# University chat bot project proposal

**By Dexter R Shepherd**

# Contents

## Overview

A problem for students is having too much information and recourses thrown at them, and not knowing who to turn to for information. This results in university staff having to check their emails all the time and answer questions over and over again. Then the same events happen with the next year group. This also means students need to wait for a response. Would it not be better if these questions and answers were written in a database and kept updated by university staff? This would allow students to get instant responses, and keep university staff emails free for questions which can't be answered.

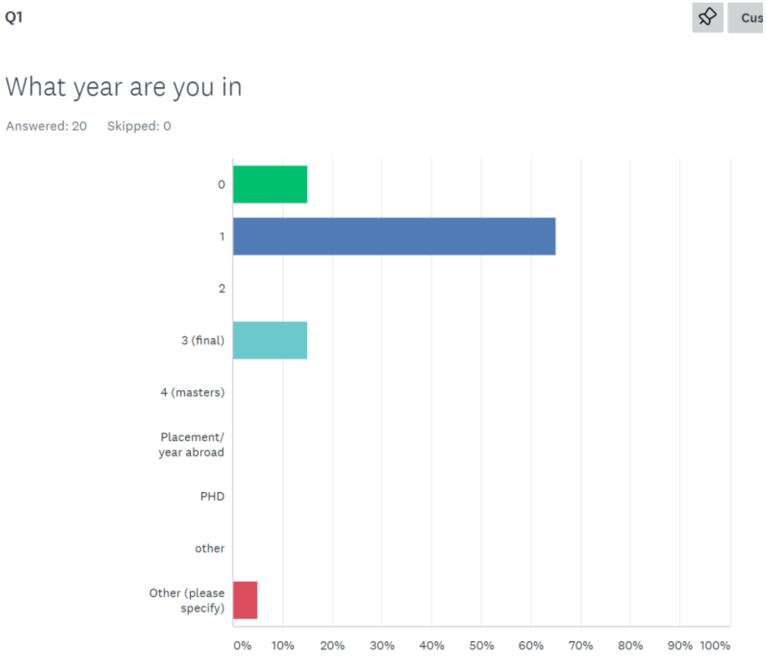
This project is for a chatbot to help students with university life. This can be questions from mental health problems, disability support all the way to where are cheap places to shop. When there is an unknown response, this gets added to a list of questions the university can keep updated with responses. This project will help students manage day to day university life, as the change from living with parents to being independent can be a struggle. It is also helps university staff as they will receive less emails.

The theory of this project could be rolled out in other applications, indeed many companies use this approach for their own services [1].

## Market Research

There are many systems like this currently out there. I looked into the Spotify chatbot which helps users find out information via a friendly non-human interfaced. I also looked into CleverBot which is a self-learning AI. It learns off of how others interact. I wouldn't be going that far with this version of the software.

I deployed a survey to find out how important the issue I am solving is, and whether people think my solution is effective and useful.



Q1

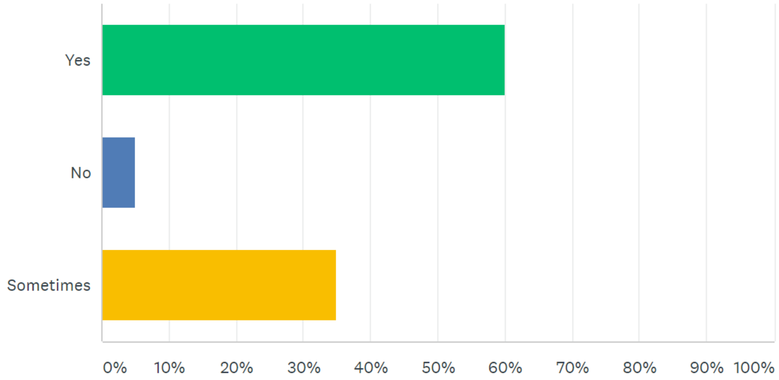
What year are you in

Answered: 20 Skipped: 0

The other was a medical student in their 6th year. My data was gathered from students, mainly first years.

## Would you say that you always know where to go to find information about university life, your course information, next steps etc...

Answered: 20 Skipped: 0



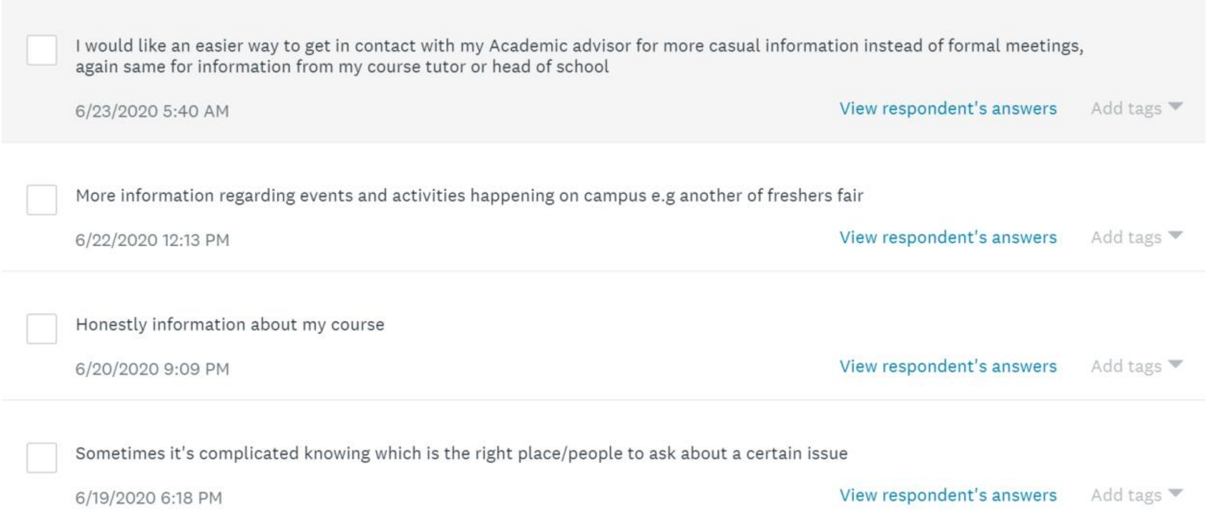
40% of students do not know or not always know where to go for information. When looking at individual trends it seemed that the higher the year you were, the more likely you were to know information.

### Q3

## If No or Sometimes, what information would you not know?

Answered: 6 Skipped: 14

The 6 responses were as followed:



□ I would like an easier way to get in contact with my Academic advisor for more casual information instead of formal meetings, again same for information from my course tutor or head of school

6/23/2020 5:40 AM

□ More information regarding events and activities happening on campus e.g another of freshers fair

6/22/2020 12:13 PM

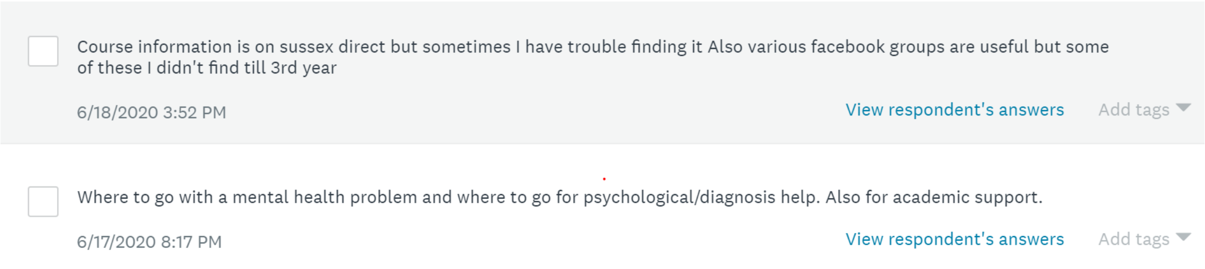
□ Honestly information about my course

6/20/2020 9:09 PM

□ Sometimes it's complicated knowing which is the right place/people to ask about a certain issue

6/19/2020 6:18 PM

View respondent's answers
Add tags
View respondent's answers
Add tags
View respondent's answers
Add tags
View respondent's answers
Add tags



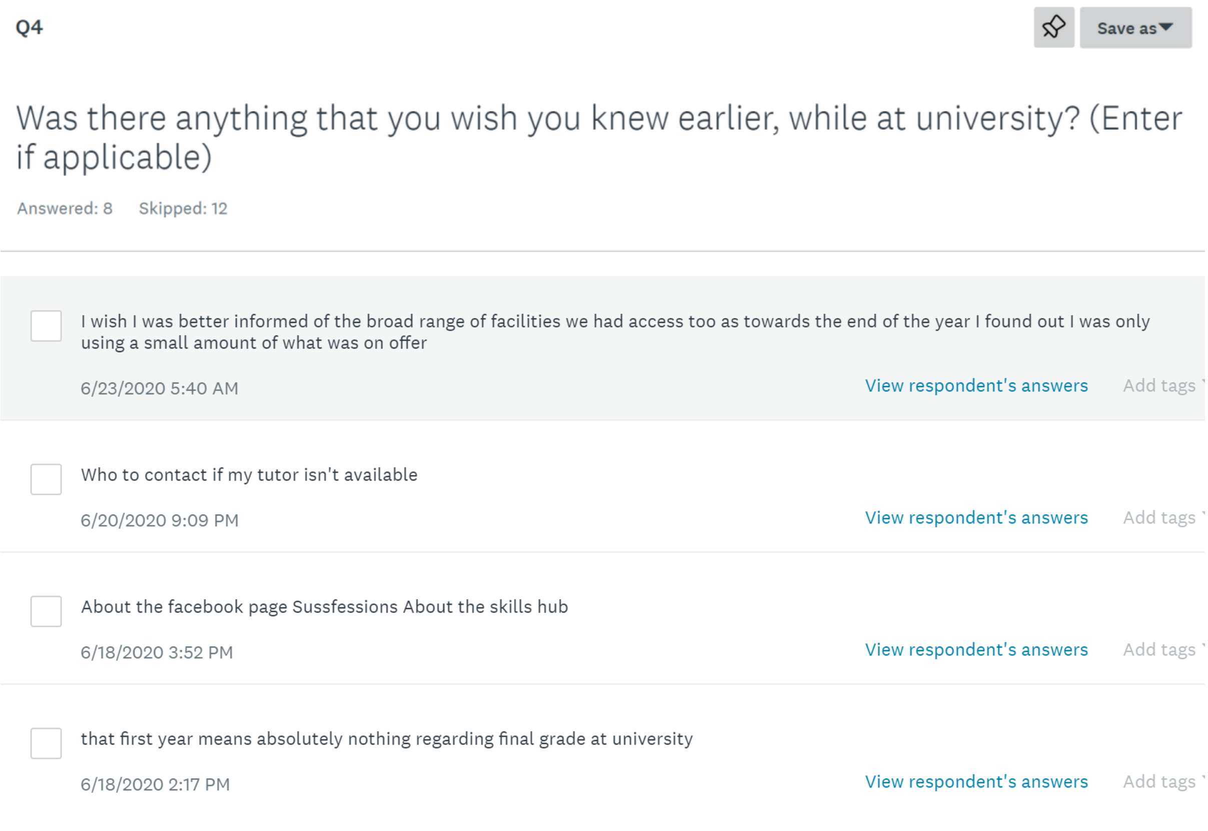
☐ Course information is on sussex direct but sometimes I have trouble finding it Also various facebook groups are useful but some of these I didn't find till 3rd year

6/18/2020 3:52 PM

☐ Where to go with a mental health problem and where to go for psychological/diagnosis help. Also for academic support.

6/17/2020 8:17 PM

The issues mainly seem to be mental health issues, where to go and what to do, sometimes involving courses. This application would aim to deliver this information quickly and accurately.



Q4

Was there anything that you wish you knew earlier, while at university? (Enter if applicable)

Answered: 8 Skipped: 12

☐ I wish I was better informed of the broad range of facilities we had access too as towards the end of the year I found out I was only using a small amount of what was on offer

6/23/2020 5:40 AM

☐ Who to contact if my tutor isn't available

6/20/2020 9:09 PM

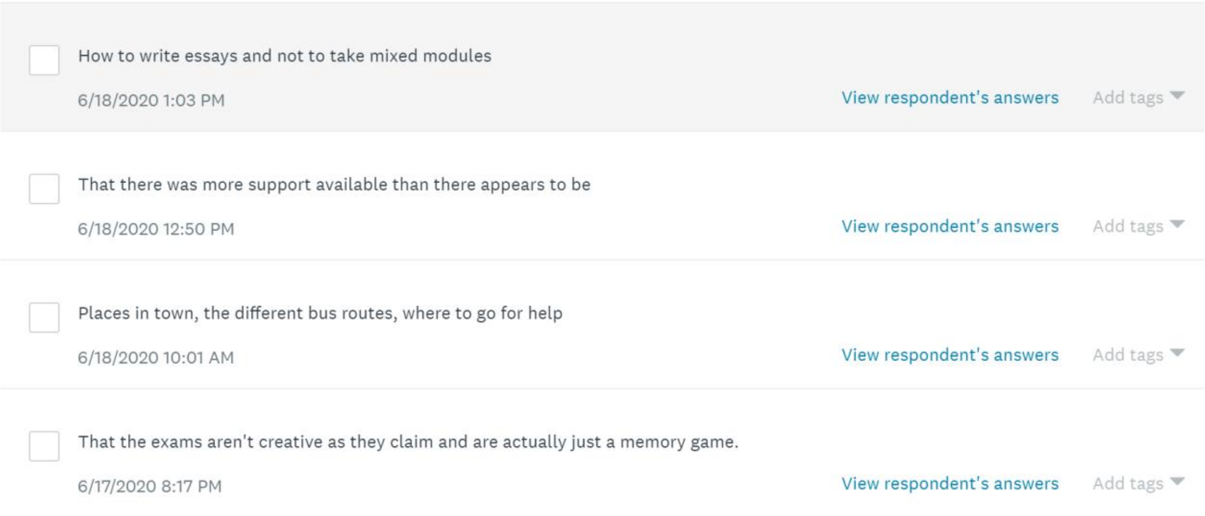
☐ About the facebook page Sussfessions About the skills hub

6/18/2020 3:52 PM

☐ that first year means absolutely nothing regarding final grade at university

6/18/2020 2:17 PM

Save as
View respondent's answers
Add tags
View respondent's answers
Add tags
View respondent's answers
Add tags
View respondent's answers
Add tags



☐ How to write essays and not to take mixed modules

6/18/2020 1:03 PM

☐ That there was more support available than there appears to be

6/18/2020 12:50 PM

☐ Places in town, the different bus routes, where to go for help

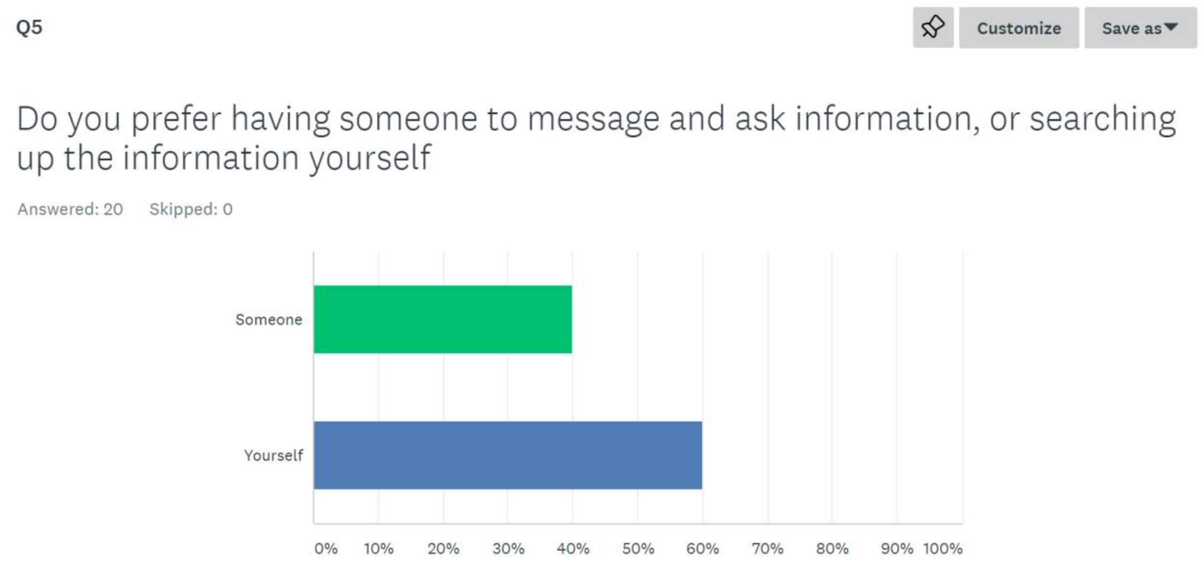
6/18/2020 10:01 AM

☐ That the exams aren't creative as they claim and are actually just a memory game.

6/17/2020 8:17 PM

View respondent's answers
Add tags
View respondent's answers
Add tags
View respondent's answers
Add tags
View respondent's answers
Add tags

The key issues are general information about university, both life and academic. The bot would be there to provide information in a friendly way, quickly and effectively. The answers need to be short but insightful, and could forward people using links to pages like Sussfessions.

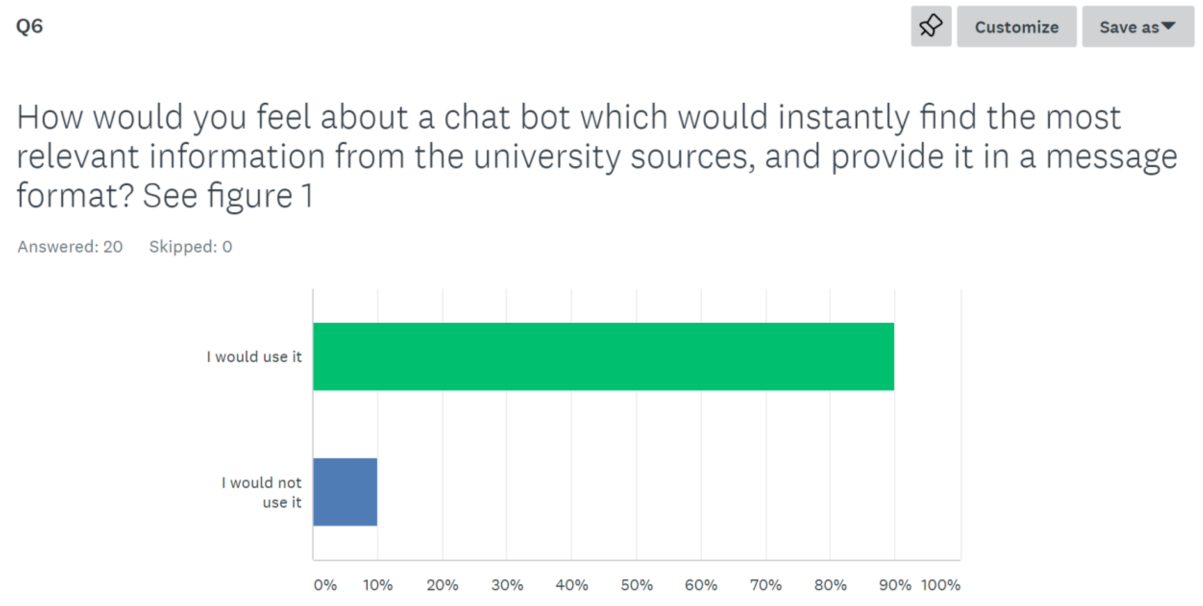


Q5

Do you prefer having someone to message and ask information, or searching up the information yourself

Answered: 20 Skipped: 0

After deploying question 5 I realized it was down to interpretation. Finding out information using a bot could both be considered talking to someone and finding out information yourself. As we see it is a 60% and 40% split it will not impact my research anyway.

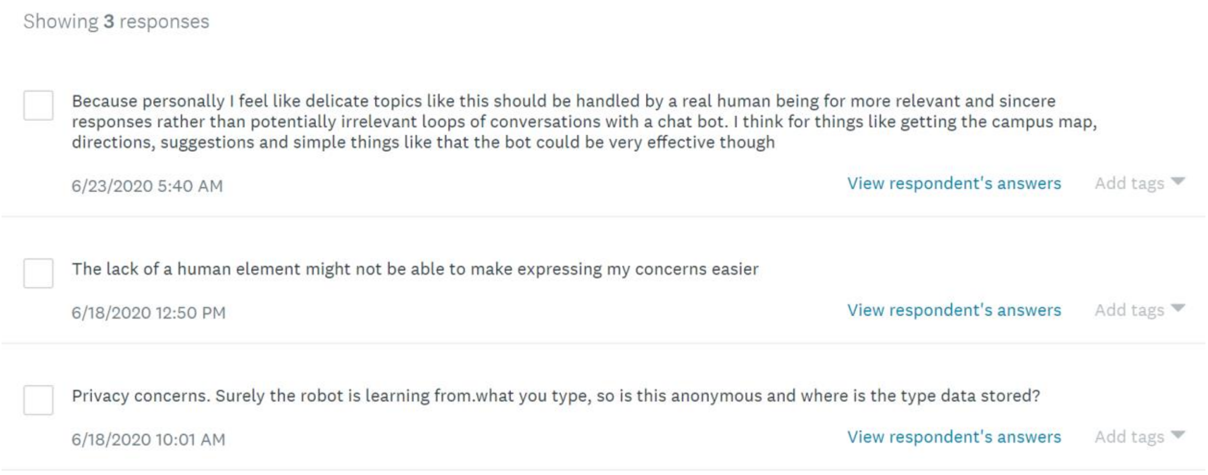


Q6

How would you feel about a chat bot which would instantly find the most relevant information from the university sources, and provide it in a message format? See figure 1

Answered: 20 Skipped: 0

We got 90% of people saying they would use this application out of the study of students. This is a strong result in favour of this application.



Showing **3** responses

□ Because personally I feel like delicate topics like this should be handled by a real human being for more relevant and sincere responses rather than potentially irrelevant loops of conversations with a chat bot. I think for things like getting the campus map, directions, suggestions and simple things like that the bot could be very effective though

6/23/2020 5:40 AM

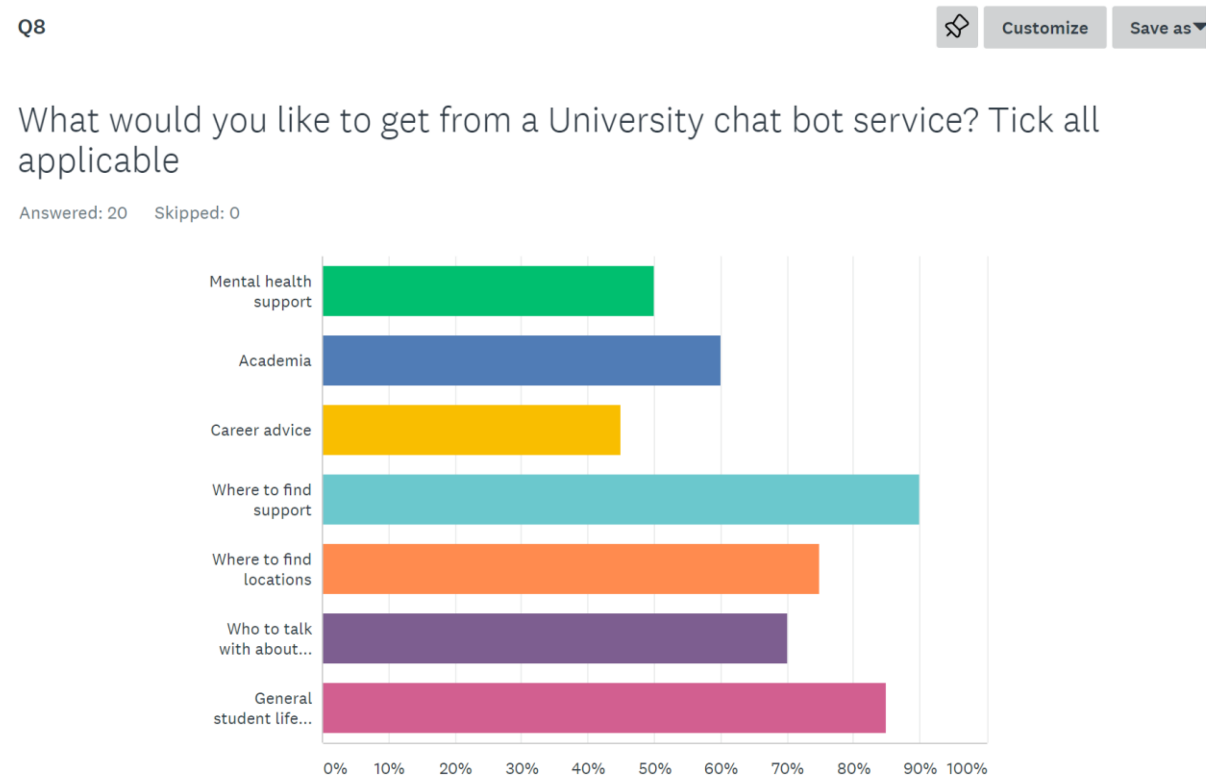
□ The lack of a human element might not be able to make expressing my concerns easier

6/18/2020 12:50 PM

□ Privacy concerns. Surely the robot is learning from.what you type, so is this anonymous and where is the type data stored?

6/18/2020 10:01 AM

The first response raises some good issues, hence the bot should only be here to provide guidance which may involve directing people to the student life centre for mental health support. He privacy concerns are not an issue at this stage, as data is managed in an input and output way. However, the element of self-learning and patching information together is not a bad idea. Perhaps even an emailing system which emails information it thinks you will like. Keeps students up to date with information they will like. Maybe even the use of cookies. This is just an idea for a future system.



Q8

What would you like to get from a University chat bot service? Tick all applicable

Answered: 20 Skipped: 0

Where to find information are popular choices with the results. The least popular are career advice and mental health support. This makes sense as the bot cannot truly feel in the way a human does hence the advice will be insincere. But for general information the bot is popular.

# Technicalities

## System requirements

The system would need constant running and a substantial amount of RAM to hold the data. If the Server uses Ubuntu, the RAM needs 1024MB of RAM for the OS [2], and for a big data base it would need up to 4GB of RAM. I estimate this would be fine. Multiple cores would be required for multiple processing of request and processing. If many people are to use the system at one time, potentially the server will need more RAM and need to monitor this. If the RAM gets too big, then the server-side code will need to alert people that the server is too busy currently. Being the design is efficient, this will not be a problem.

## Potential security threats

The main potential security threat would be trolls. Damage to the system wouldn't be inflicted, but the staff who monitor responses would get many spam responses if people kept asking "stupid" questions. This could be avoided through using Sussex emails to sign in and all responses are monitored this way, however this would then stop non students finding out information about the university. Another potential solution to trolls is to ban IP addresses on devices if they have spammed the system with inappropriate content. There does lie the underlying solution of just the monitoring staff deleting spam when it arrives. The AI does not learn naturally therefore University reputation remains untarnished if spammed with bad replies. It is down to the monitoring staff to add it for the rude responses to show up.

## Grammar and spelling

There will be potential issues with grammar and spelling. Universities within the UK are very diverse and English is not the first language of everyone. Being this application is accessed through mobiles, much of the spelling issues will be solved through auto correct, but  spelling and grammar checks could be implemented into the software to make sure everything is correct. The software can also take extra care when processing information, if it is only one word out then it could still go with that option.

## Language

The plan is to use Python due to the large amount of machine learning support and support for dynamic variables. Other languages like Java and C are not as helpful with non-fixed sized arrays. Python supports object oriented formats of code as well as event driven making it perfect for this project.

JavaScript will be used on the client side in order to communicate with the AI, which will be server side.

# Test plan

The following tests are designed to know when the system works at the different areas of development

## Bot code

| *Test No* | *Test* | *Expected outcome* |
|---|---|---|
| *1* | AI splits a paragraph down into manageable sentences and performs spell check. | An array of sentences made from the paragraph, with spelling mistakes corrected to a good degree. |

| | | |
|---|---|---|
| 2 | The AI splits a sentence into a statement or a question or neither | If neither nothing is returned. If a question it will try and find an answer. If a statement it will thank the user for their feedback. |
| 3 | A statement is saved | It appears within the statement folder in the json format. |
| 4 | A question which has been added will find a response. | The result set to that question is outputted |
| 5 | A question is not found but there are saved questions about that topic. | The system will add the question to a confused file, to be added by the admin at another point. Then will find something like that and state "I am not sure at the moment but here is something similar". |
| 6 | A question which has never been asked and is not entered | The system will add it to the confused file and apologise for it does not have an answer at this point. |
| 7 | Data already in the confused file is added | The data will not add it again but will increase its priority |
| 8 | A paragraph is entered with a vague sentence following another sentence. "what mental health services are there? Where are they" | The paragraph is split, and the previous topic remains the current topic to help find out what the second sentence means. "where are they" + "mental health services" |
| 9 | | |

## Client to Server and Server to Client

| *Test No* | *Test* | *Expected outcome* |
|---|---|---|
| 1 | The client connects to the server | A test code is sent and pinged back |
| 2 | The client sends the text input out to the server | A response is sent back based on the input |
| 3 | The client stores the subjects of the past and the response. | Local variables are shown in the browser console. |
| 4 | The client responds negative feedback to a message | The system receives information about what is wrong and it adds it to be added in the confused file. |

## Admin side

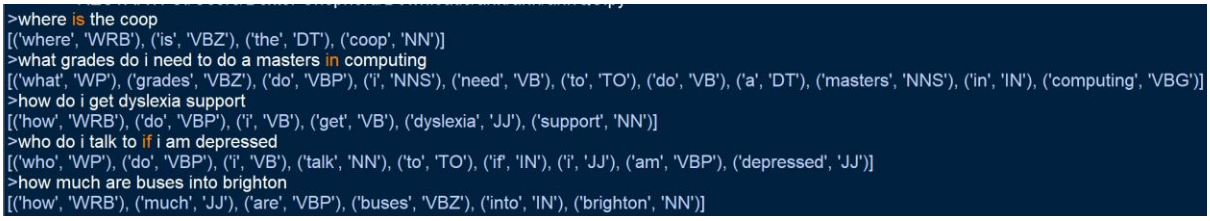
| *Test No* | *Test* | *Expected outcome* |
|---|---|---|
| 1 | Makes user sign in with correct credentials. | If the password and username is wrong then it returns wrong. If not then it lets the user in. |
| 2 | The user can receive the top amount of questions to add. | The user sends a request via the GUI and gets a response |

| | | |
|---|---|---|
| 3 | The user can answer questions. | The question and answer are submitted and trained into the system. |
| 4 | The user disconnects | Their IP is no longer saved in the server side code and they will have to re-enter their password and username. |
| 5 | The user can delete responses which are incorrect | The user selects delete and it no longer appears in the data. |

# Development

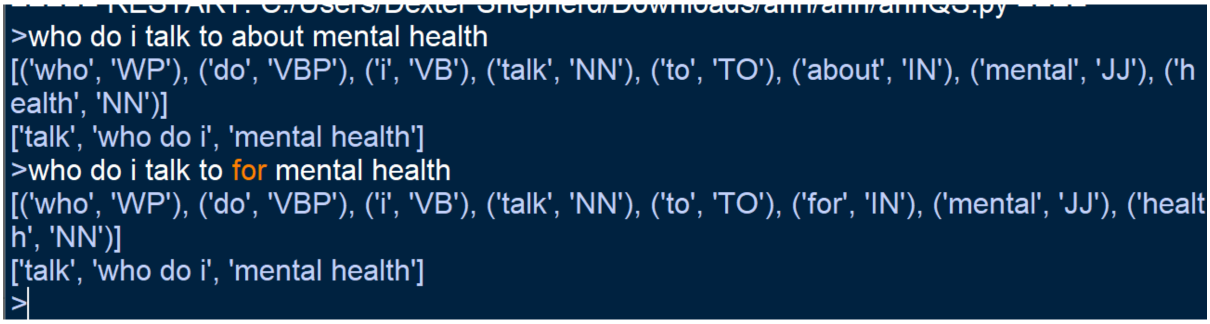
## Natural language processing

For the splitting of language I have two options: To make a complex graph data structure system to split sentences down to meaning, or use a python library which already processes information. I downloaded the NLTK library [3][4] which tokenizes words in order to develop meaning of sentences.



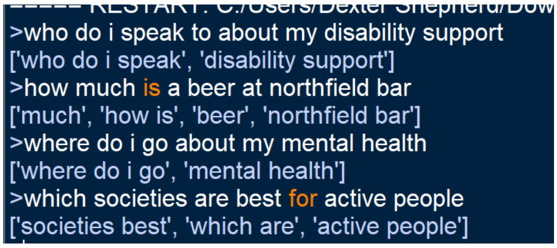
```
>where is the coop
[('where', 'WRB'), ('is', 'VBZ'), ('the', 'DT'), ('coop', 'NN')]
>what grades do i need to do a masters in computing
[('what', 'WP'), ('grades', 'VBZ'), ('do', 'VBP'), ('i', 'NNS'), ('need', 'VB'), ('to', 'TO'), ('do', 'VB'), ('a', 'DT'), ('masters', 'NNS'), ('in', 'IN'), ('computing', 'VBG')]
>how do i get dyslexia support
[('how', 'WRB'), ('do', 'VBP'), ('i', 'VB'), ('get', 'VB'), ('dyslexia', 'JJ'), ('support', 'NN')]
>who do i talk to if i am depressed
[('who', 'WP'), ('do', 'VBP'), ('i', 'VB'), ('talk', 'NN'), ('to', 'TO'), ('if', 'IN'), ('i', 'JJ'), ('am', 'VBP'), ('depressed', 'JJ')]
>how much are buses into brighton
[('how', 'WRB'), ('much', 'JJ'), ('are', 'VBP'), ('buses', 'VBZ'), ('into', 'IN'), ('brighton', 'NN')]
```

The above shows a screenshot of the library splitting information down into tokens. "NN" represents the subject. What the system will need to do is split up a sentence into relevant information. Using the NLTK library documentation I was able to decide upon the rules which build up significant nodes to the language. "where is the coop" would be "where is" "coop". Alternatively with a more complex sentence we would have "who do I talk to if I am depressed" which would break down to "who do" "I talk to" "I am depressed". The system will need to take the meaning, in the case of our second example is the person to talk to involving depression, and its being queried with "who". Taking a sentence like "who do I talk to about mental health", or "who do I go to for mental health" mean the same thing.



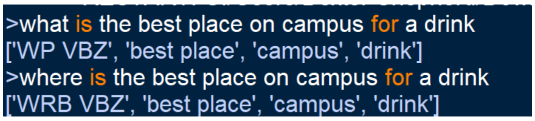
```
==== RESTART: C:/Users/Dexter Shepherd/Downloads/ann/ann/annQS.py ====
>who do i talk to about mental health
[('who', 'WP'), ('do', 'VBP'), ('i', 'VB'), ('talk', 'NN'), ('to', 'TO'), ('about', 'IN'), ('mental', 'JJ'), ('h
ealth', 'NN')]
['talk', 'who do i', 'mental health']
>who do i talk to for mental health
[('who', 'WP'), ('do', 'VBP'), ('i', 'VB'), ('talk', 'NN'), ('to', 'TO'), ('for', 'IN'), ('mental', 'JJ'), ('healt
h', 'NN')]
['talk', 'who do i', 'mental health']
>
```

Here we see the algorithm splitting down two different sentence filtering out meaning to its bare minimum. Below shows a series of sentences inputted (in white) and the node creation of the language class (in blue):



```
===== RESTART: C:/Users/Dexter Shepherd/Dow
>who do i speak to about my disability support
['who do i speak', 'disability support']
>how much is a beer at northfield bar
['much', 'how is', 'beer', 'northfield bar']
>where do i go about my mental health
['where do i go', 'mental health']
>which societies are best for active people
['societies best', 'which are', 'active people']
```

Each item in the array will represent a node of relevance relating to the answer. It will find the meaning based on the data it has been given. Using tokens instead of words will improve effectiveness when finding if two sentences which are different, mean the same thing.



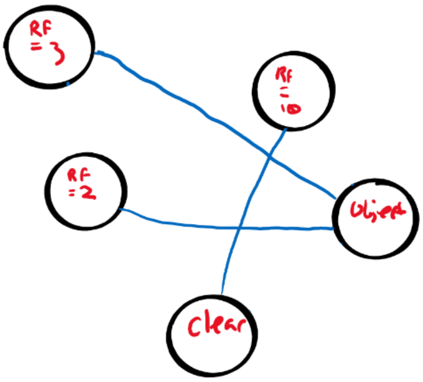
```
>what is the best place on campus for a drink
['WP VBZ', 'best place', 'campus', 'drink']
>where is the best place on campus for a drink
['WRB VBZ', 'best place', 'campus', 'drink']
```

The algorithm when searching will find that not all nodes are the same, but the only difference still has the tokenized similarity, therefore having a higher chance.

## Organization of data

### Data structures

The data will be organized into .json files and when opened, into graph data structures using dictionaries to link vertices with edges. Data will be entered in to a “confused” file if not found in the data, which will be the file that the university read when adding their own answers.

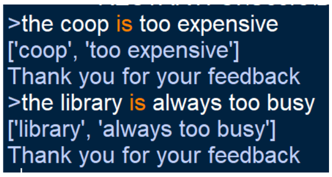


These nodes would hold different words and connect with a strength (which increases every time it happens) to either a question node or a statement. This allows the code to work out whether a sentence is a question. “what” in position 1 and “is” in position 2 is a common way to begin a sentence. “the” in position 1 and “is” in position 3 is a common way to state something about an object in position 2.

For the question file it will have nodes linking to a response. If all the nodes linking to this response are matched, or similar, then the response is likely to be the correct outcome. The system could find the most likely, but also state “this is how questions like this were answered” to alert the user that this is not the exact answer.

The statement file will be slightly different where the nodes will attach to each other rather than to any specific answer. This is to work out associations of information. “the shop is expensive” will need to convert down to “shop” and “expensive”. This will build up the more this happens. Then when this graph is processed it will show a strong connection between these two, alerting the university that most people find the shop too expensive.

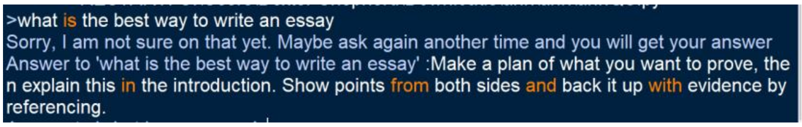
By splitting out the irrelevant nodes and only keeping subjects I am able to get information which is relevant.



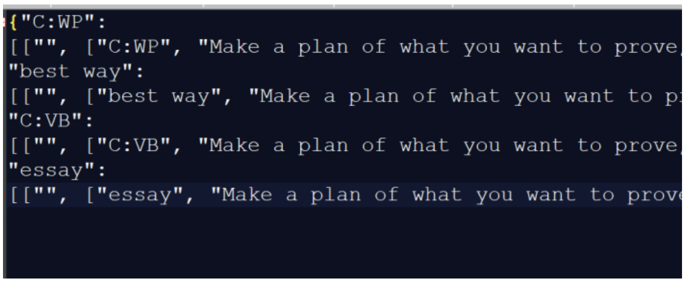
```
>the coop is too expensive
['coop', 'too expensive']
Thank you for your feedback
>the library is always too busy
['library', 'always too busy']
Thank you for your feedback
```

It associates the information which is relevant in the array seen, and the nodes would then all be connected together in the graph. It will then return the message “thank you for your feedback” as there is no other way.

I write some test code which would ask me to clarify each item in the confused file. It would then add this to the question graph and save it:

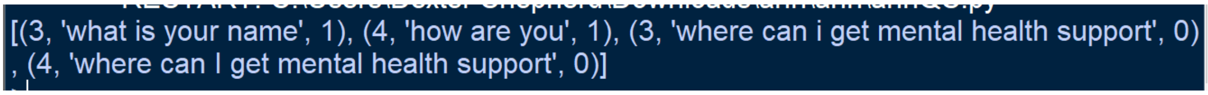


```
>what is the best way to write an essay
Sorry, I am not sure on that yet. Maybe ask again another time and you will get your answer
Answer to 'what is the best way to write an essay' :Make a plan of what you want to prove, the
n explain this in the introduction. Show points from both sides and back it up with evidence by
referencing.
```



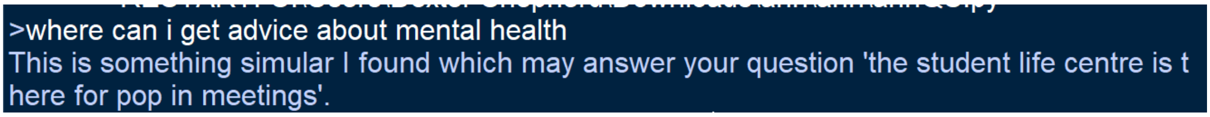
```
{"C:WP":
[["", ["C:WP", "Make a plan of what you want to prove,
"best way":
[["", ["best way", "Make a plan of what you want to p
"C:VB":
[["", ["C:VB", "Make a plan of what you want to prove,
"essay":
[["", ["essay", "Make a plan of what you want to prove
```

Within the data it saves all the nodes from the sentence pointing to the data. Using the graph class I currently made we see many redundancies in the data. This will slow up memory eventually, so I will need to review the graph code. At this time it works and that is the main importance. It was further removed from the confused data.



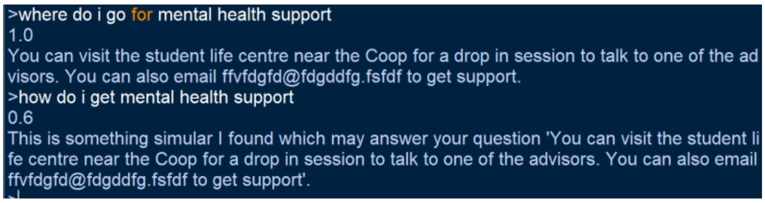
```
[(3, 'what is your name', 1), (4, 'how are you', 1), (3, 'where can i get mental health support', 0)
, (4, 'where can I get mental health support', 0)]
```

If I have taught the AI something but it is not in the data exactly, it will respond with a message like the following:



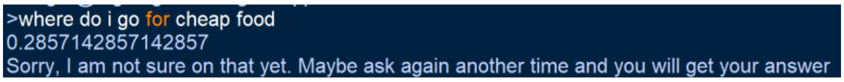
```
>where can i get advice about mental health
This is something simular I found which may answer your question 'the student life centre is t
here for pop in meetings'.
```

The data learned that “mental health” and “where” links strongly to the answer given. The extra bits of information made it so that it was not exact, but was close. I could allow the user to give feedback on whether that information is correct. If it is then it can adjust the data to link to it. If not then it can add it to the confused file. The data regarding the question and answer will need to be stored within the client bot and not the main bot.

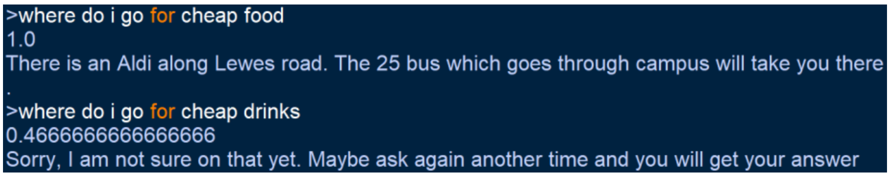


```
>where do i go for mental health support
1.0
You can visit the student life centre near the Coop for a drop in session to talk to one of the ad
visors. You can also email ffvfdgfd@fdgddfg.fsfdf to get support.
>how do i get mental health support
0.6
This is something simular I found which may answer your question 'You can visit the student li
fe centre near the Coop for a drop in session to talk to one of the advisors. You can also email
ffvfdgfd@fdgddfg.fsfdf to get support'.
```

Above shows the percentages as decimals on a sentence. I trained it on “where do I go for mental health support”. This is why the second question only has a probability of 0.6. This method prevents confusion with similar sentences.



```
>where do i go for cheap food
0.2857142857142857
Sorry, I am not sure on that yet. Maybe ask again another time and you will get your answer
```



```
>where do i go for cheap food
1.0
There is an Aldi along Lewes road. The 25 bus which goes through campus will take you there
.
>where do i go for cheap drinks
0.4666666666666666
Sorry, I am not sure on that yet. Maybe ask again another time and you will get your answer
```

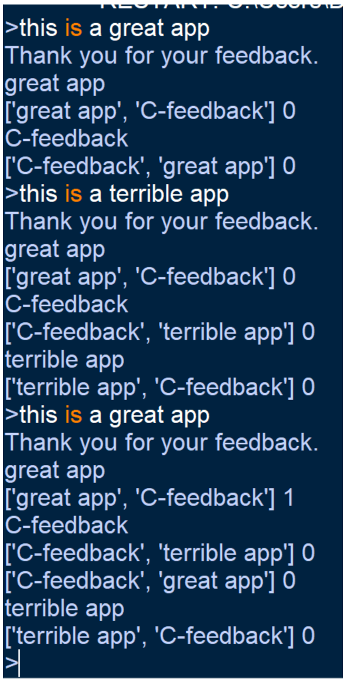
If two questions are the same but slightly different in subject, the system responds well by assigning a higher worth to subject words than to other words.

## Training in the data

The data for the question or statement analyses will be pre trained into the system. This will mean that the system will need a wide range of sentences to be trained on in order to be accurate.

The data for questions will be trained in by the university, and kept updated by the university. Question data can only be accessed for reading by the user, and accessed for writing to by the university. This keeps the integrity of the data.

Statements can be written to by the user. It records information so the university can learn from it. I have written code here to show the graph in text form every time something is added:

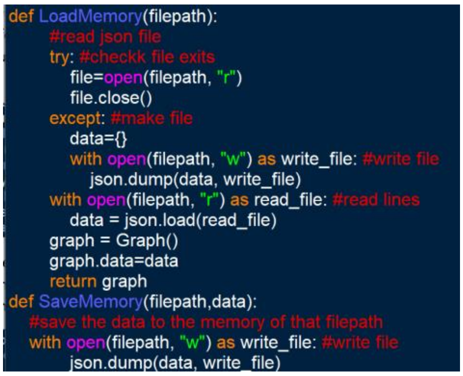


```
>this is a great app
Thank you for your feedback.
great app
['great app', 'C-feedback'] 0
C-feedback
['C-feedback', 'great app'] 0
>this is a terrible app
Thank you for your feedback.
great app
['great app', 'C-feedback'] 0
C-feedback
['C-feedback', 'terrible app'] 0
terrible app
['terrible app', 'C-feedback'] 0
>this is a great app
Thank you for your feedback.
great app
['great app', 'C-feedback'] 1
C-feedback
['C-feedback', 'terrible app'] 0
['C-feedback', 'great app'] 0
terrible app
['terrible app', 'C-feedback'] 0
>
```

It takes in the input and links it all with feedback. This means that the highest connections between feedback and nodes, and the connection between that node and other nodes will signify whether something is significant enough for the university to take action.

## Saving data

The data saving will be saved in a .json file, and retrieved into the graph based on the filename. The sentence graph, question graph and statement graph will all be saved into different files. It will have a simple start up function which retrieves the current data and saves new data as it goes along. If the datafile does not exist the function will create an empty file to read from.



```
def LoadMemory(filepath):
      #read json file
      try: #checkk file exits
         file=open(filepath, "r")
         file.close()
      except: #make file
         data={}
         with open(filepath, "w") as write_file: #write file
            json.dump(data, write_file)
      with open(filepath, "r") as read_file: #read lines
         data = json.load(read_file)
      graph = Graph()
      graph.data=data
      return graph
def SaveMemory(filepath,data):
   #save the data to the memory of that filepath
   with open(filepath, "w") as write_file: #write file
      json.dump(data, write_file)
```

The above two functions manage the loading and saving of the memory. It is saved in a dictionary format, and opened into a dictionary structure. As the graph data structures are saved as dictionaries, json files are perfect. The only concerns I would have is large amounts of data slowing down processing, or potentially running out of RAM. Future versions may need a better method of saving data such as using a folder system branching off vertices and saving edges as text files. This would only open up relevant data. This is something to consider later down the line in development.

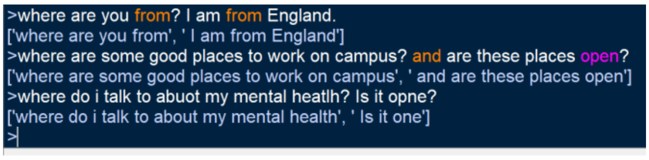
Text files rarely take up many GB of data, so as long as the operating system used to host the software is reasonable like Ubuntu which uses a low amount of RAM [2] the system should function correctly.

An alternative saving pattern which will work better will be the bot server saving every X minutes. This will stop the slowdown of processing for the user and save when necessary.

## Managing data

### Grammar and spelling checks

Using libraries, I can implement a spelling and grammar engine which should improve accuracy when messaging the bot. The main algorithm is set up for sentences, therefore the grammar system will need to split up sentences using full stops, question marks and exclamation marks. This would then filter each sentence through one by one and form a paragraph answer.

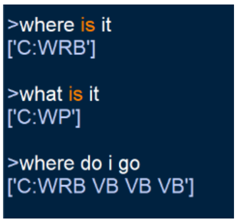


```
>where are you from? I am from England.
['where are you from', ' I am from England']
>where are some good places to work on campus? and are these places open?
['where are some good places to work on campus', ' and are these places open']
>where do i talk to abuot my mental heatlh? Is it opne?
['where do i talk to about my mental health', ' Is it one']
>
```

The spelling and grammar will convert the text to its most correct form and split it to sentences to be individually processed.

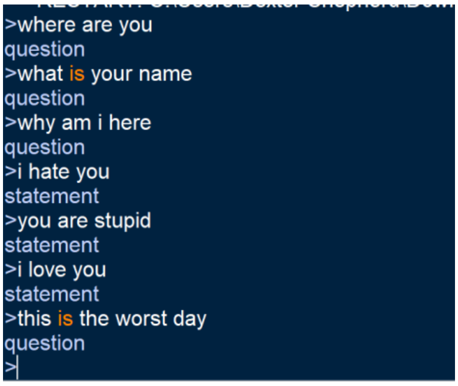
### Vague questions

In human conversation we are able to jump from subjects. “I am struggling with essay writing. Where do I get help?”. The second sentence would be confused as to what is being said. Having a subject save method will be useful, where the last sentence to be processed will keep its subjects stored. When the system gets confused because there are no subjects, it will take the subjects from before.



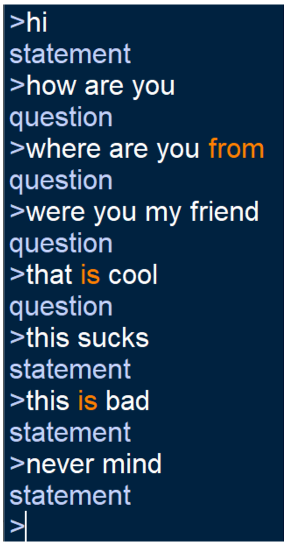
```
>where is it
['C:WRB']

>what is it
['C:WP']

>where do i go
['C:WRB VB VB VB']
```

If there are subjects but it cannot find anything like it, it will look for an exact response with the old subject. If no exact response is found then it will respond with a “I’m sorry” answer, and try find something similar to show the user. If nothing is found then it will just have an apology answer.



```
>where are you
question
>what is your name
question
>why am i here
question
>i hate you
statement
>you are stupid
statement
>i love you
statement
>this is the worst day
question
>
```

I started doing the sentence analyses to determine whether a word was a statement or a question. I was not entirely accurate. I worked on improving the training data but still was not accurate. I made a more accurate version of this code before but it was using a more complex data structure. I could potentially add a trainer of wrong sentences, or indeed allow the admin to keep up the training. I was looking at it from a stupid angle. I am trying to get every word to point to whether a sentence is or is not. Using the comparison method from the sequencer library I could improve the algorithm. The graph method just wasn’t working.

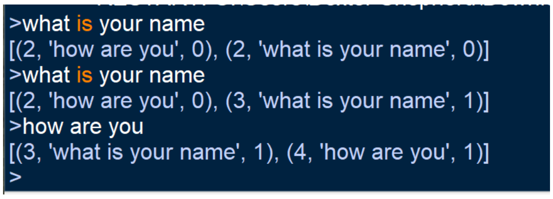


```
>hi
statement
>how are you
question
>where are you from
question
>were you my friend
question
>that is cool
question
>this sucks
statement
>this is bad
statement
>never mind
statement
>
```

The accuracy improves with this new model. I implemented this into the main bot object.

## Bot learning

The bot would enter all unknown questions into a file and await the person who monitors the system to add an answer. This could be a member of the student union. They will the be able to type an answer to a question. There will need to be a method to remove information, and also add information which deletes after a certain date. This is so information which does not need to be there for long periods of time, such as temporary events. If information changes, the staff can delete the current information and add new. This will be managed through a separate bot client called an admin Bot Client. It will be connected to via an alternative method.



```
>what is your name
[(2, 'how are you', 0), (2, 'what is your name', 0)]
>what is your name
[(2, 'how are you', 0), (3, 'what is your name', 1)]
>how are you
[(3, 'what is your name', 1), (4, 'how are you', 1)]
>
```

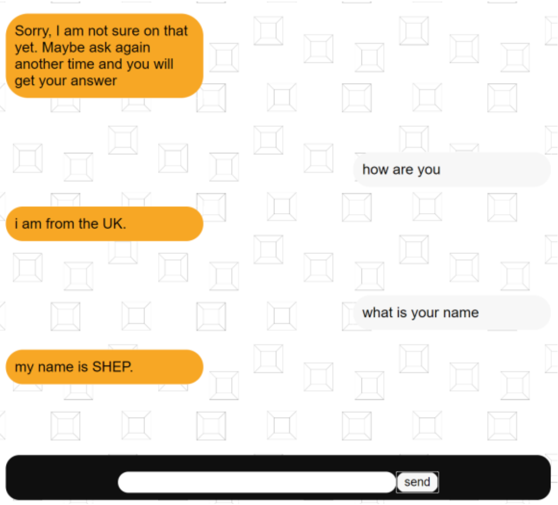
I set up the file to be a database, and wrote a class method for this. When a question is not found, it will add it to the database, and if it already exists it will increase the priority of that question.

## Client side and Server side

The client side and server side will be managed by JavaScript and Python respectively. The JavaScript will take input from the browser with a text input box, and then send this data to the server listed. If there is no response then it will need to output the server is not available at that time. If there is a response, then it will wait for a data transmission back and then display it on the browser.

After a while of setting it all up, I made a server using Ubuntu Server and Apache. I hosted a website on the server, and Python code using the Websocket library. I then used Javascript client side to connect to the Python server side and echo responses back to the browser. This is the fundamental structure that this application will use for input and output. Validation of text will be done client side to save server processing time. The server will take in text, form a response and return it.

I installed the necessary libraries and data onto the server, and adapted the code to work with the new server side processing code. It would wait for a user request on port 50007 and return a string. Then on 50008 it will wait for admin response. The testing of the client worked.



Sorry, I am not sure on that yet. Maybe ask again another time and you will get your answer

how are you

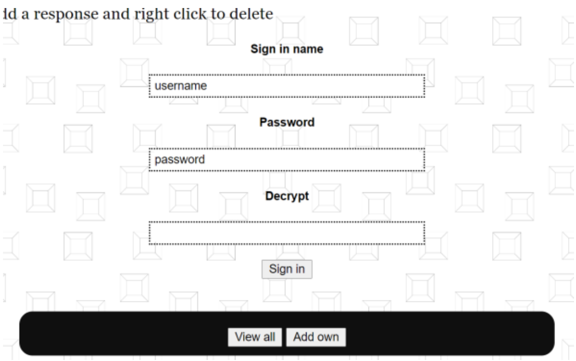
i am from the UK.

what is your name

my name is SHEP.

The issue is now updating the algorithm to work better with the stripping down of language. After going back and sorting that out the algorithm has been working better. I went on to develop the admin side of the computer system.

On the admin side you will have to sign in, all features are closed until the user signs in with the correct credentials. The username and password is currently set in the software, but I will change this down the line and use encryption of the passwords.



ld a response and right click to delete

Sign in name

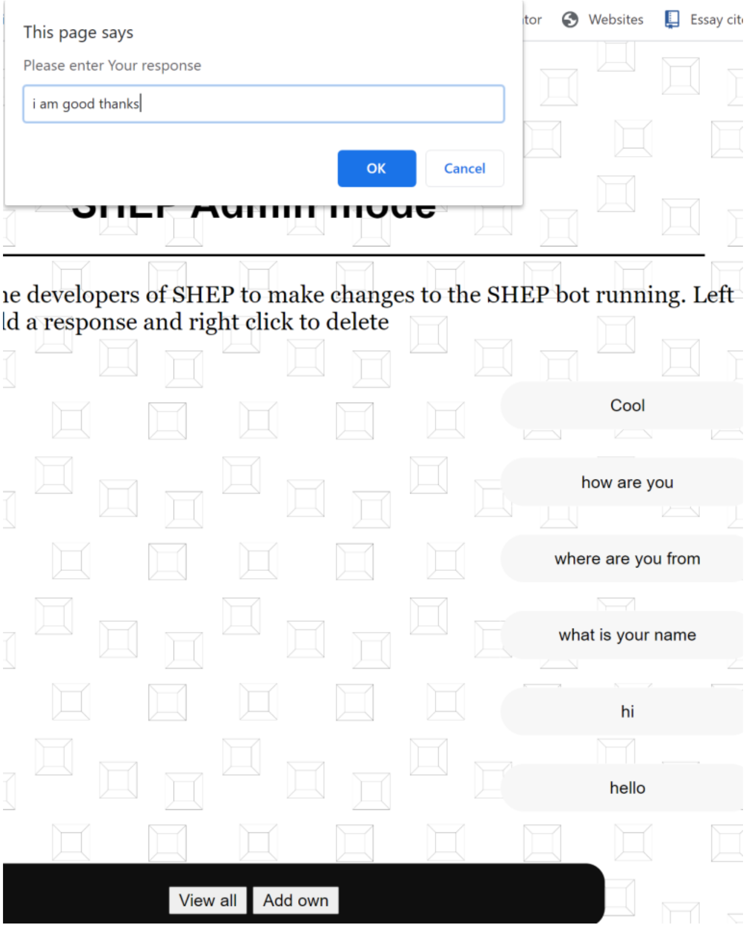
username

Password

password

Decrypt

If the password is incorrect the server will respond with an error code and the browser will notify the person.



This page says

Please enter Your response

i am good thanks

SHEP Admin mode

ne developers of SHEP to make changes to the SHEP bot running. Left
ld a response and right click to delete

Cool

how are you

where are you from

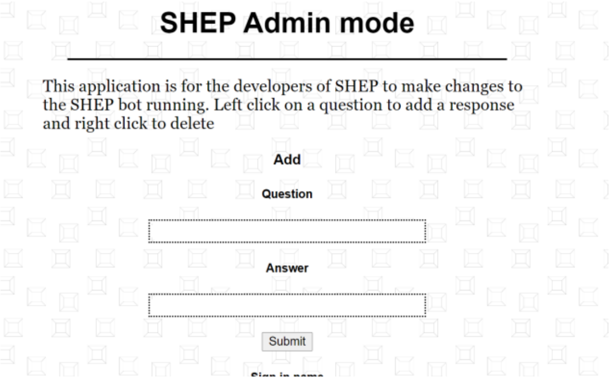
what is your name

hi

hello

The above shows the admin page once signed in. All the responses which need to be added show up, and the user can click them to get a dialogue box and add responses. They can also right click to delete them from them from the system. This will be used if someone enters a silly response.

I then proceeded to add options to add and delete.



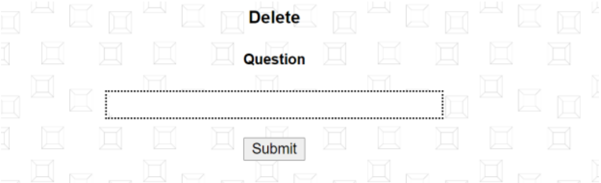
SHEP Admin mode

This application is for the developers of SHEP to make changes to the SHEP bot running. Left click on a question to add a response and right click to delete

Add

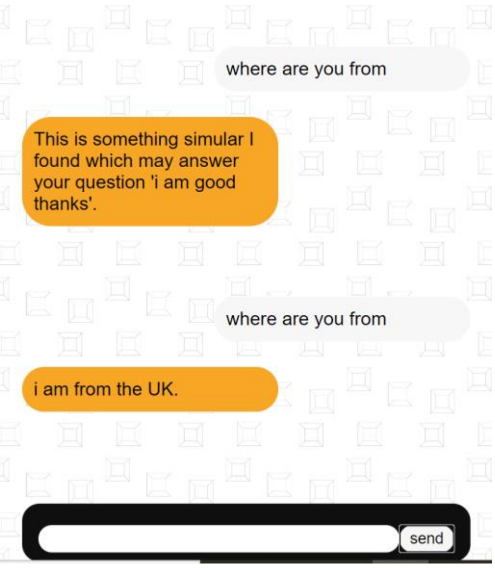
Question

Answer



Delete

Question



where are you from

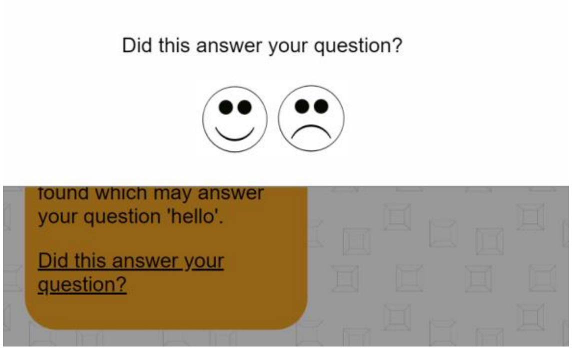
This is something simular I found which may answer your question 'i am good thanks'.

where are you from

i am from the UK.

The image above shows the client side where the first time the bot did not know, but the admin added it and the next time the person asked “where are you from” it knew.

The final addition that I made to the client side was a feedback option which.



Did this answer your question?

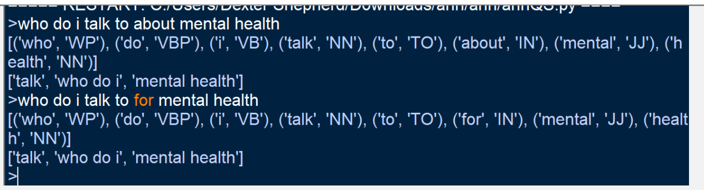
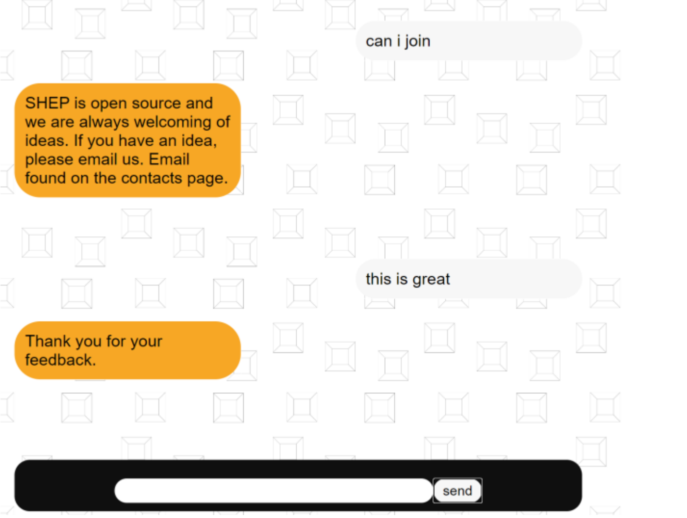
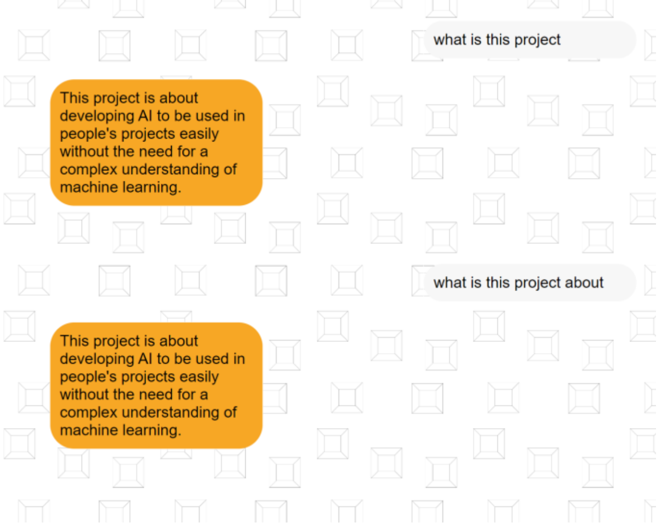
found which may answer your question 'hello'.

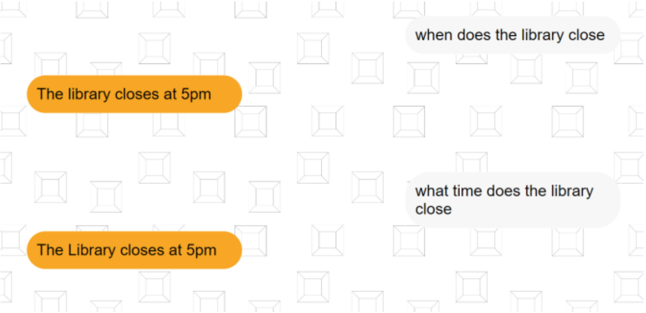
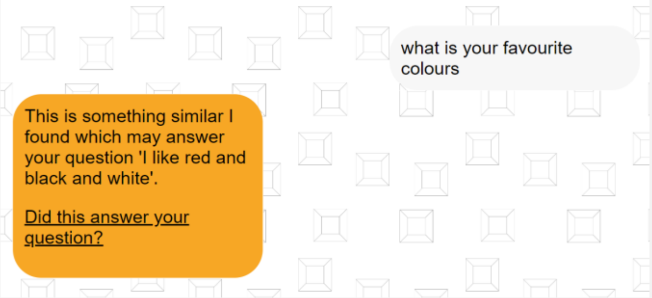
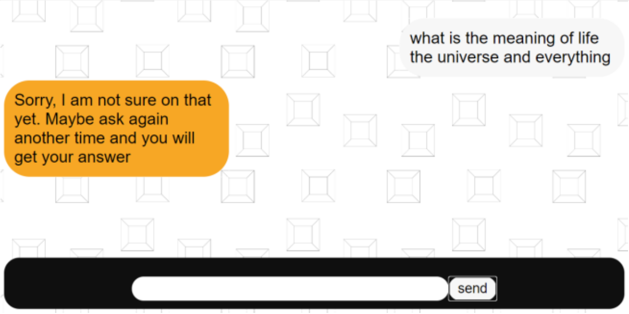
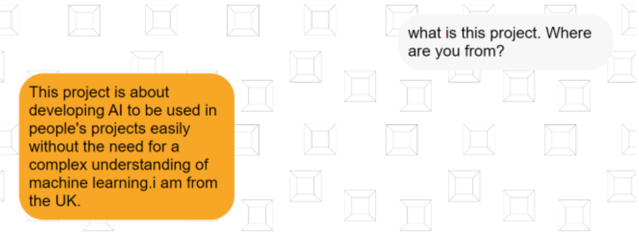
Did this answer your question?

You click the link in the comment below, and it comes up with a pop up allowing you to make a choice. Negative feedback will send the comment to the confused section. Positive feedback will work differently where the answer will be added for the question.

# Testing

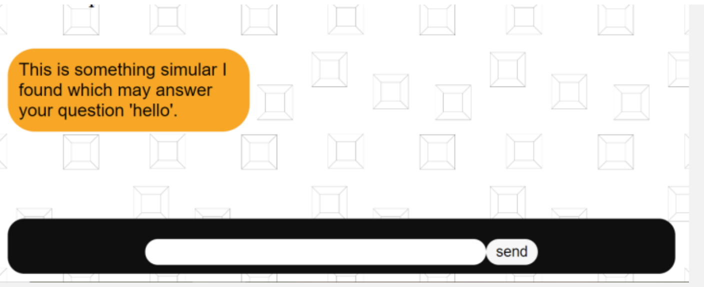
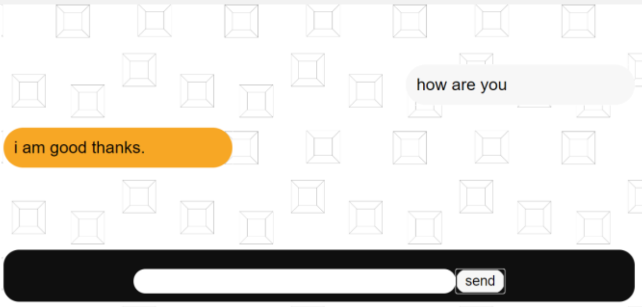
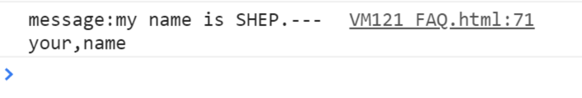
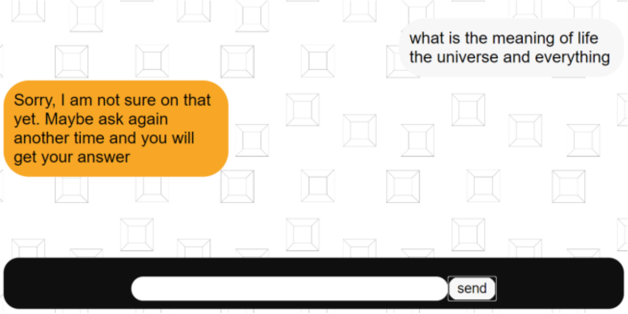
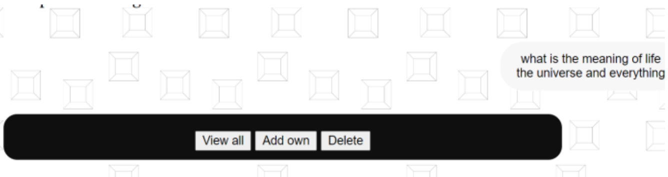
## Bot code

| *Test No* | *Test* | *Outcome* |
|---|---|---|
| *1* | AI splits a paragraph down into manageable sentences and performs spell check. | <br>===== RESTART: C:/Users/Dexter Shepherd/Downloads/ahh/ahh/ahhQS.py =====<br>>who do i talk to about mental health<br>[('who', 'WP'), ('do', 'VBP'), ('i', 'VB'), ('talk', 'NN'), ('to', 'TO'), ('about', 'IN'), ('mental', 'JJ'), ('health', 'NN')]<br>['talk', 'who do i', 'mental health']<br>>who do i talk to for mental health<br>[('who', 'WP'), ('do', 'VBP'), ('i', 'VB'), ('talk', 'NN'), ('to', 'TO'), ('for', 'IN'), ('mental', 'JJ'), ('health', 'NN')]<br>['talk', 'who do i', 'mental health']<br>><br>The sentence is broken down, and due to people having spell check on their phone the spelling engine was removed as it was unnecessary processing.<br>**PASS** |
| *2* | The AI splits a sentence into a statement or a question or neither | <br>can i join<br>SHEP is open source and we are always welcoming of ideas. If you have an idea, please email us. Email found on the contacts page.<br>this is great<br>Thank you for your feedback.<br> <br>To a reasonable degree of accuracy **PASS** |
| *3* | A statement is saved | It saves to a json file, and is loaded back in once I close the program and open it again.<br>**PASS** |
| *4* | A question which has been added will find a response. | <br>what is this project<br>This project is about developing AI to be used in people's projects easily without the need for a complex understanding of machine learning.<br>what is this project about<br>This project is about developing AI to be used in people's projects easily without the need for a complex understanding of machine learning. |

|  |  |  |
|---|---|---|
|  |  | <br>when does the library close<br>The library closes at 5pm<br>what time does the library close<br>The Library closes at 5pm<br>As you can see, my response is saved to this answer and similar meaning responses will return the answer.<br>**PASS** |
| 5 | A question is not found but there are saved questions about that topic. | <br>what is your favourite colours<br>This is something similar I found which may answer your question 'I like red and black and white'.<br>Did this answer your question?<br>**PASS** |
| 6 | A question which has never been asked and is not entered | <br>what is the meaning of life the universe and everything<br>Sorry, I am not sure on that yet. Maybe ask again another time and you will get your answer<br>The information is added to the confused file, as it appears in the admin side.<br>**PASS** |
| 7 | Data already in the confused file is added | The data will not add it again but will increase its priority |
| 8 | A paragraph is entered with a vague sentence following another sentence. “what mental health services are there? Where are they” | <br>what is this project. Where are you from?<br>This project is about developing AI to be used in people's projects easily without the need for a complex understanding of machine learning.i am from the UK.<br>**PASS** |

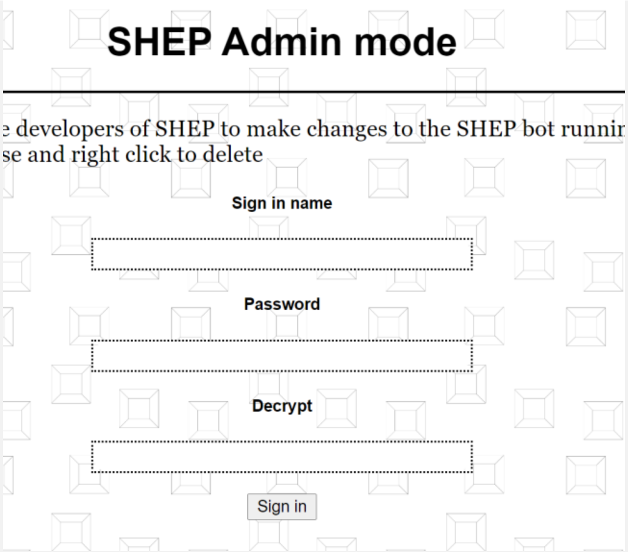
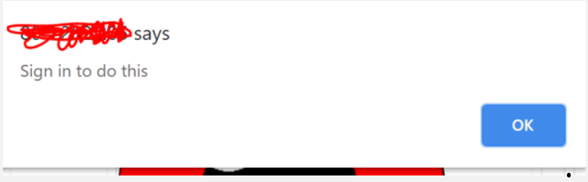
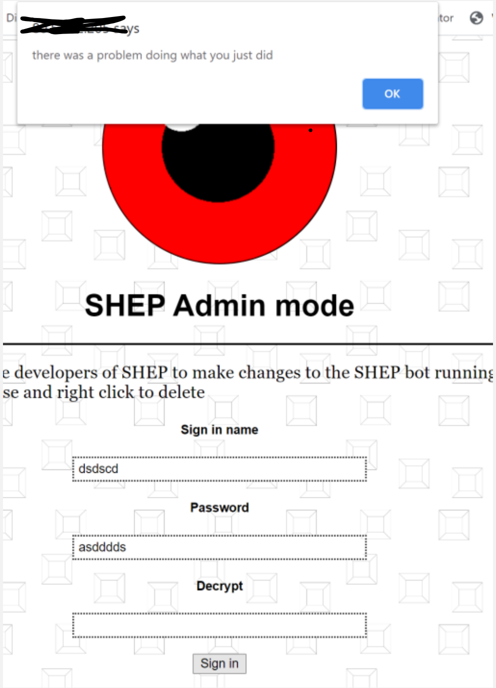
## Client to Server and Server to Client

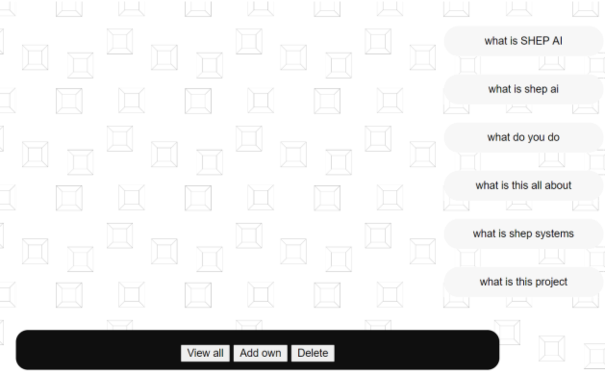
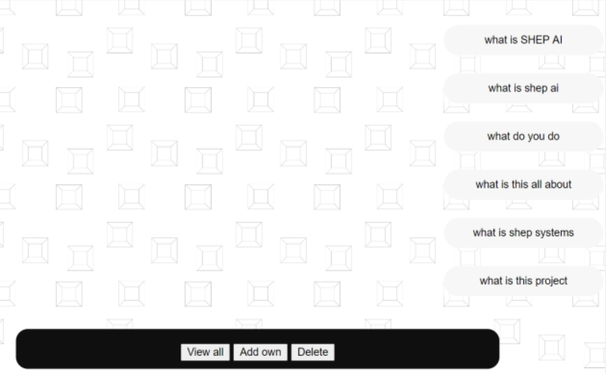
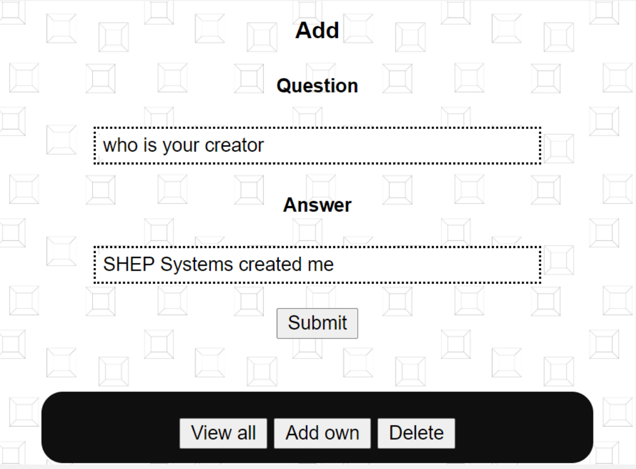
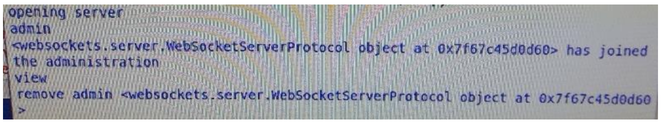
| *Test No* | *Test* | *Outcome* |
|---|---|---|

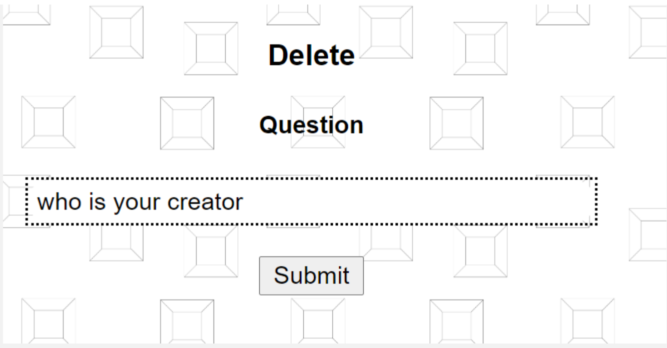
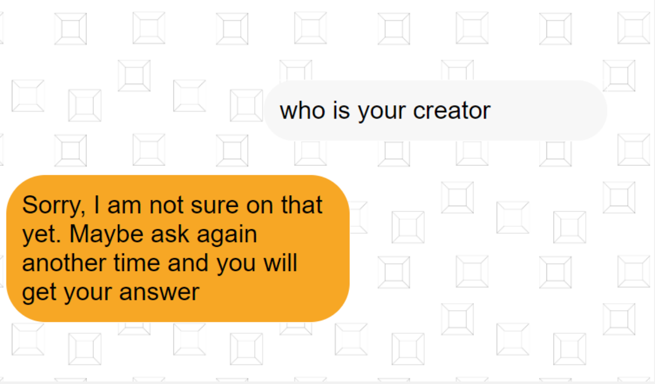
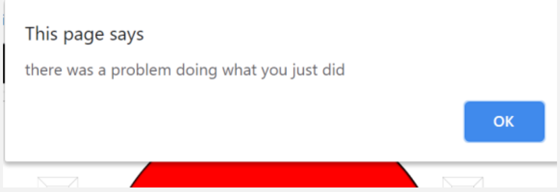
| | | |
|---|---|---|
| 1 | The client connects to the server | <br>This is something simular I found which may answer your question 'hello'.<br> <br>The ping was "hi" and got the response "hello". There is a bug to do with multiple responses which will need to be fixed. That's is separate to this test.<br>**PASS** |
| 2 | The client sends the text input out to the server | <br>how are you<br>i am good thanks.<br> <br>The response is sent back<br>**PASS** |
| 3 | The client stores the subjects of the past and the response. | Local variables are shown in the browser console.<br><br>`message:my name is SHEP.---   VM121 FAQ.html:71`<br>`your,name`<br>The subjects are being picked out<br>**PASS** |
| 4 | The client responds negative feedback to a message | After inputting information it does not know, I was able to view it in the admin mode which takes from the confused file.<br><br>what is the meaning of life the universe and everything<br>Sorry, I am not sure on that yet. Maybe ask again another time and you will get your answer<br> <br>And in admin mode:<br><br>what is the meaning of life the universe and everything<br> <br>**PASS** |

## Admin side

| Test No | Test | Outcome |
|---|---|---|

| | | |
|---|---|---|
| 1 | Makes user sign in with correct credentials. | <br>SHEP Admin mode<br>e developers of SHEP to make changes to the SHEP bot runnir<br>se and right click to delete<br>Sign in name<br>Password<br>Decrypt<br> <br>The interface is shown and you cannot click anything until the sign in is correct.<br><br>[redacted] says<br>Sign in to do this<br> <br><br>[redacted] says<br>there was a problem doing what you just did<br> <br>SHEP Admin mode<br>e developers of SHEP to make changes to the SHEP bot running<br>se and right click to delete<br>Sign in name: dsdscd<br>Password: asdddds<br>Decrypt<br> <br>Once you sign in it takes you to the main content<br>**PASS** |

|  |  |  |
|---|---|---|
| 2 | The user can receive the top amount of questions to add. | <br>what is SHEP AI<br>what is shep ai<br>what do you do<br>what is this all about<br>what is shep systems<br>what is this project<br> <br>**PASS** |
| 3 | The user can answer questions. | The user does this two ways. They can click on an item in the view mode:<br><br>what is SHEP AI<br>what is shep ai<br>what do you do<br>what is this all about<br>what is shep systems<br>what is this project<br> <br>Or they can add it manually:<br><br>Add<br>Question<br>who is your creator<br>Answer<br>SHEP Systems created me<br> <br>**PASS** |
| 4 | The user disconnects | <br>opening server<br>admin<br>&lt;websockets.server.WebSocketServerProtocol object at 0x7f67c45d0d60> has joined the administration<br>view<br>remove admin &lt;websockets.server.WebSocketServerProtocol object at 0x7f67c45d0d60<br>><br>When the user has been idle for a certain amount of time they are removed from the approved admin list. They will have to re-sign in.<br>**PASS** |

| | | |
|---|---|---|
| 5 | The user can delete responses which are incorrect | <br>Delete<br>Question<br>who is your creator<br>Submit<br>The phrase "who is your creator" existed in the system.<br><br>who is your creator<br>Sorry, I am not sure on that yet. Maybe ask again another time and you will get your answer<br>Then we saw that it was removed. When I tried it again and submitted the same thing twice I got an error message as expected.<br><br>This page says<br>there was a problem doing what you just did<br>OK<br>You cannot delete what is not there<br>**PASS** |

Bot: **PASS**

Client: **PASS**

Admin: **PASS**

All initial tests passed and work as expected with only a few issues.

## *Bugs I found while testing*

**ISSUE**

When adding sentences with the same output "hi"="hello", "hey"="hello" , The system would get confused and say "something I found like" due to the number of inputs connected to the same answer decreasing the overall chance of being the output. This method was used as words will be used on many responses, whereas responses tend to be more unique. Of course, this is not all the time so will need fixing.

**FIX** ☑

To fix this I would add in a validation technique. When an answer is added, if it already exists it will be given a count code to make it a different node. The client side will be programmed to ignore this.

**ISSUE**

Sometimes I will make a statement and it will give me a response that I have given it feedback even though the statement wasn't for feedback.

**FIX** ☑

A quick fix is to class sentence such as "hello" as questions for the sake of a friendly user interface and more personal user experience.

**ISSUE**

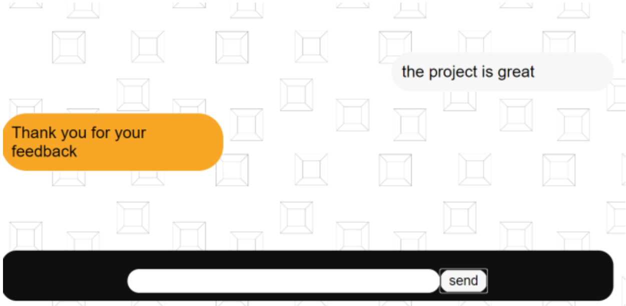
There is no way for the user to point out false information.

**FIX** ☑

Add a report false information which will send a report form to the admin.

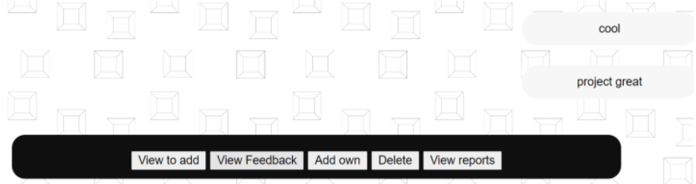
**ISSUE**

There is no way to view the feedback from the user



the project is great

Thank you for your feedback



cool

project great

**FIX** ☑

Add it to the admin page, where the user can request to see the feedback from users.

**ISSUE**

I can copy and paste large sentences and break the code

**FIX** ☑

Add validation and not allow sending of strings over a size of 500 characters.

**ISSUE**

Accidental deletion of the confused file data is irreversible without typing in what you deleted.

**FIX** ☑

I can develop an undo button, and use a fixed size stack data structure to hold each item of data

**ISSUE**

The positive feedback does nothing. If an answer is correct the user should be able to add it

**FIX** ☑

I can simply add the same algorithm as the admin bot to add a response. If this response is wrong then someone can report it as false information.

**ISSUE**

Cannot delete feedback if potentially inappropriate feedback or old feedback

**FIX** ☑

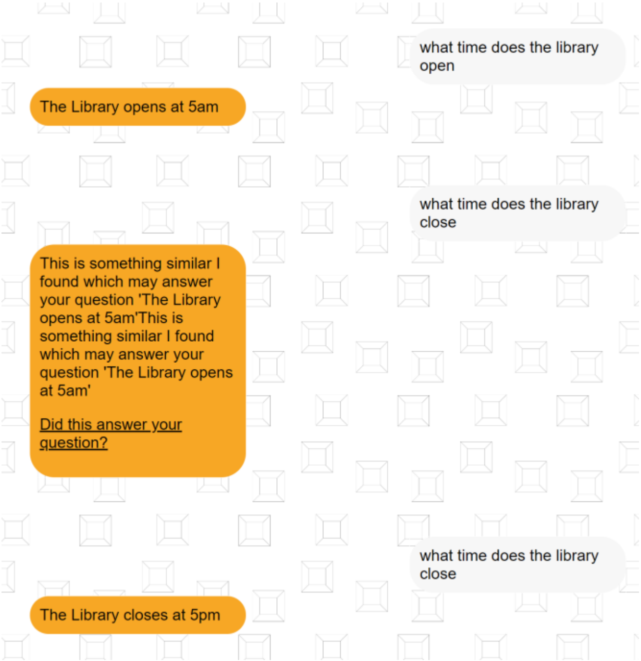
Add a delete feedback method

**ISSUE**

The system cannot handle conjunction of sentences. “When does the library open and when does it close”. If there has been two separate parts of information stored for these two questions it will find neither due to the sequence matcher length of sentence calculation.

**FIX**

Add in a way that the nodes of the system will be split into sub sentences where if there is no exact response to the system, it will split up the nodes into the sub sentences and accumulate responses based on these inputs. If it is missing subjects in one of the sub sentences, it will look for it in the other assuming the subject is remaining the same. “When does the library open and when does it close” will split to “When” “Does” “Library” “Opens”, and “Does”, “Close”. Being both items of information are trained into the system, it will pick out both as “when does the library opens” and “does close”+”library” will get the response wanted.



what time does the library open

The Library opens at 5am

what time does the library close

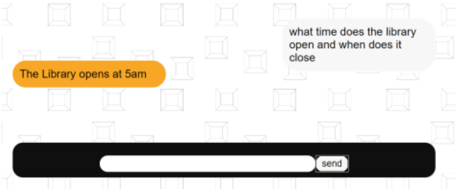
This is something similar I found which may answer your question 'The Library opens at 5am'This is something similar I found which may answer your question 'The Library opens at 5am'

Did this answer your question?

what time does the library close

The Library closes at 5pm

At this moment in time I have improved the algorithm from what it once was. But it still is lacking. Originally the code would not understand anything. Now it will respond with one of the sentences.

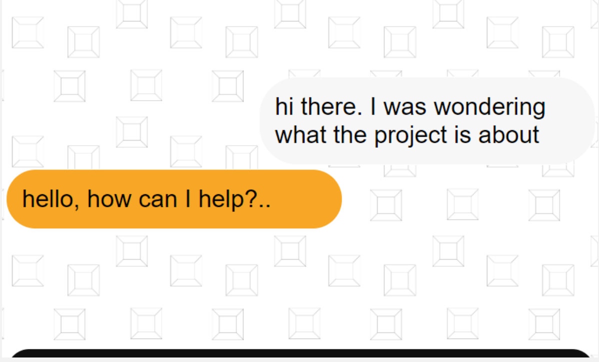
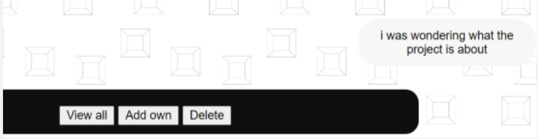
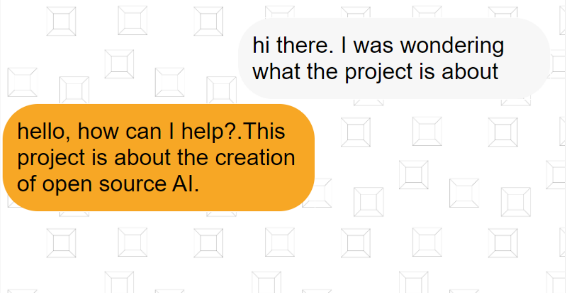


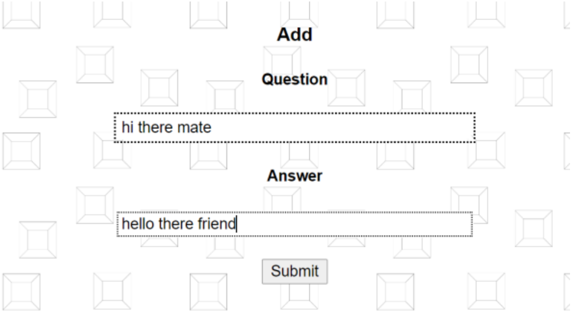
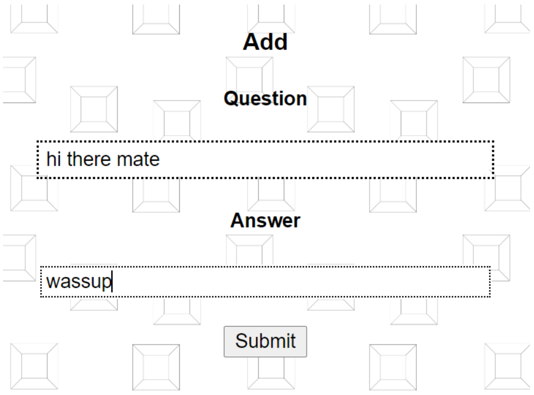
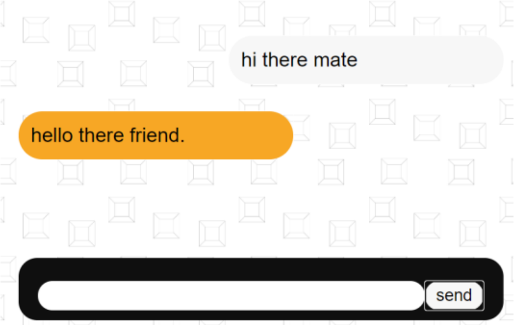
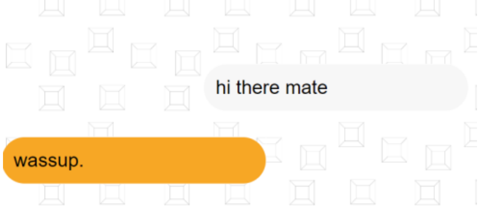
what time does the library open and when does it close

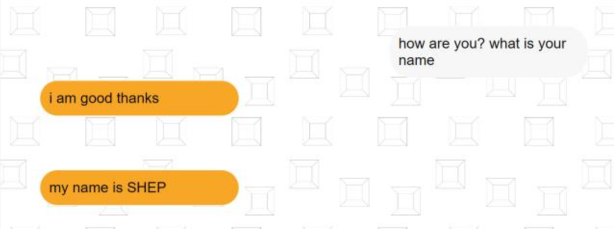
The Library opens at 5am

Of course, this is not what I intended to do, but it is how it has ended up. I will have to work further on this and go at it from a different perspective.

*Rigorous testing*

| TEST NO | TEST | OUTCOME |
|---|---|---|
| 1 | What happens if I enter two sentences, one added and one not. | <br>hi there. I was wondering what the project is about<br>hello, how can I help?..<br>The code copes by only responding to one, and adding the unknown one to the admin.<br><br>i was wondering what the project is about<br> <br>Next time I enter it:<br><br>hi there. I was wondering what the project is about<br>hello, how can I help?.This project is about the creation of open source AI.<br>The responses add up.<br>I am happy with how it deals with this method. |

| | | |
|---|---|---|
| **2** | Add two same questions with different answers and then delete them | <br>Add<br>Question<br>hi there mate<br>Answer<br>hello there friend<br>Submit<br><br>Add<br>Question<br>hi there mate<br>Answer<br>wassup<br>Submit<br>Are both added. The normal client side finds one of the answers:<br><br>hi there mate<br>hello there friend.<br>send<br>And when deleted it only deletes one of them<br><br>hi there mate<br>wassup.<br>Ideally it would delete both, although the user should try not to add two messages which are the same. I have no desire to change this unless it becomes a bigger problem.<br><br>UPDATE: I added a loop to the deletion algorithm, so it will delete all answers linked to the question. It became an issue when I was updating false information, and it kept giving me false information. |
| **3** | What if I have a long response and I give feedback on a sentence. | The code gets confused and easily breaks. This is a major issue which needs sorting. I could potentially split messages up into separate responses. |

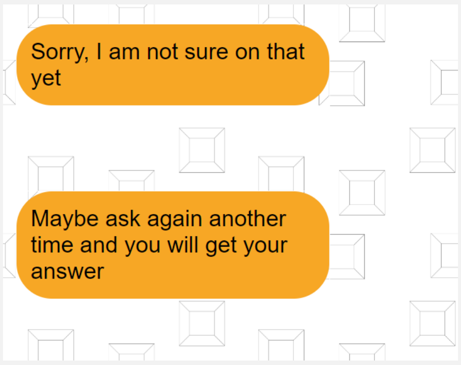


how are you? what is your name

i am good thanks

my name is SHEP

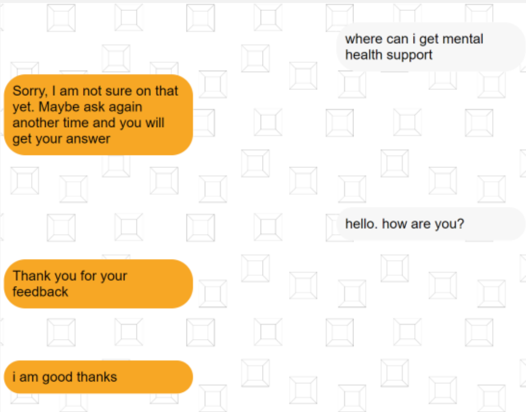
This will then be more accurate deciding which sentence the user is referring to. Another problem is this…



Sorry, I am not sure on that yet

Maybe ask again another time and you will get your answer

I could change the symbol in the main algorithm so it splits sentences using something else such as "$".



where can i get mental health support

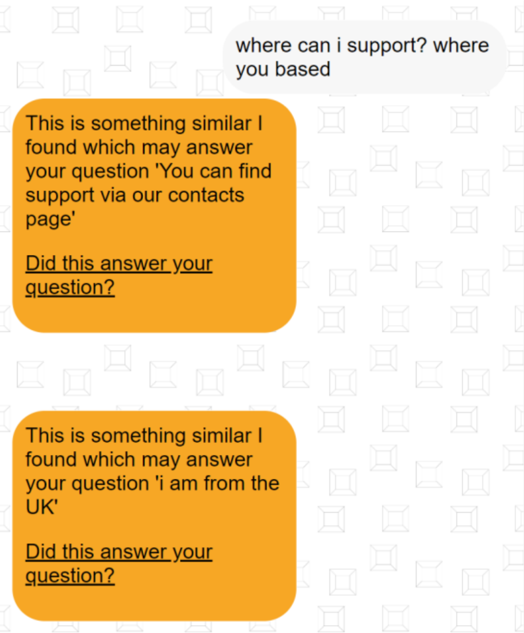
Sorry, I am not sure on that yet. Maybe ask again another time and you will get your answer

hello. how are you?

Thank you for your feedback

i am good thanks

And that did the trick.



where can i support? where you based

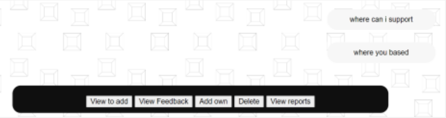
This is something similar I found which may answer your question 'You can find support via our contacts page'

Did this answer your question?

This is something similar I found which may answer your question 'i am from the UK'

Did this answer your question?

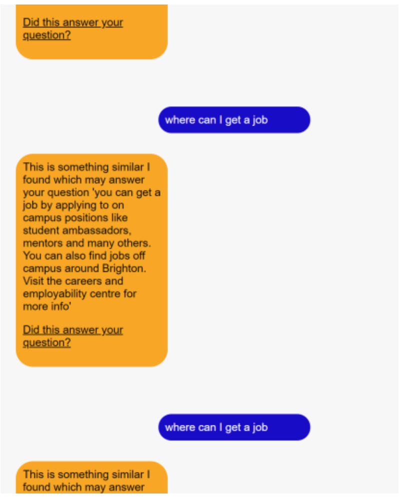
I could add negative feedback to both of these and get the responses I wanted.



where can i support

where you based

After getting everything working, it then stopped working… The system lost it’s ability to understand information.



Did this answer your question?

where can I get a job

This is something similar I found which may answer your question 'you can get a job by applying to on campus positions like student ambassadors, mentors and many others. You can also find jobs off campus around Brighton. Visit the careers and employability centre for more info'

Did this answer your question?

where can I get a job

This is something similar I found which may answer

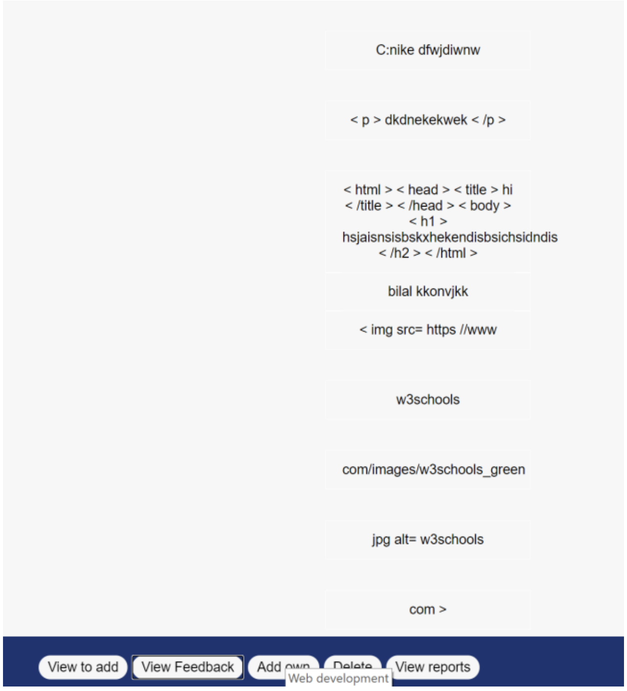
This got me thinking whether the use of a graph could either be changed to work better, or perhaps thrown out altogether. This could also be to do with the checking algorithm and the language analyser. I got to developing a better language analyser. I improved the accuracy of the language analyser but still had the problem with the code not adding the new information. After tracing through errors and RAM usage, I found it was all to do with the adding of new information. It worked, and then stopped without any changes happening. Even after deleting the data and starting again the problem persisted. I changed the sequence matcher code, as it did not take into consideration orders of sentences can change. I could add back the sequence similarity to be a variable on the percentage similarity overall, if I come into complications.

# Deployment of the system

## Trials

The first trial is the alpha, where people who I know will trial out the code, and find bugs or issues with it. They will then submit a feedback form for me to improve the client experience. I will then take it to a beta trial which would mean someone else will manage the admin, and the new fixes will be added to the client side.

Prior to the main testing I have set up a extreme testing approach, where I will let fellow Computer Science students purposely try and break my code. This is so I can fix up any weak areas of the program.



C:nike dfwjdiwnw

< p > dkdnekekwek < /p >

< html > < head > < title > hi
< /title > < /head > < body >
< h1 >
hsjaisnsisbskxhekendisbsichsidndis
< /h2 > < /html >

bilal kkonvjkk

< img src= https //www

w3schools

com/images/w3schools_green

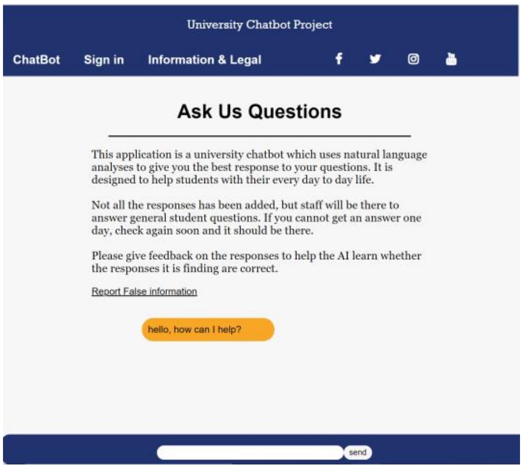
jpg alt= w3schools

com >

These tests seemed to have gone well. The first student was unable to break my program.

### *ALPHA*

The alpha trial was done to test out the program with untrained users. This is to find any bugs, security issues, and functionality problems. I set up the following interface:



Ask Us Questions

This application is a university chatbot which uses natural language analyses to give you the best response to your questions. It is designed to help students with their every day to day life.

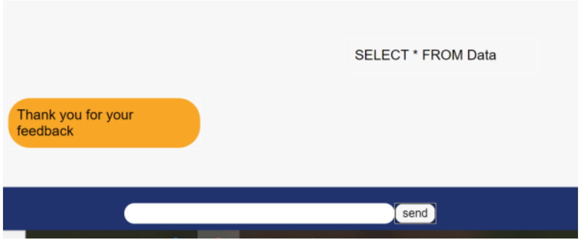
Not all the responses has been added, but staff will be there to answer general student questions. If you cannot get an answer one day, check again soon and it should be there.

Please give feedback on the responses to help the AI learn whether the responses it is finding are correct.

Report False information

hello, how can I help?

This uses my university's organizational colour scheme, to match their website. The interface is friendly and easy to use. It provides all the information needed and some links to the SHEP social medias for publicity. I will also add in a link to a survey when I have made that survey.

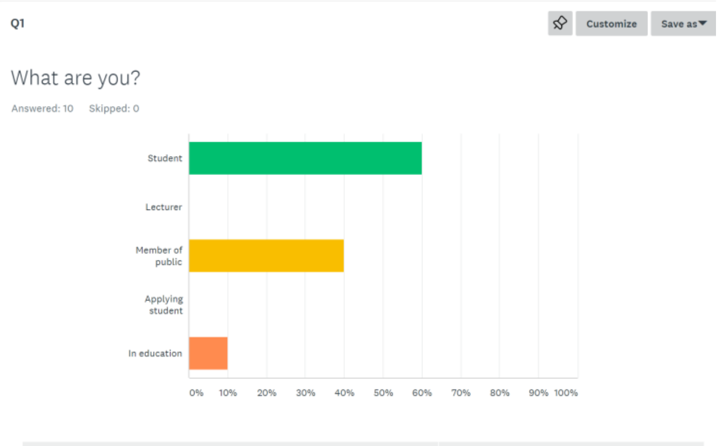


SELECT * FROM Data

Thank you for your feedback

The second tester attempted to put in SQL injections, but due to the nature of the linguistic analyses the information was secure.

I put out a survey and here were the results:

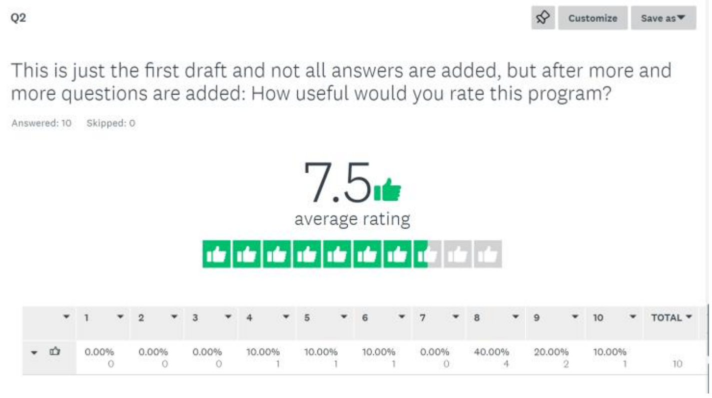


Q1

What are you?

Answered: 10 Skipped: 0

The main testers were students but also members of the public to make sure it is usable by all.



Q2

This is just the first draft and not all answers are added, but after more and more questions are added: How useful would you rate this program?

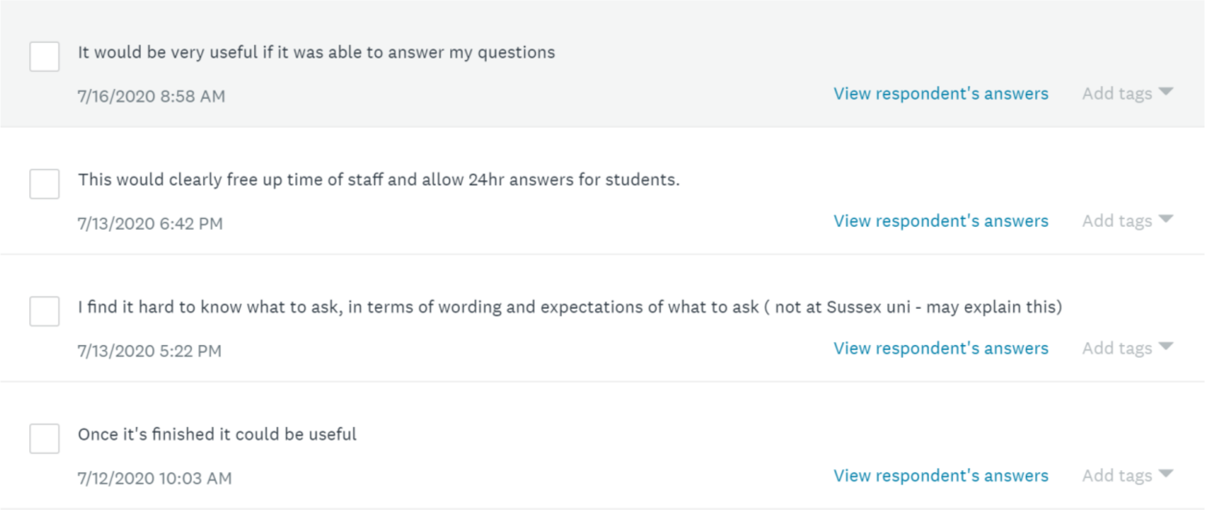
Answered: 10 Skipped: 0

7.5

average rating

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.00%<br>0 | 0.00%<br>0 | 0.00%<br>0 | 10.00%<br>1 | 10.00%<br>1 | 10.00%<br>1 | 0.00%<br>0 | 40.00%<br>4 | 20.00%<br>2 | 10.00%<br>1 | 10 |

Overall people seemed happy with the project. Below shows the responses about why they gave the previous answer.



□ It would be very useful if it was able to answer my questions

7/16/2020 8:58 AM

□ This would clearly free up time of staff and allow 24hr answers for students.

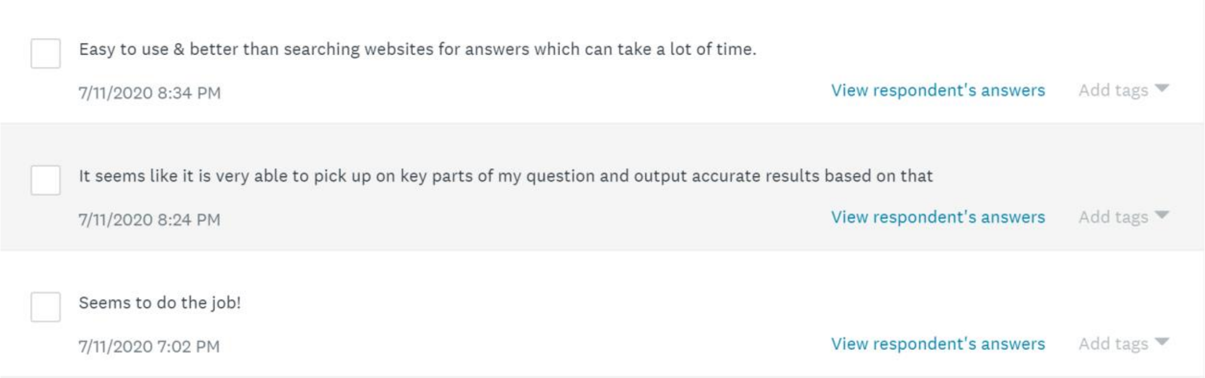
7/13/2020 6:42 PM

□ I find it hard to know what to ask, in terms of wording and expectations of what to ask ( not at Sussex uni - may explain this)

7/13/2020 5:22 PM

□ Once it's finished it could be useful

7/12/2020 10:03 AM



□ Easy to use & better than searching websites for answers which can take a lot of time.

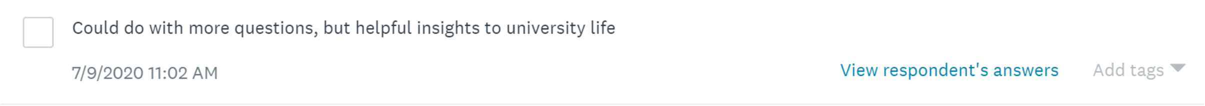
7/11/2020 8:34 PM

□ It seems like it is very able to pick up on key parts of my question and output accurate results based on that

7/11/2020 8:24 PM

□ Seems to do the job!

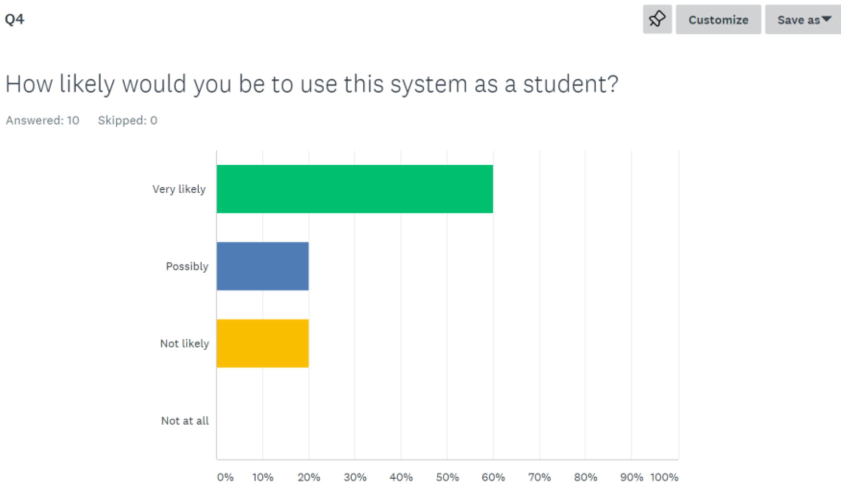
7/11/2020 7:02 PM



□ Could do with more questions, but helpful insights to university life

7/9/2020 11:02 AM

The summery of these comments are that people are they like the 24-hour quick answers to save the time and researching. However, the negative comments are more about the lack of questions added, which I expected to be the case. This will be fixed when the university add all their information.

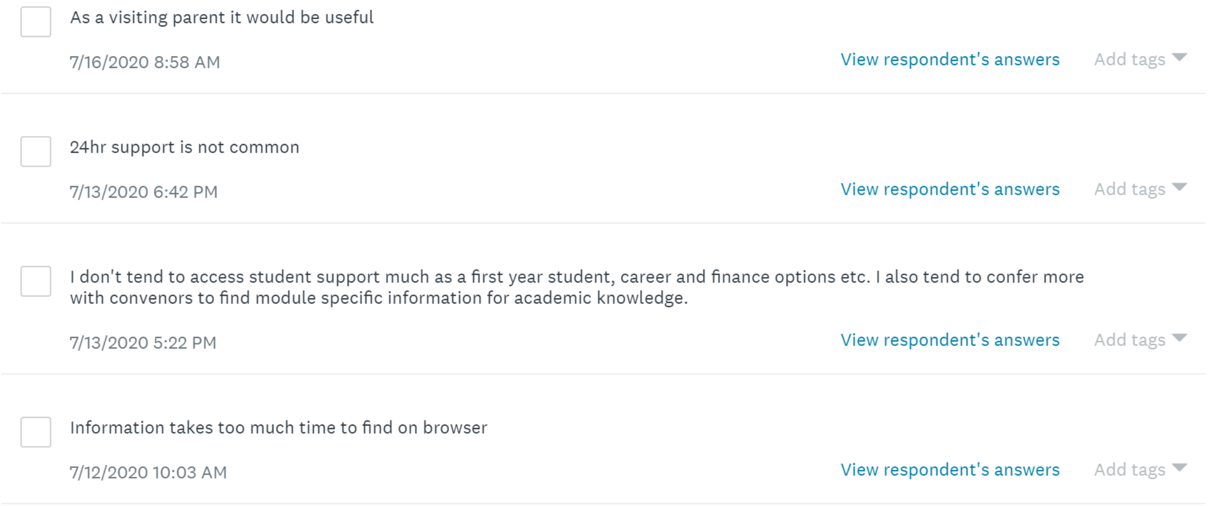


Q4

How likely would you be to use this system as a student?

Answered: 10 Skipped: 0

Most people would potentially use this system.


View respondent's answers
Add tags
View respondent's answers
Add tags
View respondent's answers
Add tags
View respondent's answers
Add tags

□ As a visiting parent it would be useful

7/16/2020 8:58 AM

□ 24hr support is not common

7/13/2020 6:42 PM

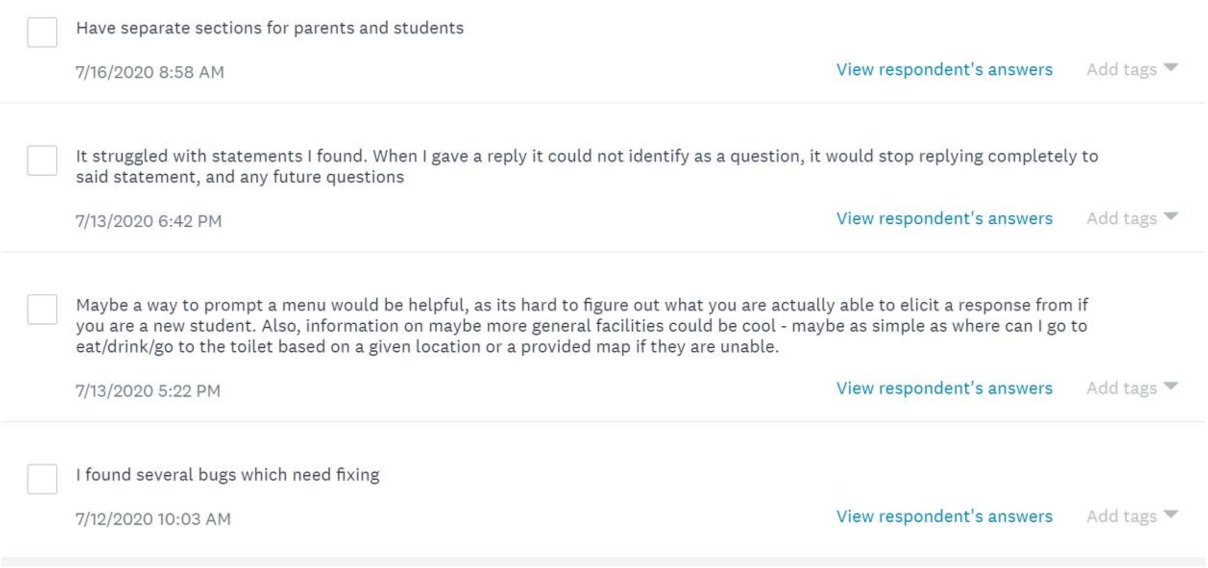
□ I don't tend to access student support much as a first year student, career and finance options etc. I also tend to confer more with convenors to find module specific information for academic knowledge.

7/13/2020 5:22 PM

□ Information takes too much time to find on browser

7/12/2020 10:03 AM

Here are a few of the comments that we got. People like the speed efficiency, 24-hour support and it being easier than an email. Some people are on top of things so do not need it.



□ Have separate sections for parents and students

7/16/2020 8:58 AM

□ It struggled with statements I found. When I gave a reply it could not identify as a question, it would stop replying completely to said statement, and any future questions

7/13/2020 6:42 PM

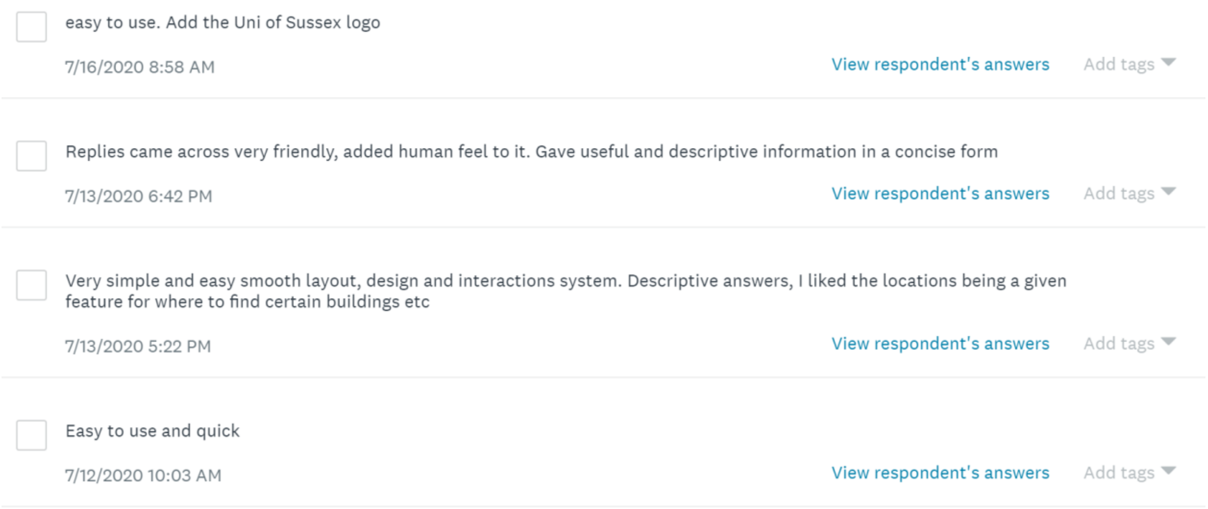
□ Maybe a way to prompt a menu would be helpful, as its hard to figure out what you are actually able to elicit a response from if you are a new student. Also, information on maybe more general facilities could be cool - maybe as simple as where can I go to eat/drink/go to the toilet based on a given location or a provided map if they are unable.

7/13/2020 5:22 PM

□ I found several bugs which need fixing

7/12/2020 10:03 AM

The comments on areas to improve were interesting. I like the idea of having a prompt of questions you can ask. That is something I could use within the next user interface. Some people found bugs which they reported at the end. The question and statement interaction will get better the more people ad to it, as it converts phrases added to questions that can be answered.



□ easy to use. Add the Uni of Sussex logo

7/16/2020 8:58 AM

□ Replies came across very friendly, added human feel to it. Gave useful and descriptive information in a concise form

7/13/2020 6:42 PM

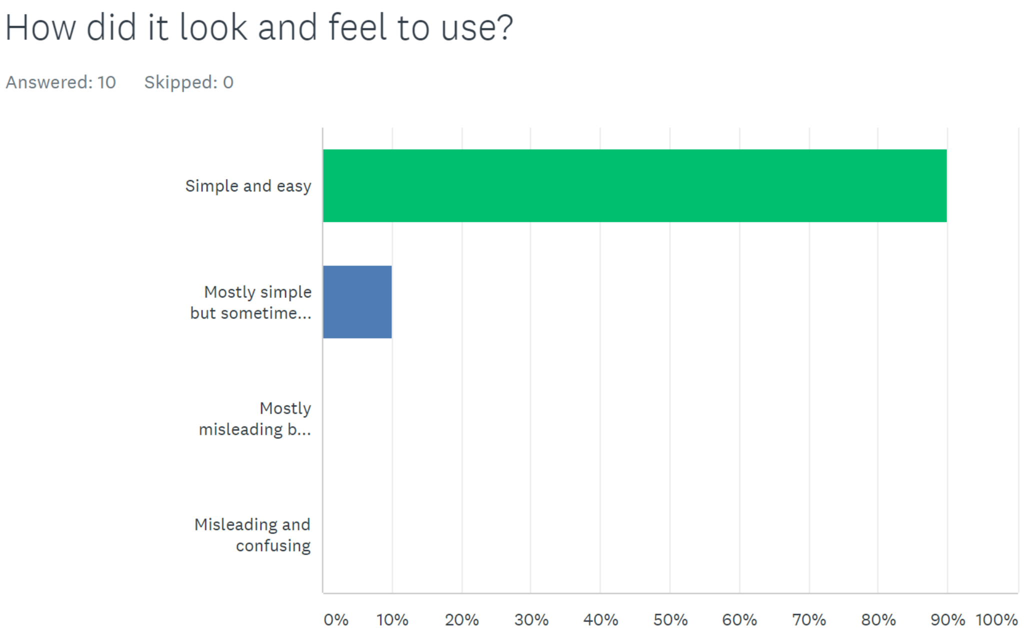
□ Very simple and easy smooth layout, design and interactions system. Descriptive answers, I liked the locations being a given feature for where to find certain buildings etc

7/13/2020 5:22 PM

□ Easy to use and quick

7/12/2020 10:03 AM

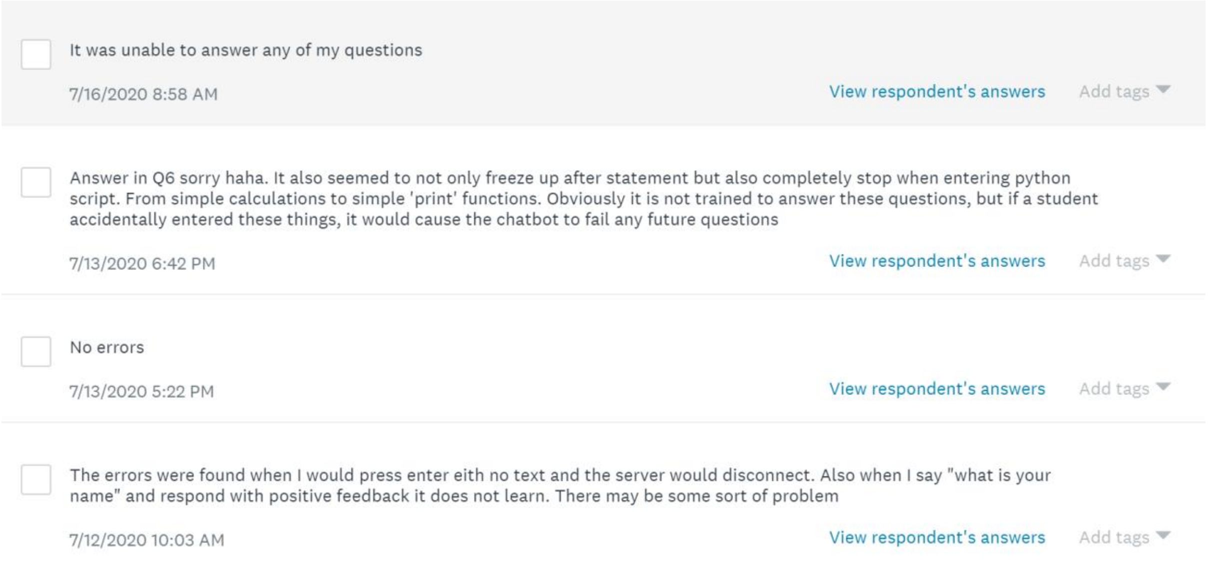
People liked the simple and easy layout.



How did it look and feel to use?

Answered: 10 Skipped: 0

This was confirmed in this question.



□ It was unable to answer any of my questions

7/16/2020 8:58 AM

□ Answer in Q6 sorry haha. It also seemed to not only freeze up after statement but also completely stop when entering python script. From simple calculations to simple 'print' functions. Obviously it is not trained to answer these questions, but if a student accidentally entered these things, it would cause the chatbot to fail any future questions

7/13/2020 6:42 PM

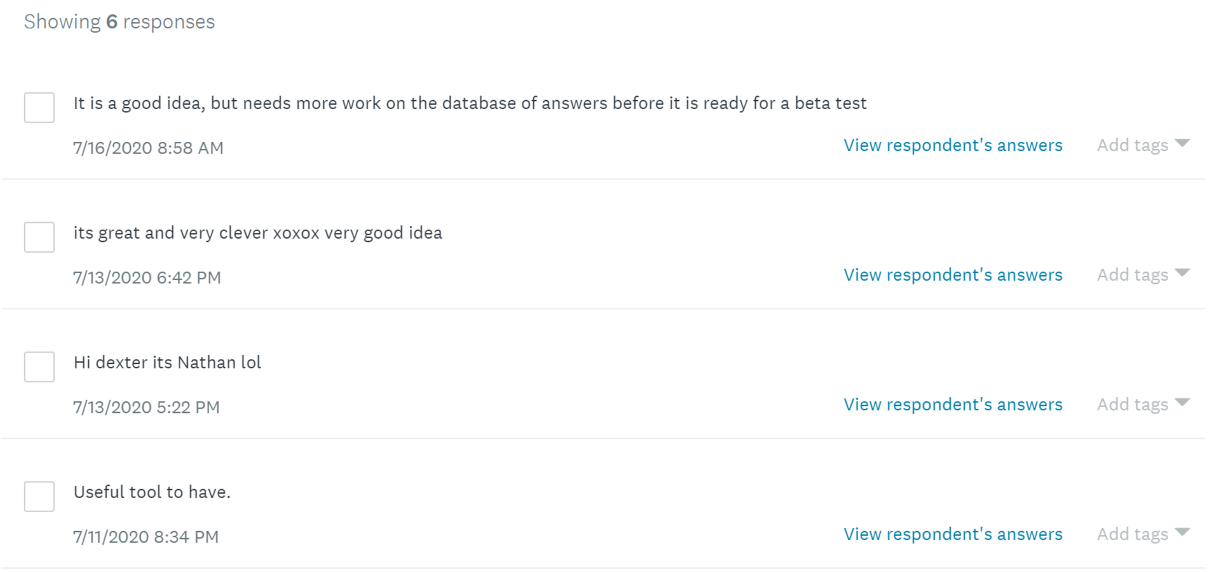
□ No errors

7/13/2020 5:22 PM

□ The errors were found when I would press enter eith no text and the server would disconnect. Also when I say "what is your name" and respond with positive feedback it does not learn. There may be some sort of problem

7/12/2020 10:03 AM

There were a few errors within the code, which I got to fixing right away.



Showing **6** responses

□ It is a good idea, but needs more work on the database of answers before it is ready for a beta test

7/16/2020 8:58 AM View respondent's answers Add tags

□ its great and very clever xoxox very good idea

7/13/2020 6:42 PM View respondent's answers Add tags

□ Hi dexter its Nathan lol

7/13/2020 5:22 PM View respondent's answers Add tags

□ Useful tool to have.

7/11/2020 8:34 PM View respondent's answers Add tags

People gave overall positive feedback and liked the project. I was happy with the feedback and knew what I had to do to get it ready for the next stage. Fix the bugs and make improvements to its ease of use.

### *BETA*

This is to test the code after the additions/deletions made in the alpha stage have been implemented. It is to test out with a wider range of people. Then to gather feedback on improvements.

## Set up

The system would be implemented on a server as a python file which would receive data sent from a client-side web application and return the appropriate response. The Python language would require the following libraries installed:

- JSON
- NLTK
- spellChecker
- sqlite3
- websocket

The file paths specified to where the passwords are kept, and where the files are kept will need to be changed when on a new system, for the code to work.

## Legal

NLTK is an open source software which is distributed under the Apache license version 2.0 [5]. The apache licence allows users to use or modify software in their own projects.

The spellChecker library comes under the MIT License which is another open source license for anyone to use.

Python and the inbuild libraries belong to the Python Software Foundation, and are licensed under the License agreement for python 3.8.3 [6] where Python is made for the royalty free distribution world-wide.

Websocket has demands that its own requirements are kept within distribution of the source code [7]. Their License will have to be appended to our own license.

My own software licence will explain this and relinquish responsibility to the user if this software is misused.

## Future ideas

### Applications of software

In the future this sort of system could be deployed further than current universities, it could move to schools as a parent information device. It could be used by the Army for recruitment questions. It could be used as an NHS 111 service and go as far to book appointments for people.

The council could use it to gather issues in statements, and answer questions around the city. Local businesses could pay to be recommended. The code could be modified to allow the reporting of crime ad submission of evidence.

This frame work has many applications if taken further.

Ways to develop the framework would be to add some self-learning elements. For example making similarity between grammar in questions and answers to help it make more accurate responses. Also to find similar meanings

The self-learning could also store data locally in cookies, and retrieve interests to promote events/ articles/ information related and of interest.

### Algorithm changes

I would like to follow up this project with a more self-sufficient program which learns from it's mistakes. When the user responds with negative feedback because it has found the wrong information, correlations as to why this has taken place should be calculated, to prevent this grammatical error from happening again. This would give the system its own self learning understanding of language and using it to improve.

Further steps would be to find its own responses. Chat with it to learn off of people. A "gossip" bot. Students learn off of one another and people gossiping about what to do. A gossip bot would find ways in which people talk to one another, find out about what is going on and help students be a part of that. Of course such applications hold many problems such as trolls and personal data being shared, it would require quite a lot of thought to make this safe.

# References

[2] https://askubuntu.com/questions/552095/how-much-ram-does-ubuntu-use, 2015. *Ubuntu RAM usage,* s.l.: s.n.

[3] https://en.wikipedia.org/wiki/Natural_language_processing, 2 June 2020. *Natural language processing,* s.l.: Wikipedia.

[1] https://www.computerweekly.com/opinion/Its-good-to-chat-but-who-to-The-role-of-chatbots-in-digital-transformation, 2020. The role of chatbots in digital transformation. *Computer Weekly.*

[4] https://www.nltk.org/api/nltk.html, 2020. *NLTK documentation.*

[5] https://www.apache.org/licenses/LICENSE-2.0, 2004, *Apache software License*

[6] https://docs.python.org/3/license.html, 2001, *Python Software foundation*

[7] https://websockets.readthedocs.io/en/stable/license.html, 2013*, Aymeric Augustin and contributors*