

Chatbot Assignment

Chatbot Description:

As discussed in class, a chatbot is a conversational tool designed to efficiently perform routine tasks. It provides users with the ability to save time and have an efficient conversation that closely mimics human interaction. The chatbot created for this project was built from scratch, using the code from the previous assignment as a foundation. The code utilized various Python libraries, such as urllib, BeautifulSoup, nltk, and pickle, to extract information from the Vogue.com webpage and create a unique and professional chatbot.

The chatbot's code begins by defining the function `exact_links(url)`, which takes a URL as input and extracts both internal and external URLs from the Vogue webpage. These URLs are then stored in different domains. BeautifulSoup is used to parse the HTML of the web page and extract the URLs from anchor tags. The next function, `clean_text(text)`, is where the extracted text is cleaned by removing new lines and other unwanted characters. The cleaned text is then saved in a variable called "texts", and the top used terms are extracted and stored in a pickle file for future use.

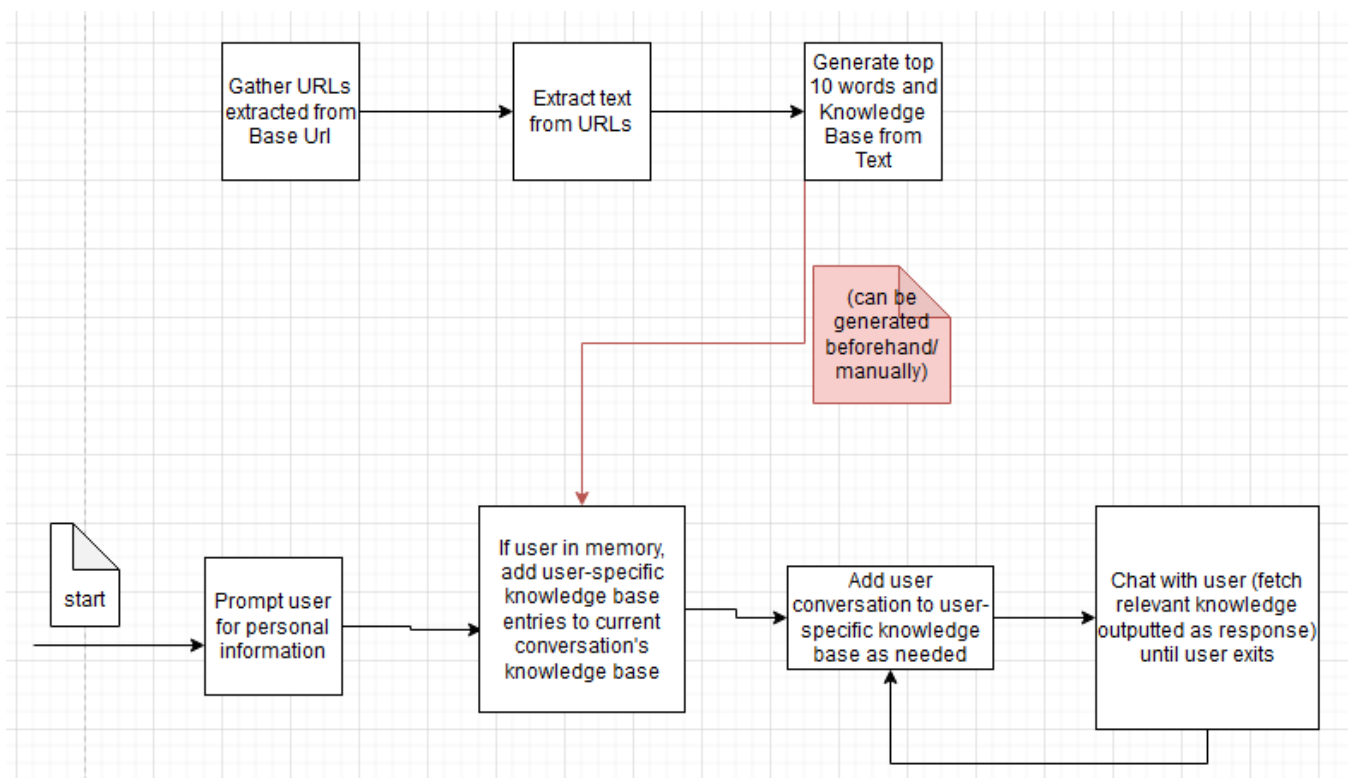
The chatbot utilizes various natural language processing (NLP) techniques, including tokenization, stopwords removal, and word counting. Tokenization is the process of splitting text into individual words or tokens, which is achieved using the `word_tokenize()` function from the nltk library in Python. Stopwords, which are common words like "the", "and", "in", etc., that do not carry much meaning, are removed from the text using the predefined list of stopwords provided by the nltk library for the English language. Word counting is performed using the Counter class from the collections module in Python, which counts the occurrences of each word in the text. This is used to identify the most common words, which are considered as top words and used for creating the knowledge base of the chatbot. The chatbot also stores the conversation

between the user and the chatbot, naming the file.txt by the user name, it asks the user for their name, their age, and if they like fashion or not. The user can enter any of the terms used in the knowledge base. Chatbot also display errors message such as if the user enter a string or character instead of an int when the chatbot asks for the user's age, same for yes/no questions, the chatbot will display an error message, and ask the user to enter yes or no responses.

Overall, the chatbot's code has been enhanced to be more unique and professional by incorporating additional details and explanations about the NLP techniques used, the libraries utilized, and the specific functions implemented for extracting, cleaning, and processing the text data from Vogue.com webpage.

Logic Diagram:

The following is a diagram of the overall logic of the code.



Sample Dialog Interaction:

This is a screenshot of sample dialog of the program where the user is chatting with the chatbot. First it prompts for the user specific information, and then proceeds to the chatbot functionality of the program.

```
Hello, my name is Chatbot Vogue. To whom am I speaking to today? Nadia
Nice to meet you, Nadia! How old are you? 23
Do you like fashion? (yes/no) no
ChatbotVogue: No problem. How can I help you today?
You: what is fashion
Chatbot: Fashion is a form of self-expression and autonomy at a particular period and place and in a specific context
Do you want to exit? (yes/no) no
Chatbot: Great! Any other vogue terms you would like to know about?
You: tell me about an artist
Chatbot: The Metropolitan Museum of Art in New York City has a Costume Institute that showcases fashion as art.
Do you want to exit? (yes/no) yes
Chatbot: My pleasure! It was nice talking to you. Goodbye!
```

Knowledge Base:

Examples of entries in the knowledge base used in the code. This knowledge base can be automatically generated using the text from the gathered links or generated and then cleaned to provide a better experience. This knowledge base can then be exported for use in other instantiations of the program or if you do not want to generate the knowledge base everytime.

1. Fashion:
 - a. Fashion is a form of self-expression and autonomy at a particular period and place and in a specific context.
 - b. Fashion consists of clothing, footwear, lifestyle, accessories, makeup, hairstyle, and body posture.
 - c. The term implies a look defined by the fashion industry as that which is trending.

2. Makeup:

- a. Makeup is a form of self-expression that has been around for thousands of years.
- b. Cosmetics such as lipstick or powder applied to the face, used to enhance or alter the appearance.
- c. Makeup mainly is used to change or enhance the way we look, to feel more confident and also to hide our imperfection.

3. Shades:

- a. Shades are a type of eyewear designed to protect the eyes from sunlight.
- b. Sunglasses can reduce the risk of developing cataracts and other eye diseases
- c. A shade is when a color remains its original hue but has been darkened.

[...]

User Models:

Here are two samples of user profiles that could have been created from the chatbot conversation. The name the user enters can be used to check if the user has previously used the chatbot before and access the userID. The UserID can then be used to access other personal information and user-specific information.

Users: {"Mustafa": "User1", "Afaf": "User2"}

- User1: {"Name": "Mustafa", "Age": "21", "Fashion": "Yes", "KnowledgeBaseID": "123"}
- User2: {"Name": "Afaf", "Age": "22", "Fashion": "Yes", "KnowledgeBaseID": "321"}

(where the knowledge base id could help the program reference the correct user-specific knowledge base to add the overall knowledge base)

Chatbot Evaluation:

Our approach to the creation of the chatbot as well as the extraction of information allows for different strengths and weaknesses. It is highly customizable as the data extraction is done by the code which is only provided a base URL which to start from. Although one website is given, the code extracts more urls and can do so extensively to find and store more information. The code performs a variety of NLP techniques to filter and process the data which can be beneficial in creating value from a variety of content which may be different from each other.. The user is also able to ask personalized questions related to the content of the training data so that the user can receive responses they may be more interested in (rather than the chatbot guiding the discussion based on the knowledge it already has). The chatbot can also store information about the user so that it can store and remember content from its conversation with the user and distinguish it from other users that use the chatbot.

Some weaknesses include that the base URL is hardcoded in the file (although this may also be seen as a positive as it is easy to change the content on which the chatbot is trained on). The domain is also, then, somewhat restricted to this domain as the additional URLs stem from this initial webpage. The collection of links is also limited to a hardcoded upper threshold but can be adjusted as needed (such as when scaling the size of the desired output knowledge base). Since the chatbot is relatively simple in NLP techniques used, it could also face challenges when faced with more complex inquiries from the user.

Overall, it is easy to interact with the chatbot as it mimics human interaction to provide an engaging experience for the user. By addressing the mentioned weaknesses, the chatbot could improve in its capabilities and become more skilled at responding to a variety of dialogue. The knowledge base could also be expanded by running the code

with the intent of targeting different domains and combining the knowledge bases that each instantiation generates