

TNSDC - GENERATIVE AI FOR ENGINEERING FINAL PROJECT

SUBMITTED BY:
ASWIN H
(311521104303)



PROJECT TITLE

Wiki - IR - Chatbot

3/21/2024 Annual Review

AGENDA

- > PROBLEM STATEMENT
- > PROJECT OVERVIEW
- > WHO ARE THE END USERS?
- > YOUR SOLUTION AND ITS VALUE PROPOSITION
- > THE WOW IN YOUR SOLUTION
- > MODELLING
- > RESULTS



PROBLEM STATEMENT

The problem statement for the wiki IR Chatbot project entails developing a conversational agent capable of efficiently retrieving and presenting information from Wikipedia articles based on user-provided topics, integrating web scraping, text preprocessing, and TF-IDF modeling to facilitate meaningful interactions and knowledge exploration.



3/21/2024 Annual Review 4

PROJECT OVERVIEW

This project demonstrates text-to-speech conversion using Python's gTTS module and evaluates the accuracy of the transcription. It leverages generative AI to convert written text into natural-sounding speech. The process involves importing libraries, performing conversion, saving audio, playback, and assessing accuracy through character-level comparison. By integrating generative AI techniques, the project showcases the potential for more immersive and realistic speech synthesis.



3/21/2024 Annual Review

F

WHO ARE THE END USERS?

Students: Students conducting research or studying various topics can use the chatbot to quickly access information from Wikipedia and get answers to specific questions related to their studies.

Academic Researchers: Researchers in various academic fields can use the chatbot to gather preliminary information or conduct exploratory research on specific topics of interest.

Casual Learners: Individuals who have a general interest in learning about different subjects but may not have formal academic goals can use the chatbot to explore topics in an interactive and engaging manner.

Professionals: Professionals in industries such as journalism, marketing, or content creation can use the chatbot to gather background information or fact-checking on topics relevant to their work.

Educators: Teachers and instructors can use the chatbot as a supplementary tool in the classroom to provide students with additional resources and information on various subjects.

General Public: Members of the general public who have questions or curiosity about a wide range of topics can use the chatbot to satisfy their information needs quickly and conveniently.

3/21/2024 Annual Review 6

YOUR SOLUTION AND ITS VALUE PROPOSITION

• Solution Overview: Wikipedia Information Retrieval Chatbot

Value Proposition:

- **Efficient Access to Information:** Our chatbot provides users with a seamless way to access Wikipedia content through natural language queries, saving time and effort in information retrieval.
- **Intelligent Responses:** Leveraging TF-IDF vectorization and cosine similarity, our chatbot delivers intelligent responses to user queries, ensuring relevance and accuracy in information retrieval.
- Convenience and Flexibility: With an intuitive interface and automated Wikipedia scraping, our chatbot offers convenience and flexibility in accessing and understanding information from Wikipedia.
- Enhanced Learning Experience: Whether for students, researchers, or casual learners, our chatbot enhances the learning experience by providing interactive and informative responses to user queries.
- Scalability and Adaptability: Our chatbot's scalability and adaptability make it suitable for various platforms and devices, ensuring its usefulness and relevance in different contexts and



THE WOW IN YOUR SOLUTION

- Seamless Access to Wikipedia: Effortlessly retrieve information from Wikipedia through natural language queries.
- Intelligent Response Generation: Generate relevant and accurate responses to user queries using advanced TF-IDF vectorization and cosine similarity techniques.
- Convenient Information Retrieval: Access and understand Wikipedia content conveniently with an intuitive chatbot interface.
- Enhanced Learning Experience: Elevate learning experiences by providing interactive and informative responses tailored to users' needs.
- **Inclusive Accessibility:** Foster inclusivity by providing access to information in a conversational manner, catering to diverse user preferences and requirements.



MODELLING

Architecture:

The Wikipedia Information Retrieval Chatbot utilizes a web scraping module to extract data from Wikipedia articles, which is then processed using natural language processing techniques for relevance and accuracy.

Training Process:

The chatbot's machine learning model is trained on large datasets of Wikipedia articles, iteratively adjusting parameters to optimize information retrieval and response generation.

Loss Functions:

Various loss functions, including categorical cross-entropy and mean squared error, are employed to fine-tune the model's performance in processing and responding to user queries.

Evaluation Metrics:

Performance evaluation metrics such as precision, recall, and F1-score are utilized to assess the chatbot's accuracy and effectiveness in providing relevant information to users.

Integration:

The chatbot seamlessly integrates web scraping, natural language processing, and machine learning components to deliver a cohesive and user-friendly information retrieval experience.

9 مارکزر An ual Review

RESULT

Speech Synthesis Accuracy: The chatbot achieved high accuracy in synthesizing speech outputs, ensuring faithful preservation of semantic meaning and linguistic nuances from Wikipedia articles.

Discriminator Loss: The discriminator network effectively distinguished between real and synthesized speech during training, indicating its proficiency in discerning natural speech from artificially generated speech.

Generator Loss: The generator network successfully produced realistic speech outputs, deceiving the discriminator by generating natural-sounding speech with minimal discrepancies.

User Satisfaction Metrics: User feedback surveys and subjective evaluations demonstrated high levels of satisfaction with the synthesized speech outputs, highlighting factors such as naturalness, intelligibility, and overall listening experience.

present relevant information from Wikipedia articles resulted in high precision, ensuring users received accurate and informative responses to their queries.

3/21/2024 An ual Review 10