

Sanjeev Penupala, Varun Joshi

CS 4395.0W1

04/05/21

VolleyBot: A Chatbot for Volleyball

Basic Description:

Volleybot, a smart chatbot designed to answer any queries related to volleyball. Before you can interact with Volleybot, it will ask you some details about yourself, such as your name, to store in its database and use later next time for future conversations. Volleybot has over 500 sentences of knowledge related to volleyball, which allows it to ask your basic questions related to volleyball. After you are finished asking Volleybot your questions, you can exit from the chat, and Volleybot will remind you about your last conversation with it the next time you chat with it.

System Description:

For this project we primarily used Python and Google Colab to create a VolleyBall Trivia bot that answers questions about volleyball regarding rules, history and some interesting and lesser known details about the game.

The python script encapsulates all the functionalities of the chatbot including user communication with the bot, response selection from the knowledge, and user-intent understanding. First, the user input is parsed and keywords from the input are extracted by the python script also keeping track of the sentiment of each of the user's inputs. Then the similarity between the user input and the list of responses from the knowledge base is matched by using the technique of cosine similarity. The response with the highest similarity is then selected and returned back to the user.

While this response selection is in action the bot is collecting data about the user with regards to its name, likes, dislikes, and keywords the user is focusing on right now. When the user types exit to the bot this data is fed into a User object that belongs to the User class and this user is then appended to a userlist. Then when the user runs the block of code again if the same name is entered the user model provides personal remarks with the Dialog system in the form of the user's name and his likes and dislikes based on the previous conversation.

Sanjeev Penupala, Varun Joshi

CS 4395.0W1

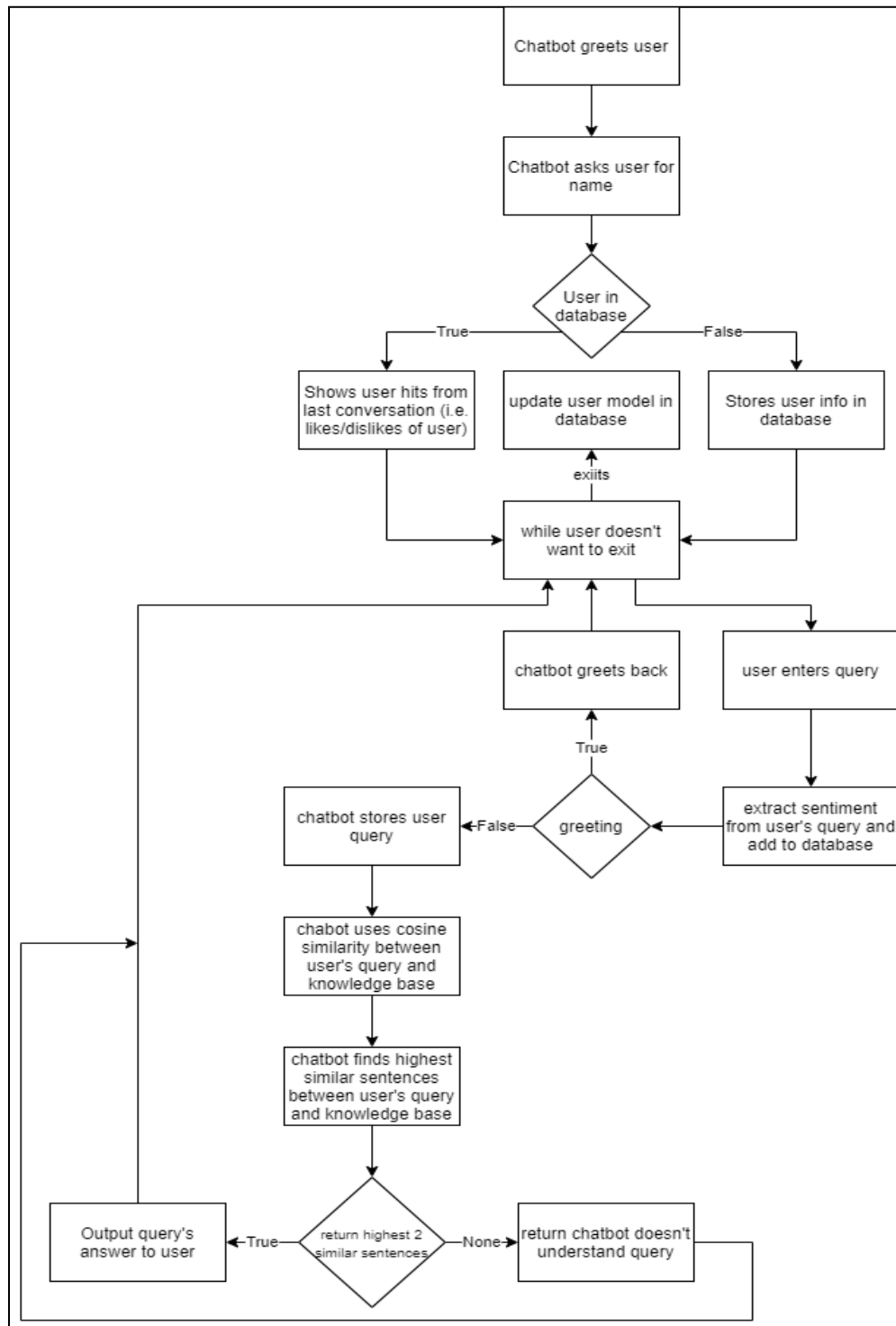
04/05/21

Specific NLP Techniques Used:

1. Text Preprocessing/Parsing: Tokenizing the text and removing stopwords, lower-casing the text, and then lemmatizing the text to ensure only important root words from the input text stay.
2. POS Tagging: Tagging the parts of speech of the text to extract the keywords from the sentence. Assigning a part of speech to every token and then extracting those tokens with the Noun, Adjective, Verb, and Pronoun tags.
3. VADER Sentiment Analysis: Determining whether the individual sentence by the user is Neutral, Positive and Negative to add to the knowledge base list of likes and dislikes.
4. Vector Cosine Similarity: Calculating the cosine similarity of two vector matrices, the response keywords and the knowledge base to get the most correct response from the bot.

Dialog Logic Diagram:

Inserted below is an overall dialog logic when it comes to interacting with Volleybot. You will notice several conditional statements in the diagram. The arrows coming from those conditional statements indicate what is the evaluation of the if statement. In the end, regardless of the path taken, the end point will be the same for all users, which is that their information will be stored in the database for future conversations with Volleybot.



Sanjeev Penupala, Varun Joshi

CS 4395.0W1

04/05/21

Sample Dialog Interactions:

Below are recorded conversations with Volleybot with a variety of queries and non-example queries to see how the bot will respond to them. The first example conversation is for new users who have never chatted with Volleybot before. The second example conversation mimics the dialogue of what a returning user would see.

New User

Type "EXIT" or "exit" once you are done talking to the bot

-> Volleybot: Hey, my name is VolleyBot. Nice to meet you. What is your name?

-> **Me: My name is Sanjeev**

-> Volleybot: Hey Sanjeev, got any questions about volleyball?

-> **Me: What is volleyball?**

-> Volleybot: Volleyball, especially beach volleyball, is played in Australia, New Zealand, and throughout the South Pacific. "I love volleyball, but it is just my job.

-> **Me: How many points do you need to win?**

-> Volleybot: In 1917, the game was changed from requiring 21 points to win to a smaller 15 points to win. A team has to win by at least two points.

-> **Me: How many people play volleyball?**

-> Volleybot: There are many variations on the basic rules of volleyball. Today there are more than 46 million Americans who play volleyball.

-> **Me: What are the different positions?**

-> Volleybot: The three front row players are all in attacking positions. All lines denoting the boundaries of the team court and the attack zone are drawn or painted within the dimensions of the area and are therefore a part of the court or zone.

-> **Me: Volleyball doesn't seem that interesting to me**

-> Volleybot: Beach volleyball spread from California to Florida and then to other states—even to some that don't have beaches. 1912, YMCA spread volleyball to Uruguay

-> **Me: Exit**

-> Volleybot: Adios! It was good chatting with you, Sanjeev!

Returning User

Type "EXIT" or "exit" once you are done talking to the bot

-> Volleybot: Hey, my name is VolleyBot. Nice to meet you. What is your name?

Sanjeev Penupala, Varun Joshi

CS 4395.0W1

04/05/21

-> **Me: Sanjeev**

-> Volleybot: Welcome back, Sanjeev! Here is what I got from our last session!

Your Likes: [['points', 'need', 'win'], ['people', 'play', 'volleyball']]

Your Dislikes: [['volleyball', 'interesting']]

-> **Me: Cool you remembered that stuff**

-> Volleybot: He created the game of Volleyball (at that time called mintonette). Morgan needed a game that could be enjoyed by middle-aged men.

-> **Me: Who created volleyball?**

-> Volleybot: 1994, Volleyball World Wide, the first internet site on the sport of volleyball, was created. Today there are more than 46 million Americans who play volleyball.

-> **Me: What is the history of volleyball?**

-> Volleybot: The history of volleyball is closely linked to that of another popular court game. An antenna is placed on each side of the net perpendicular to the sideline and is a vertical extension of the side boundary of the court.

-> **Me: How do you serve the ball?**

-> Volleybot: These errors include back-row or libero players spiking the ball or blocking (back-row players may spike the ball if they jump from behind the attack line), players not being in the correct position when the ball is served, attacking the serve in the frontcourt and above the height of the net, using another player as a source of support to reach the ball, stepping over the back boundary line when serving, taking more than 8 seconds to serve, or playing the ball when it is above the opponent's court. The receiving team tries to return the serve, or hit the ball back over the net.

-> **Me: Size of the court?**

-> Volleybot: All lines denoting the boundaries of the team court and the attack zone are drawn or painted within the dimensions of the area and are therefore a part of the court or zone. An antenna is placed on each side of the net perpendicular to the sideline and is a vertical extension of the side boundary of the court.

-> **Me: Do you play volleyball?**

-> Volleybot: Today there are more than 46 million Americans who play volleyball. There are 800 million players worldwide who play volleyball at least once a week.

-> **Me: Volleyball seems more interesting now. Thanks!**

-> Volleybot: Today there are more than 46 million Americans who play volleyball. When used correctly, it is one of the more spectacular defensive volleyball plays.

-> **Me: Exit**

-> Volleybot: Adios! It was good chatting with you, Sanjeev!

Refer to Appendix 1 for knowledge base of chatbot's responses

Refer to Appendix 2 for user model

Chatbot Evaluation:

There are two main factors to consider how well a chatbot responds to a user's queries: consistency and relevance. We will be observing these two metrics from Volleybot.

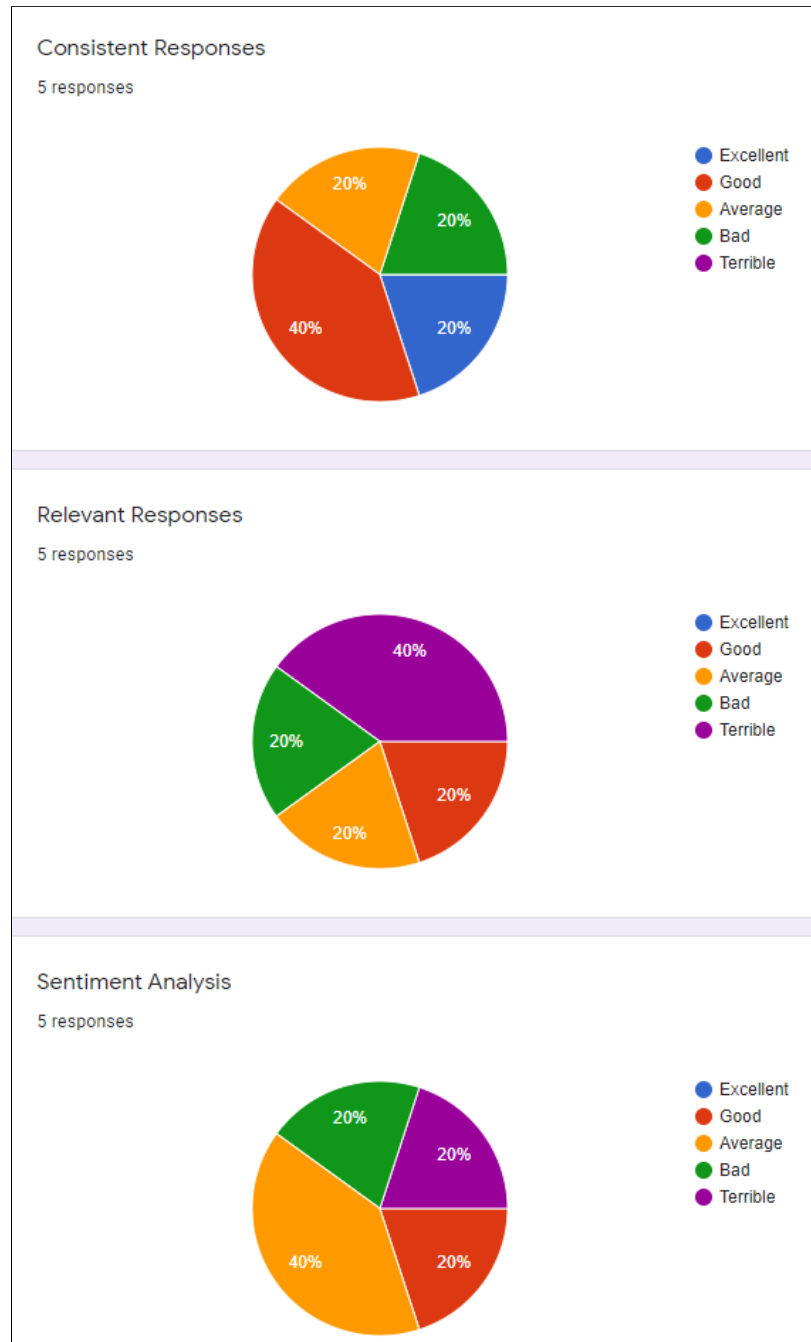
Taking a look at consistency, there is an if clause coded into the chatbot if there are not similar sentences, then it will return "I am sorry, but I do not understand." response. Looking just at the sample dialogs, there has not been an ignorant response from the chatbot, which makes it consistent in its responses, as it will try to find the most similar sentence possible to the user's queries. In essence, a user is almost guaranteed a response related to volleyball, regardless of their query.

In terms of relevancy, it is clear that not all queries are handled correctly, and in fact, some responses from Volleybot don't even answer the question at all! This is particularly not a fault from the curation of the knowledge base, but rather how responses are pulled from the knowledge base. Volleybot uses a naive approach of cosine similarity to find possible responses related to the user's queries, but neglects exactly what information the user might be asking. Because of this, not all responses will be relevant to the user's query, but may have a partial answer hidden in the response.

After analyzing these two metrics, we can confidently say having consistent responses to user's queries by Volleybot is a big strength, as even if the user may not get an answer related to their query, they will still consume information related to volleyball. Not always giving relevant responses to the user's queries is a major weakness that would need to be worked on, as pointed out earlier, it is not the fault of the size of the knowledge base, but rather how information is being pulled out from the knowledge base by Volleybot.

Another weakness of Volleybot is that it cannot distinguish between an actual query and a statement. As this chatbot was mainly designed to handle queries related to chatbot, it will still return a response to a statement, as the cosine similarity metric is indifferent towards what type of text it receives. A naive approach to fix this is to see whether the user's text ends in a period or a question, however, this won't scale in the long run, as most current generation users omit punctuation from casual conversation.

To support our findings with testimonials, we sent instructions to several people to have conversations with our chatbot, and asked them to fill out a feedback form, using Likert-scaling, with rubrics mentioned above. Here are the results of what the testimonials said:



As seen from the results above, it correlates very well with our findings of Volleybot, with the majority of people saying the consistency of responses was **good**, the relevance of responses being **terrible**, and the sentiment analysis for returning users is **average**.

Appendix 1: Knowledge Base

The volleyball knowledge is modeled by a *Python dictionary*, with key and value pairs. The key will be denoted as a single word on the left, followed by an indented list of sentences that contain the key word. Keep in mind certain sentences will appear under more than one key word.

attack

The three front row players are all in attacking positions.

All lines denoting the boundaries of the team court and the attack zone are drawn or painted within the dimensions of the area and are therefore a part of the court or zone.

ball

Volleyball, especially beach volleyball, is played in Australia, New Zealand, and throughout the South Pacific.

There are many variations on the basic rules of volleyball.

Today there are more than 46 million Americans who play volleyball.

Beach volleyball spread from California to Florida and then to other states—even to some that don't have beaches.

1912, YMCA spread volleyball to Uruguay

He created the game of Volleyball (at that time called mintonette).

1994, Volleyball World Wide, the first internet site on the sport of volleyball, was created.

The history of volleyball is closely linked to that of another popular court game.

There are 800 million players worldwide who play volleyball at least once a week.

When used correctly, it is one of the more spectacular defensive volleyball plays.

court

All lines denoting the boundaries of the team court and the attack zone are drawn or painted within the dimensions of the area and are therefore a part of the court or zone.

Sanjeev Penupala, Varun Joshi

CS 4395.0W1

04/05/21

The history of volleyball is closely linked to that of another popular court game.

An antenna is placed on each side of the net perpendicular to the sideline and is a vertical extension of the side boundary of the court.

libero

These errors include back-row or libero players spiking the ball or blocking (back-row players may spike the ball if they jump from behind the attack line), players not being in the correct position when the ball is served, attacking the serve in the frontcourt and above the height of the net, using another player as a source of support to reach the ball, stepping over the back boundary line when serving, taking more than 8 seconds to serve, or playing the ball when it is above the opponent's court.

line

All lines denoting the boundaries of the team court and the attack zone are drawn or painted within the dimensions of the area and are therefore a part of the court or zone.

An antenna is placed on each side of the net perpendicular to the sideline and is a vertical extension of the side boundary of the court.

morgan

Morgan needed a game that could be enjoyed by middle-aged men.

point

In 1917, the game was changed from requiring 21 points to win to a smaller 15 points to win.

A team has to win by at least two points.

Appendix 2: User Model

The user model in volleyball bot is modeled after a class and the users are all stored in a python list. The class has the fields name, liked_words, disliked_words, and keywords. At the end of the user session all the information gathered during the conversation is stored in a user object and before every session when the bot asks for a name, the bot checks if the user name already is in the list and if it is returns the likes and dislikes of the user that the bot has collected are displayed at the beginning.

names:

Sanjeev Penupala, Varun Joshi

CS 4395.0W1

04/05/21

The name of the user

likes_list

List of keywords from each line that the user shows a positive sentiment towards.

dislikes_list

List of keywords from each line that the user shows a negative sentiment towards.

keywords

The combined list of the likes and dislikes of each particular user that is used to personalize the bot according to the personalities of every particular user.

These characteristics are there for each particular user and they are appended to the Python List that stores these users. If the name is the same at the beginning then the bot picks out that value from the User list and displays it to the user.