by Label Materia Collective.

Haley-Abigail: I have ran out of questions for you.  
Haley-Abigail: Now it's your turn to ask!  
Haley-Abigail: These are the following terms I know alot about:  
Haley-Abigail: Type in 'bye' to end this conversation  
\* Stardew Valley  
\* ChuckleFish  
\* Eric Barone  
\* ConcernedApe  
\* mobile  
\* music  
\* multiplayer  
\* album  
\* android  
\* version  
\* iOS  
\* update  
Haley-Abigail: What would you like to know about one of the terms?  
John Oliver: How is the iOS port?  
Haley-Abigail: Now, Barone is the publisher of Stardew Valley on all platforms, save for iOS and Android.  
Haley-Abigail: What would you like to know about one of the terms?  
John Oliver: how are the mobile controls?  
Haley-Abigail: Around 41% of those who purchased the game on mobile were based in the US.  
Haley-Abigail: What would you like to know about one of the terms?  
John Oliver: bye  
Haley-Abigail: Goodbye.

## Sample Dialog 2:

```
Halley-Abigail: Hi! I'm Halley-Abigail, a fan-bot of Stardew Valley. I love to talk about Stardew Valley! What is your name?
You: My name is Bob Jones
Halley-Abigail: Hello Bob Jones. Do you like Stardew Valley?
Bob Jones: Not really. I never played it in my life.
Halley-Abigail: Really? That's too bad. Why don't you like Stardew Valley?
Bob Jones: I don't like the art
Halley-Abigail: I see... You dislike its art.
Halley-Abigail: Do you know who ConcernedApe, otherwise known as Eric Barone, is?
Bob Jones: Nope
Halley-Abigail: He's the sole creator of Stardew Valley!
Halley-Abigail: What would you like to know about Eric Barone.
Bob Jones: What city is he from?
Halley-Abigail: Eric is a solo game developer based in Seattle.
Halley-Abigail: Do you know about ChuckleFish?
Bob Jones: Nope
Halley-Abigail: It was the original publisher of Stardew Valley.
Halley-Abigail: What would you like to know about ChuckleFish.
Bob Jones: Are they still the developer of Stardew Valley?
Halley-Abigail: Indie publisher Chucklefish has proudly announced the launch of agricultural life sim Stardew Valley on iOS.
Halley-Abigail: Have you ever played Stardew Valley?
Bob Jones: Nope
Halley-Abigail: You should play it!
Halley-Abigail: Stardew Valley is on many platforms. Do you want to know which ones?
Bob Jones: No, which ones?
Halley-Abigail: Stardew Valley is available on iOS, Android, Windows, macOS, Linux, PlayStation 4, Xbox One, Nintendo Switch, Playstation Vita, and PC.
Halley-Abigail: Do you own one of these platforms?
Bob Jones: Yes
Halley-Abigail: Which platforms do you own?
Bob Jones: I have a PC, iOS device, and a Nintendo Switch.
Halley-Abigail: You should play Stardew Valley on iOS or Nintendo Switch or PC.
Halley-Abigail: Do you like the music of Stardew Valley?
Bob Jones: Nope, I never heard of it.
Halley-Abigail: You should give it another chance!
Halley-Abigail: They even have a physical vinyl album!
```

```
Halley-Abigail: I have ran out of questions for you.
Halley-Abigail: Now it's your turn to ask!
Halley-Abigail: These are the following terms I know alot about:
Halley-Abigail: Type in 'bye' to end this conversation
  * Stardew Valley
  * ChuckleFish
  * Eric Barone
  * ConcernedApe
  * mobile
  * music
  * multiplayer
  * album
  * android
  * version
  * iOS
  * update
Halley-Abigail: What would you like to know about one of the terms?
Bob Jones: What is the latest update?
Halley-Abigail: That being said, the update is very close to being ready... the core multiplayer functionality is entirely operational, but the update
Halley-Abigail: What would you like to know about one of the terms?
Bob Jones: bye
```

## Google Survey:

<https://forms.gle/zSfR4s9FuYgKLDb9>

# Stardew Valley Chatbot Survey

Sign in to Google to save your progress. [Learn more](#)

\* Indicates required question

How was your overall experience with the chatbot? \*

	1	2	3	4	5	
Very unfun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very fun

Did you like the length of the chatbot? \*

	1	2	3	4	5	
Too short	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too long

How well did the chatbot respond to your questions/answer? \*

	1	2	3	4	5	
Poorly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too long

Any other criticism?

Your answer

Response 1 (a friend who has no programming knowledge):

How was your overall experience with the chatbot? \*

	1	2	3	4	5	
Very unfun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Very fun

Did you like the length of the chatbot? \*

	1	2	3	4	5	
Too short	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Too long

How well did the chatbot respond to your questions/answer? \*

	1	2	3	4	5	
Poorly	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Well

Any other criticism?

None, it was fun

Haley-Abigail: Stardew Valley is on many platforms. Do you want to know which ones?

Bob: yes

Haley-Abigail: Do you own one of these platforms?

- Fixed this grammatical mistake



Response 2 (a friend who has no programming knowledge):

How was your overall experience with the chatbot? *						
	1	2	3	4	5	
Very unfun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Very fun

Did you like the length of the chatbot? *						
	1	2	3	4	5	
Too short	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Too long

How well did the chatbot respond to your questions/answer? *						
	1	2	3	4	5	
Poorly	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Well

Any other criticism?						
my name was not recognized						

- Fixed the function to better recognize some names

Response 3 (a friend who has some programming knowledge):

How was your overall experience with the chatbot? \*

	1	2	3	4	5	
Very unfun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Very fun

Did you like the length of the chatbot? \*

	1	2	3	4	5	
Too short	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Too long

How well did the chatbot respond to your questions/answer? \*

	1	2	3	4	5	
Poorly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Well

---

Haley-Abigail: Pleased to meet you! I'm Haley-Abigail, a fan-bot of Starde

You: Hi Haley-Abigail

Haley-Abigail: Hello Haley-Abigail. Do you like Stardew Valley?

Haley-Abigail: I'm not Haley-Abigail

- Fixed this case when they greet the chatbot

Haley-Abigail: Sorry about that, I couldn't get your dislikes. Can you repeat your dislikes please?

Haley-Abigail: I like turtles

Haley-Abigail: I see... You dislike its turtles.

- Fix the case where the user gives a like if the user gives a dislike and vice versa

### Advantages:

- Easily scalable, if I needed to add more cases, I could easily do so.
- The chatbot remember users and gives information from its response.
- It's fairly lengthy and keeps a conversation for a while.

### Disadvantages:

- My chatbot doesn't process Non-Sentence (one word response that well)
  - Ex: This response doesn't know the name of the user is Apple until a sentence is formed

```
Haley-Abigail: Greetings! I'm Haley-Abigail, a fan-bot of Stardew Valley. I love to talk about
You: Apple
Haley-Abigail: Sorry about that, I couldn't get your name. Can you repeat your name please?
You: My name is Apple
Haley-Abigail: Hi Apple. Do you like Stardew Valley?
```

- The rules for everything is very easily breakable
  - Ex. This doesn't make grammatical sense, from both the user and chatbot
  - It also doesn't extract the right information.

```
Haley-Abigail: Really? That's too bad. Why don't you like Stardew Valley?
Apple: I dislike the art and the fact that its done
Haley-Abigail: I see... You dislike its art and fact.
```

- The responses of the chatbot makes no sense most of the time. This could be improved with a larger corpus/ML.
- The programming quality of this chatbot is very poor. Some functions could be created to make the code more manageable and more readable.

### Appendix A: An example of a sample user model: “John Smith”

```
Haley-Abigail: Pleased to meet you! I'm Haley-Abigail, a fan-bot of Stardew Valley. I love to talk about Stardew Valley! What is your name?
You: John Smith
Haley-Abigail: Welcome back John Smith!
Haley-Abigail: From our last conversation, I remembered that you liked Stardew Valley.
Haley-Abigail: When we talked last, I remembered that you liked the characters and music and art and atmosphere about Stardew Valley.
Haley-Abigail: Are there any additional likes about Stardew Valley since our last conversation John Smith?
John Smith: I like the community
Haley-Abigail: You additionally like community about Stardew Valley.
Haley-Abigail: I also remembered that you disliked the price about Stardew Valley.
Haley-Abigail: Are there any additional dislikes about Stardew Valley since our last conversation John Smith?
John Smith: Nope.
Haley-Abigail: Have you played Stardew Valley on PC since the last time we talked John Smith?
John Smith: I have played it. Yes
```

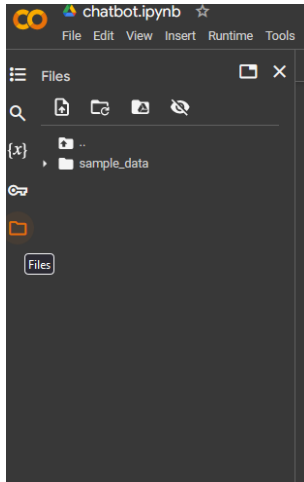
### Appendix B: Knowledge base

```
Stardew Valley 125
ChuckleFish 11
Eric Barone 72
ConcernedApe 8
mobile 8
music 12
multiplayer 15
album 8
android 6
version 10
iOS 11
update 14
```

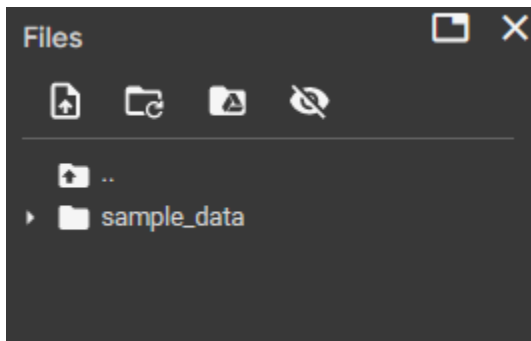
Where “Stardew Valley” is a term and “125” is the number of sentences with the term “Stardew Valley”.

## Appendix C: How to access the Chatbot via Google Colab:

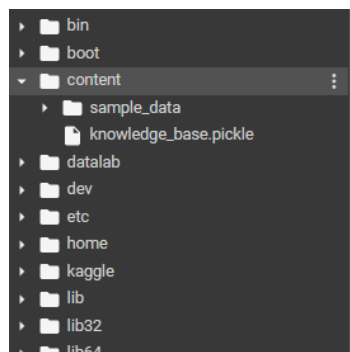
1. Access the chatbot.IPYNB on Google Colab
  - [https://colab.research.google.com/drive/15d8JcjMKWMmWgN\\_aNFULO9PUUIrFrgI2?usp=sharing](https://colab.research.google.com/drive/15d8JcjMKWMmWgN_aNFULO9PUUIrFrgI2?usp=sharing)
2. Click on the folder on the left panel



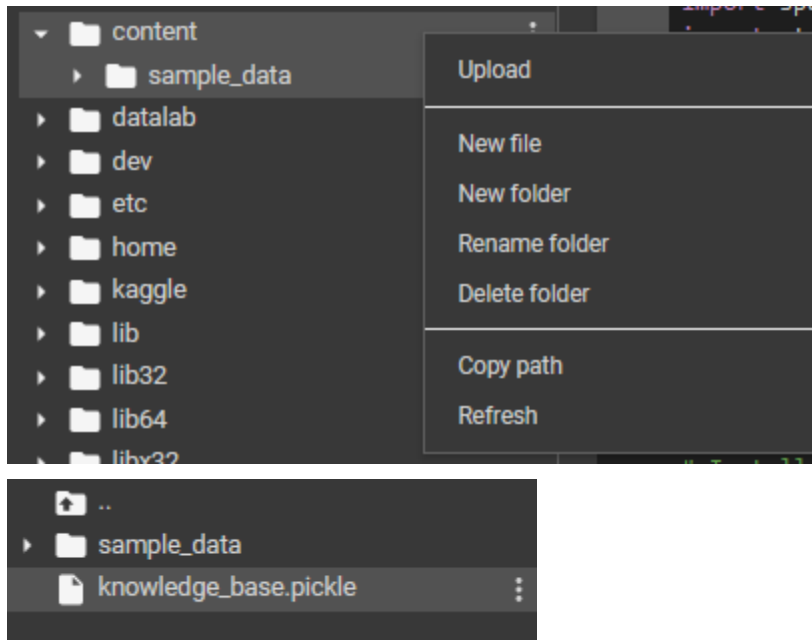
3. Sign in with your personal Google account if prompted
4. You should see something similar to this.



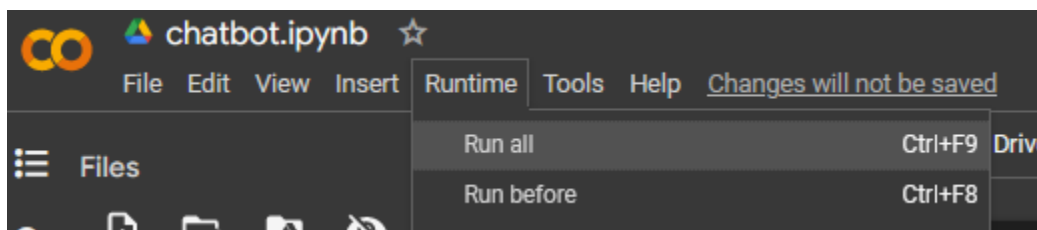
- If not click on content folder.



5. Right click and click upload. Upload “*knowledge\_base.pickle*” from the uploaded files.



6. Click Runtime -> Run All



7. Talk to the bot!
  - (Note: to access the previous user models upload the folder “user\_models”)
  - (Note 2: Can still run the chatbot.py as well, since its equivalent)

## Appendix D: readme.MD contents

```

① readme.MD
1  Part one: webscraper.IPYNB and extract_important_terms.IPYNB
2  Part two: chatbot.IPYNB (or chatbot.py)
3
4  Dependancies:
5  pip install nltk
6  pip install -q spacy
7  python -m spacy download en_core_web_sm
8  pip3 install vaderSentiment
9  pip install -U scikit-learn
10
11  Run "python chatbot.py" to run the chatbot.

```