CORPUS WINES CHAT BOT

Corpus Wine Company Chatbot

1. Overall Approach

The Corpus Wine Company chatbot is designed to assist users by answering questions about the company's wine products. The chatbot leverages a pre-defined corpus of questions and answers, processing user queries to provide accurate and relevant responses. The overall approach includes the following steps:

1. **Preprocessing:** Tokenizes and normalizes the text of both the corpus and user queries by removing stopwords and non-alphanumeric characters.
2. **Vectorization:** Converts text data into numerical representations using TF-IDF (Term Frequency-Inverse Document Frequency) vectors.
3. **Similarity Matching:** Computes cosine similarity between the user's query and the corpus questions to determine the most relevant answer.
4. **Session Management:** Maintains conversation history using unique session IDs.

2. Frameworks/Libraries/Tools Used

- **NLTK (Natural Language Toolkit):** Used for text preprocessing tasks such as tokenization and stopword removal.
- **scikit-learn:** Employed for TF-IDF vectorization and cosine similarity calculations to match user queries with the corpus.
- **Streamlit:** Provides a web-based interface for users to interact with the chatbot.
- **uuid:** Generates unique session identifiers to manage and store user session data.

3. Problems Faced and Solutions

1. **Text Preprocessing Issues:**
    - **Problem:** Initial difficulties with tokenization and stopword removal.
    - **Solution:** Refined the preprocessing function to ensure accurate tokenization and removal of irrelevant words.
2. **Vectorization and Similarity Threshold:**
    - **Problem:** Determining an appropriate threshold for cosine similarity to filter out non-relevant responses.
    - **Solution:** Conducted experiments to set an optimal threshold that balances relevance and precision.
3. **Session Management:**
    - **Problem:** Ensuring reliable storage and retrieval of conversation history.
    - **Solution:** Implemented a local dictionary to manage session data and handle multiple user sessions effectively.

4. Future Scope:

1. **Expansion of Corpus:**
    - Adding more questions and answers to cover a broader range of topics related to the wine products.
2. **Advanced NLP Techniques:**
    - Integrating more sophisticated NLP models, such as BERT or GPT, to improve understanding and response accuracy.
3. **Feedback Mechanism:**
    - Implementing a feedback system where users can rate responses, allowing for continuous improvement of the chatbot.
4. **Backend Integration:**
    - Connecting the chatbot to backend systems to provide real-time information about wine inventory, availability, and promotions.
5. **Multi-Language Support:**
    - Adding support for multiple languages to cater to a global audience.