

Chatbot Project: CarBot

System Description

Motivation

I have been a car enthusiast for a long time, and I have always loved learning more about cars and how they work. When I was first trying to get into cars, I found the task very daunting and even embarrassing as a beginner. It can be difficult to teach yourself something as technical as the complex and extensive engineering that goes into cars. Learning from people is a good approach, but not everyone has somebody around them that knows enough about cars to teach them. Many people could also be anxious or embarrassed to ask someone they know. I have always wanted to help people find a love for cars like I did, which is why I had the idea for CarBot.

Overview

CarBot is a chatbot geared towards helping people who want to learn the basics of how cars work. It can answer questions about some of the major beginner to intermediate concepts such as engines, transmissions, fuel, racing, and 7 more topics. The Bot is written in python and uses the Chatterbot library to generate responses using a machine learning model. The model was trained using a knowledge base scraped from <https://auto.howstuffworks.com/car.htm>. By training the model on some broad questions and their possible responses, the Chatterbot model learns how to respond to questions based on the conversation flows in the training data. There is more information about the knowledge base in Appendix A. The bot is capable of basic personalization by storing information about the user and their preferences. The preferences are stored in User models and used to train the model in a more personalized way. More information about these User models is in Appendix B.

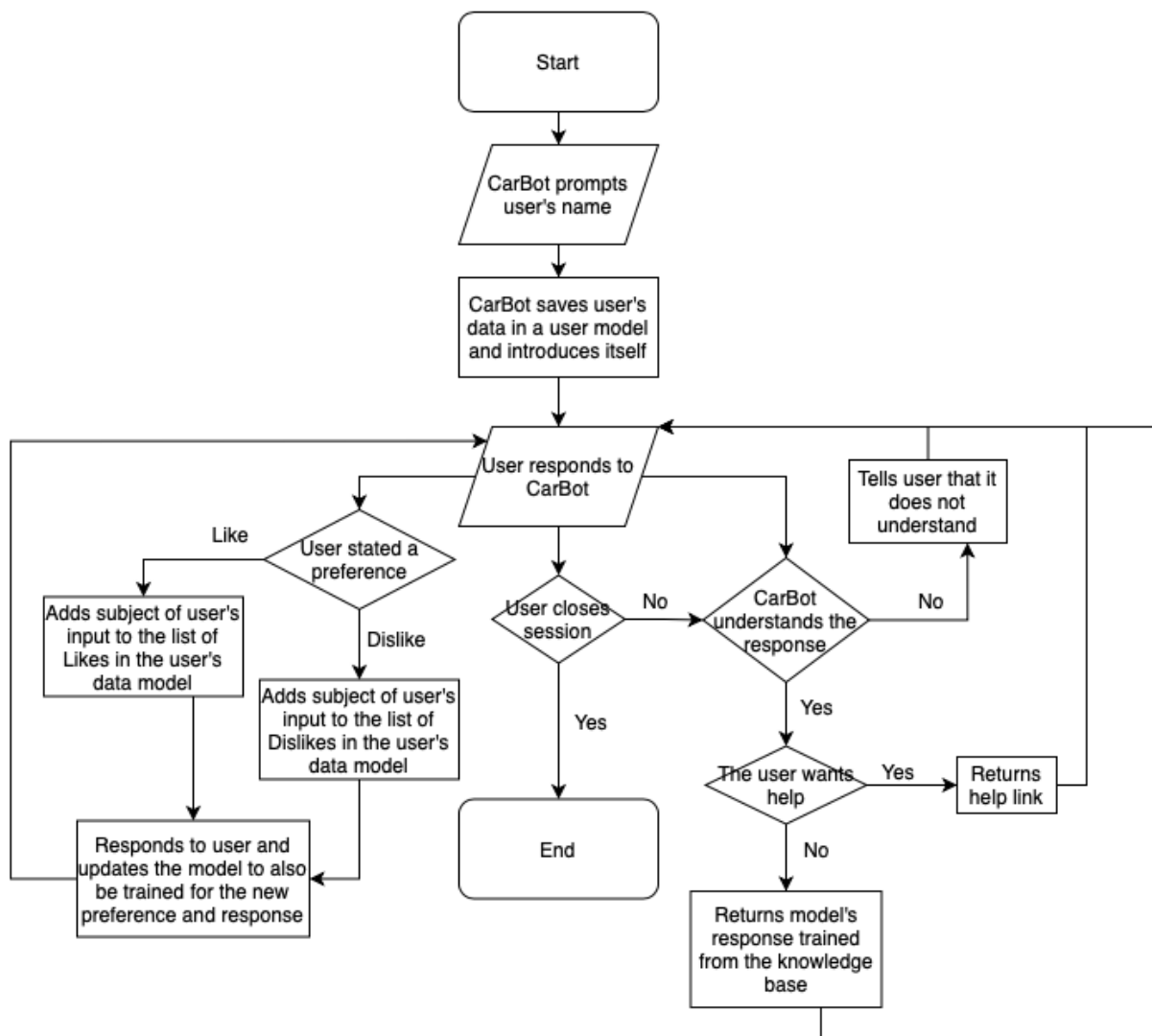
NLP Concepts Used

- Name Entity Recognition (NER)
 - NER is the step of text processing that focuses recognizing significant entities in a body of text. These entities can be people, organizations, locations, etc. For the purposes of identifying the user's name, NER was used to identify the part of their input that is the name. Since there is no guaranteed user response format when they are prompted for their name, by checking for the PERSON entities, the user's name could be extracted regardless of what else the response contained.
- Tokenizing User Response
 - Tokenizing bodies of text is an integral step in NLP. Natural language tends to include a lot of content that is not particularly useful in language processing. Therefore, it is important to convert bodies of text into a form that NLP algorithms can use. This was done by converting the words from a user's responses into tokens that ignored any stopwords (words removed during processing because they are not important) and punctuation. By using NLTK's `word_tokenize()` function, user responses are converted to lists of the tokens

that carry significant meaning. This eliminates any negative influence of discrepancies in user input while the program processes the input.

- Parts-of-speech Tagging (POS Tagging)
 - Identifying the parts of speech of the tokens can give significant context to the subjects, actions, and other parts of a body of text. Here, NLTK's `pos_tag()` function was used to find the parts of speech of the extracted tokens in order to identify the subject of a user's input. This helps identify their likes and dislikes.

Diagram



Sample Interactions

```
CarBot: Hello! I am CarBot. What is your name?
user:  Rahul

CarBot: Hello Rahul, I know about how cars work, so feel free to ask questions. I am here to inform you about engines, transmissions, racing, and other things to do with cars!

CarBot: If you need additional help, type 'help'. Type 'close' if you would like to end the conversation.

CarBot: How can I help you today?

Rahul: how does an engine work?

CarBot: the throttle valve controls the power and speed of the engine by limiting how much air the engine takes in.

Rahul: how can cars make more power?

CarBot: engine questions and answers

Rahul: how is power made?

CarBot: in a car, you need a clutch because the engine spins all the time, but the car's wheels do not.

Rahul: what does the clutch do?

CarBot: the clutch is a device that lets you connect and disconnect the engine and the transmission.

Rahul: what is a transmission?

CarBot: the transmission allows the gear ratio between the engine and the drive wheels to change as the car speeds up and slows down.

Rahul: how do i change the gear ratio?

CarBot: these clutches engage and disengage various sets of planetary gears , which are made of a central drive gear surrounded by gears that orbit it.

Rahul: I like driving fast cars
List Trainer: [#####] 100%

CarBot: i know a lot about cars

Rahul: do you know about race cars?

CarBot: i like cars

Rahul: what makes race cars different from normal cars?

CarBot: a champ car engine is obviously very different from the engine you find in a normal car.

Rahul: what makes race cars different?

CarBot: compared to a normal engine, the engine in a champ car has larger pistons and the pistons travel a shorter distance up and down on each stroke.

Rahul: what do pistons do?

CarBot: I'm sorry, I don't know what you're asking.

Rahul: what makes an engine work?

CarBot: the throttle valve controls the power and speed of the engine by limiting how much air the engine takes in.

Rahul: i don't like truck engines
List Trainer: [#####] 100%

CarBot: how to do an engine oil analysis

Rahul: i don't like trucks
List Trainer: [#####] 100%

CarBot: I understand, duly noted.
```

```
Rahul: i don't like tractors either
List Trainer: [#####] 100%

CarBot: I understand, duly noted.

Rahul: do you know about trucks?

CarBot: I understand, duly noted.

Rahul: i don't understand

CarBot: I'm sorry, I don't know what you're asking.

Rahul: that's okay

CarBot: I'm sorry, I don't know what you're asking.

Rahul: why do trucks use diesel?

CarBot: in a car, you need a clutch because the engine spins all the time, but the car's wheels do not.

Rahul: hmmm what is the difference between diesel and gasoline?

CarBot: diesel fuel has a higher energy density than gasoline, so a diesel engine gets better mileage.

Rahul: how can a car get more power?

CarBot: engine questions and answers

Rahul: have you ever heard of turbochargers?

CarBot: the car engine information channel explores engine performance, maintenance and new fuel technologies.

Rahul: what is a turbocharger anyways?

CarBot: I'm sorry, I don't know what you're asking.

Rahul: what is a turbocharger?

CarBot: the turbocharger pressurizes the air flowing into the engine's cylinders so that the engine can burn more fuel during each stroke.

Rahul: wow that's really cool! thank you for your time!

CarBot: of course!

Rahul: bye!

CarBot: good bye!

Rahul: close

CarBot: It was great meeting you! Talk to you later!
```

Evaluation

Overview

Overall, the bot is decent for a very simple bot made with a relatively crude training knowledge base. The bot does a fairly good job of answering questions about the 11 topics as well as responding to some basic formalities. The bot is definitely lacking, as even though it handles some variations in input from the prompts specified in the knowledge base, it still struggles with complex responses and prompt topics that might not have been covered in the base. As a very basic informational chatbot, CarBot works fairly well, but there is definitely room for improvement.

Strengths

The bot is able to give fairly reliable and understandable responses to basic questions about the 11 basic topics the knowledge base covers. It also does a good job of responding to formalities. It is also able to handle variations in some variation in the content of the user responses, such as where the user says, “hmmm what is the difference between diesel and gasoline?” and CarBot gives an accurate response. The wording and interjection do not directly match what is in the knowledge base, but the model is able to handle those variations. Additionally, CarBot provides an emergency “panic” option when the user types in “help”. It returns the auto.howstuffworks.com website link so the user can find what they need if the bot is not providing it properly. The knowledge base is also quite flexible and contains easy to understand information about a lot of the basic questions that might be asked about a person who has only heard car terminology without knowing what it means.

Weaknesses

This bot is very simple, and the knowledge base was processed very minimally. Some basic, broad question prompts were added. However, there is still data in the knowledge base that isn’t easily accessible to the user because the training conversation flow does not include a sample prompt for that information response. This is because some of the information requires more complex queries from the user for the answer to make sense, but these are too specific to reliably show up for the user. There is also information missing in the knowledge base due to a topic not being chosen or the website simply lacking that information. The responses are also only one sentence which means some relevant information to the user’s question may not show up since it's a separate sentence. Additionally, the User preferences are not actually used for any sort of answer tailoring yet and you cannot reference existing user models so those could be improved, as well as the bot’s ability to process complex and conversational responses.

Appendix A: Knowledge Base

The scraping was done by a slightly modified version of the Web Scraper from my Portfolio using the domain: <https://auto.howstuffworks.com/car.htm>. The base is still made up of common key terms and sentences that contain them, but with some broad questions manually added in to help train the Chatterbot model on possible conversation flows. Just giving the bot answer data does not train it for the other parts of the conversation that the bot will need to process, such as questions and prompts from the user. The questions were kept fairly broad so that the model could handle minor discrepancies in user wording as long as the user's input is similar enough.

Samples

```
How does an engine work?
The throttle valve controls the power and speed of the engine by limiting how much air the engine takes in.
Many cars use actuators powered by engine vacuum to open and close the throttle.
If the lead vehicle slows down, or if another object is detected, the system sends a signal to the engine or braking system to decelerate.
The throttle valve controls the power and speed of the engine by limiting how much air the engine takes in.
Everything from the engine to the tires is its own special universe of design and engineering.
Why do we need a clutch?
In a car, you need a clutch because the engine spins all the time, but the car's wheels do not.
How does an engine make power?
The throttle valve controls the power and speed of the engine by limiting how much air the engine takes in
This works in a similar way to the brake booster, which provides power to your brake system.
The throttle valve controls the power and speed of the engine by limiting how much air the engine takes in.
When most or all of the friction material is gone, the clutch will start to slip, and eventually it won't transmit any power from
It takes a great deal of power to overcome all of this drag, and that's one reason why Champ Cars need such amazing engines.
The extra fuel means extra power per stroke.
What is methanol?
Methanol can run at much higher compression ratios, meaning that you can get more power from the engine on each piston stroke.
```

Appendix B: User Models

The bot stores User information in text files that contain a user's name and their likes and dislikes. These preferences are used to then train the model further to add some personalization to the bot. Since this bot is a simple informational one, based on questions about a topic, the personal information isn't used much other than for making the experience feel a bit more personal.

Samples

```
User Name:  Rahul
Likes:      ['driving cars']
Dislikes:   ['truck engines', 'trucks', 'tractors']
```

```
User Name:  Matt
Likes:      ['driving to work', 'reading', 'trucks']
Dislikes:   ['cars', 'pasta']
```

```
User Name:  Dylan
Likes:      ['big cars', 'sandwiches', 'driving to work']
Dislikes:   ['small cars', 'france']
```