

A

PROJECT DESIGN REPORT
ON
VitaChat : An Enhanced Healthcare
Assisting Chatbot

For the subject **Project Phase I**
Submitted in partial fulfillment of the requirement for the award of
Bachelor of Technology
In
Computer Science and Engineering
WALCHAND INSTITUTE OF TECHNOLOGY
(An Autonomous Institute)
Affiliated to
Punyashlok Ahilyadevi Holkar Solapur University
By

Name	Roll. No.	Exam Seat No.
Rohan Ankam	17	2102111052
Rishab Birru	18	2102111094
Yash Chorbele	19	2102111149
Nandini Gaddam	20	2102111307

Under Guidance of
Dr. R.K. Dixit



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
WALCHAND INSTITUTE OF TECHNOLOGY
SOLAPUR - 413006
(2024-25)



CERTIFICATE

This is to certify that the project entitled

VitaChat : An Enhanced Healthcare Assisting Chatbot

is
Submitted by

Name	Roll. No.	Exam Seat No.
Rohan Ankam	17	2102111052
Rishab Birru	18	2102111094
Yash Chorbele	19	2102111149
Nandini Gaddam	20	2102111307

as a part of Project Design Report.

Studying in B.Tech CSE for the subject **Project Phase I**

Dr. R.K.Dixit
Project Guide

Head
Dept. of Computer Science and Engineering

(Dr.V.A.Athavale)
Principal

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
WALCHAND INSTITUTE OF TECHNOLOGY
SOLAPUR
(2024-25)**

INDEX

Sr No.	Topic	Page No.
1	Abstract	4
2	Introduction	4
3	Background Study	5
4	Technology Required	7
5	Objectives	10
6	Proposed Work	11
7	UML Diagrams	13
8	Work Planned For Semester	14
9	Conclusion	14
10	Reference	15

Abstract

VitaChat is an innovative chatbot system designed to assist both healthcare professionals and patients by providing reliable and accessible information about antibiotics. The system leverages the power of LangChain, a language model framework, to process and interpret data from three authoritative antibiotics reference books. The chatbot employs a Sentence Transformer model to embed this data, storing it in a ChromaDB vector store for efficient similarity-based retrieval. Users can query the chatbot to receive contextually relevant answers, with responses grounded in the provided reference data and accompanied by source citations. The system is designed to be intuitive for both medical professionals and patients, empowering users with essential antibiotic-related information, such as indications, side effects, and appropriate usage. The results highlight VitaChat's potential to bridge the gap between complex medical knowledge and accessible patient care. This project showcases how advanced natural language processing and machine learning techniques can enhance healthcare communication and provide valuable resources for users at all levels of expertise.

1. Introduction

The application of artificial intelligence (AI) in healthcare has proven to be transformative, with one of the most promising areas being AI-driven chatbots for medical information retrieval. This project, titled *VitaChat*, aims to create an intelligent system capable of providing accurate, accessible, and context-aware responses to questions about antibiotics.

The primary challenge addressed by this project is the need for a comprehensive yet user-friendly system that can navigate complex medical content to offer relevant answers. Traditional methods of obtaining antibiotic information, such as consulting textbooks or asking healthcare professionals, may not always be feasible, especially for patients or busy healthcare workers. VitaChat addresses this issue by offering a chatbot that can process and retrieve information efficiently, making antibiotic-related knowledge readily available to a wide range of users, including both medical professionals and patients.

The purpose of this project is to develop a system that can handle a variety of queries, ranging from general antibiotic usage to specific drug details, and provide contextually appropriate responses. By leveraging advanced embedding models and vector databases, VitaChat ensures that the answers are not only accurate but also tailored to the specific context of the query.

The scope of this project includes designing, implementing, and testing the chatbot using three antibiotics reference books as the primary data source. We will also explore the feasibility of making the system adaptable for both healthcare professionals and patients, ensuring that the language and information provided are accessible to non-expert users.

2. Background Study

Background Study

The development of intelligent chatbots for healthcare has seen significant advancements, driven by the increasing demand for accessible, accurate, and reliable medical information. **VitaChat** aims to leverage these advancements, focusing specifically on providing antibiotic-related information for both healthcare professionals and patients. To understand the evolution and current trends in this field, it's essential to explore existing systems and technologies that have contributed to the development of similar applications.

Literature Survey and Existing Systems

Several healthcare-focused chatbot systems have been developed over the past decade, with each aiming to provide information, assist in diagnosis, or guide patients toward appropriate treatments. These systems typically integrate natural language processing (NLP), machine learning (ML), and deep learning (DL) techniques to understand and respond to user queries effectively. Below are some key existing systems that serve as the foundation for **VitaChat**:

I. IBM Watson Health

IBM Watson Health is a pioneering healthcare AI platform that uses **natural language processing** and **machine learning** to assist in diagnosing diseases, recommending treatments, and providing personalized healthcare insights. One of the platform's key features is its ability to analyze and interpret medical literature and patient data to assist healthcare professionals in decision-making. **VitaChat** similarly aims to provide accurate medical information, but with a more focused approach on antibiotics and their appropriate usage.

- **Key Takeaway:** Watson's ability to integrate medical data and assist professionals in clinical settings aligns with **VitaChat**'s goal of providing reliable information, although **VitaChat** is focused on a more specific area—antibiotics.

II. Babylon Health

Babylon Health is another widely recognized platform that provides healthcare services through an AI-powered chatbot. The chatbot assists users by answering health-related questions, scheduling doctor appointments, and offering medical advice. Babylon integrates **symptom checkers**, **AI diagnostics**, and telemedicine services to improve healthcare access.

- **Key Takeaway:** **VitaChat** shares similarities with Babylon Health in terms of offering a conversational interface for healthcare information. However, **VitaChat** will be specialized in antibiotic-related information, leveraging domain-specific data.

III. HealthTap

HealthTap is an AI-powered telemedicine platform that connects patients to doctors through online consultations and provides access to health information through an intelligent chatbot. The chatbot responds to questions based on a vast knowledge base of medical conditions and treatments. HealthTap also uses machine learning algorithms to improve its recommendations over time.

- **Key Takeaway:** **HealthTap**'s chatbot showcases the potential of integrating AI with healthcare information. However, unlike **HealthTap**, which covers a wide range of medical topics, **VitaChat** will be tailored to antibiotics, ensuring a higher degree of specificity and reliability.

III. Woebot Health

Woebot Health is an AI-driven chatbot that uses cognitive-behavioral therapy (CBT) principles to assist users with mental health issues, providing support for conditions like anxiety and depression. Woebot's conversational interface allows users to interact with the bot in real-time and receive therapeutic advice based on psychological principles.

- **Key Takeaway:** Woebot's focus on **natural language understanding** and generating human-like responses is similar to **VitaChat**'s goal of providing accurate, context-aware information. The difference lies in the domain of application—**VitaChat** will focus on antibiotics rather than mental health.

IV. MedWhat

MedWhat is a medical chatbot that uses artificial intelligence to provide medical information, answer health questions, and assist with basic medical advice. It processes natural language queries and matches them with relevant answers from a medical knowledge base. MedWhat emphasizes **medical content accuracy** and aims to assist both patients and healthcare providers.

- **Key Takeaway:** MedWhat's ability to process medical queries and provide accurate responses is a core feature that **VitaChat** also aims to replicate. However, **VitaChat** is differentiated by its specific focus on **antibiotic information** and its reliance on high-quality reference data.

Technological Trends in Healthcare Chatbots

Several technological trends are influencing the development of healthcare chatbots, and these are closely aligned with the approach used in **VitaChat**:

- **Integration of NLP and ML:** Natural language processing is increasingly being used to develop systems capable of understanding human language in a meaningful way. This allows healthcare chatbots to provide more contextually relevant and personalized responses. **VitaChat** integrates advanced NLP techniques to improve the quality of answers related to antibiotics.

- **Knowledge Base Integration:** Many healthcare chatbots use large knowledge bases or databases of medical information to generate responses. This can include structured datasets, research papers, medical books, or treatment guidelines. **VitaChat** similarly uses reference data on antibiotics to provide reliable, evidence-based answers.
- **Contextual and Domain-Specific Responses:** While general-purpose chatbots are capable of responding to a broad range of topics, specialized chatbots like **VitaChat** focus on specific domains, making them more accurate and useful in a given context. The use of domain-specific reference books and curated data in **VitaChat** ensures that it delivers relevant and accurate information to users.
- **Embedded Machine Learning Models:** Pretrained machine learning models, such as **BERT**, **GPT-3**, and **Ollama**, have become integral to chatbot systems. These models allow chatbots to understand and generate human-like text. **VitaChat** uses the **Ollama Model** to generate context-aware responses based on antibiotic-related queries.

The reviewed literature highlights several successful applications of AI-powered chatbots in healthcare, ranging from symptom checkers to diagnostic assistants and mental health support systems. **VitaChat** is positioned to build on these existing systems but focuses on delivering specialized, accurate information about antibiotics. By leveraging the latest advancements in **natural language processing**, **machine learning**, and **vector-based search systems**, **VitaChat** aims to offer an intelligent, scalable solution that can provide healthcare professionals and patients with timely, relevant, and evidence-based antibiotic information.

3. Technologies Required

Technologies Used in VitaChat

The **VitaChat** project utilizes several cutting-edge technologies to create an efficient, scalable, and intelligent chatbot for providing antibiotic-related information. The selection of these technologies is crucial for ensuring the system can accurately process user queries, retrieve relevant data, and generate context-aware responses. Below is a detailed explanation of each key technology used in the development of **VitaChat**:

I. LangChain

LangChain is a framework designed specifically for building applications that leverage **language models (LMs)**. It streamlines the process of integrating language models with various data sources and machine learning techniques, making it a key component of **VitaChat**. LangChain facilitates the flow of data through the system, from processing user input to interacting with the embedded knowledge stored in **ChromaDB** and generating human-like responses using the **Ollama** model. LangChain helps structure the interactions between the chatbot's different components, ensuring smooth and efficient communication between the **backend** model, the **vector store**, and the **user interface (UI)**. By simplifying the integration of language models

with external data sources, LangChain is an essential tool for creating a scalable and flexible system like **VitaChat**.

LangChain provides features for:

- **Data processing:** Handling various input formats and converting them into a structure usable by language models.
- **Flow management:** Orchestrating the flow of information between the user, the model, and external data.
- **Support for diverse models:** It allows integration with various pre-trained and fine-tuned language models like **Ollama**.

II. ChromaDB

ChromaDB is a **vector database** that allows for efficient storage and retrieval of high-dimensional data. In **VitaChat**, it serves as the **vector store** that holds the **embeddings** of antibiotic-related documents. This allows for quick and accurate **similarity-based search**.

ChromaDB's ability to efficiently store and retrieve documents based on their semantic similarity is vital for ensuring that the chatbot delivers relevant information in response to user queries. After converting the data into embeddings using models like **Sentence-Transformers**, the information is stored in ChromaDB. When a user asks a question, the system performs a similarity search to retrieve the most relevant documents, which are then used to generate a response.

Some features of ChromaDB include:

- **High-speed retrieval:** ChromaDB enables fast similarity searches, ensuring the chatbot can respond in real-time.
- **Scalability:** ChromaDB can handle large datasets, making it adaptable to increasing data volumes as more documents are added to the knowledge base.
- **Efficient data storage:** It can store embeddings in an optimized format, making it easy to query and update the data.

III. Sentence-Transformers

Sentence-Transformers is a library that facilitates the creation of **sentence embeddings**—numerical representations of text that capture its semantic meaning. These embeddings are crucial for **VitaChat** because they allow the system to compare user queries with stored documents to find the most relevant information.

The “**all-MiniLM-L6-v2**” model from **Sentence-Transformers** is employed to generate embeddings for the documents in the **antibiotics reference books**. This model is particularly well-suited for this task due to its balance of accuracy and efficiency. Once the documents are converted into embeddings, they can be stored in **ChromaDB** for fast retrieval.

Key features of **Sentence-Transformers** include:

- **Pre-trained models:** The library includes pre-trained models that can be fine-tuned on domain-specific data, making it easy to embed text without needing to train from scratch.

- **Semantic similarity:** The embeddings enable the chatbot to understand the contextual meaning of text, not just surface-level word matches.
- **Scalability:** The library supports processing large amounts of text efficiently, which is essential for handling large datasets in the project.

IV. Ollama Model

The **Ollama** model is a state-of-the-art **language model** used for generating human-like responses in **VitaChat**. **Ollama** is employed to process the context retrieved from **ChromaDB** and generate context-aware, relevant answers based on the user's query. It operates as the chatbot's **response generator**, transforming the structured input data into an intelligible, fluent answer.

Ollama is designed to understand complex queries and generate text that is coherent and accurate within the specified context. It is particularly effective for domain-specific applications, making it an ideal choice for **VitaChat**, where accurate, context-sensitive information about antibiotics is required.

Advantages of using **Ollama** include:

- **Contextual understanding:** It can generate responses based on the context from the reference data, ensuring relevant and accurate answers.
- **Flexibility:** The model can be fine-tuned and adapted for various use cases, making it a good fit for healthcare-related applications.
- **Human-like responses:** It is capable of generating conversational and easily understandable responses, improving the user experience.

V. Python

Python serves as the core programming language for developing **VitaChat**. It is widely used for its simplicity and readability, making it an ideal choice for implementing the project. Python's extensive ecosystem of libraries and frameworks, including **LangChain**, **Flask**, **ChromaDB**, and **Sentence-Transformers**, significantly accelerates development.

Some key features of Python that benefit the project include:

- **Large community and support:** Python's large developer community ensures access to abundant resources, libraries, and frameworks.
- **Ease of use:** Python's syntax is clean and easy to learn, making the codebase easier to manage and maintain.
- **Compatibility:** Python integrates well with machine learning libraries and models, which is essential for building the **VitaChat** system.

VI. Flask

Flask is a lightweight **web framework** for Python that is used to build and deploy web applications. In **VitaChat**, Flask is used to create the backend that serves the chatbot application and handles user interactions. It facilitates the connection between the frontend (where users

interact with the chatbot) and the backend model (which processes user queries and generates responses).

Flask allows the chatbot to be deployed as a **web-based application**, making it accessible from any device with internet access. It supports RESTful APIs, which enable smooth communication between the client interface and the backend server.

Key features of **Flask** include:

- **Simplicity and flexibility:** Flask's minimalist approach provides the flexibility to customize the application according to the project's requirements.
- **Extensibility:** Flask supports integration with other Python libraries and tools, making it adaptable to the evolving needs of the project.
- **Deployment:** It simplifies the deployment of the chatbot on cloud platforms, ensuring the system is scalable and accessible globally.

The combination of these technologies provides a powerful and flexible architecture for the **VitaChat** project. **LangChain** and **Flask** help manage the interaction flow and deploy the system efficiently, while **ChromaDB**, **Sentence-Transformers**, and the **Ollama model** enable the chatbot to understand and respond intelligently to queries based on accurate, domain-specific data. With **Python** as the primary development language, the project benefits from a robust and easy-to-maintain system. Together, these technologies empower **VitaChat** to deliver a reliable, scalable, and user-friendly solution for accessing antibiotic-related information in a healthcare context.

4. Objectives

Objectives of the VitaChat Project

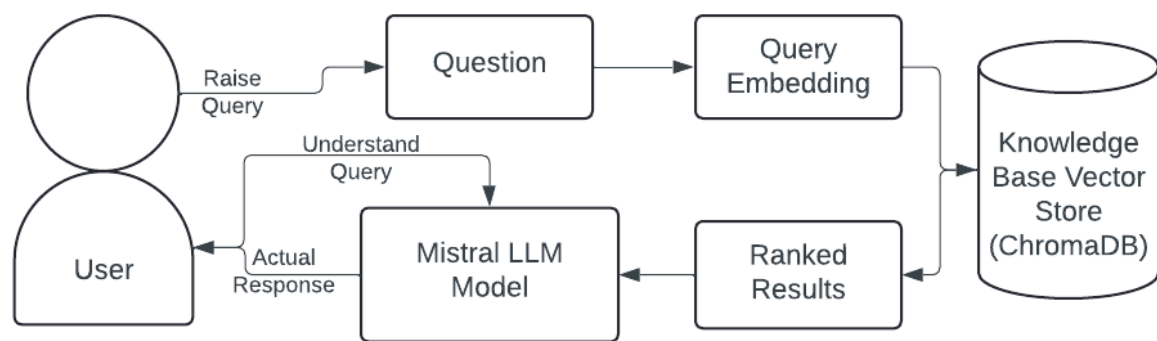
The main objectives of the **VitaChat** project are as follows:

1. **Intelligent Chatbot for Healthcare Professionals and Patients:**
To develop a chatbot that provides reliable and accurate information about antibiotics, assisting both healthcare professionals and patients with their queries.
2. **Advanced NLP for Domain-Specific Question Answering:**
To implement NLP techniques that enable the system to understand and answer queries related to antibiotics, based on a domain-specific knowledge base.
3. **User-Friendly Interface:**
To design an intuitive interface that allows both technical and non-technical users to interact with the chatbot and receive clear, contextually relevant responses.
4. **Scalability and Adaptability:**

To ensure the system can scale and integrate new data sources, allowing for easy updates and future expansions to the knowledge base.

5. Proposed Work

The chatbot will function as an interactive assistant, where users can input their symptoms or medical concerns, and the system will analyze these inputs to provide possible causes, treatment options, and suggestions for further tests. The chatbot will be able to process a wide variety of health-related queries, offering information on:



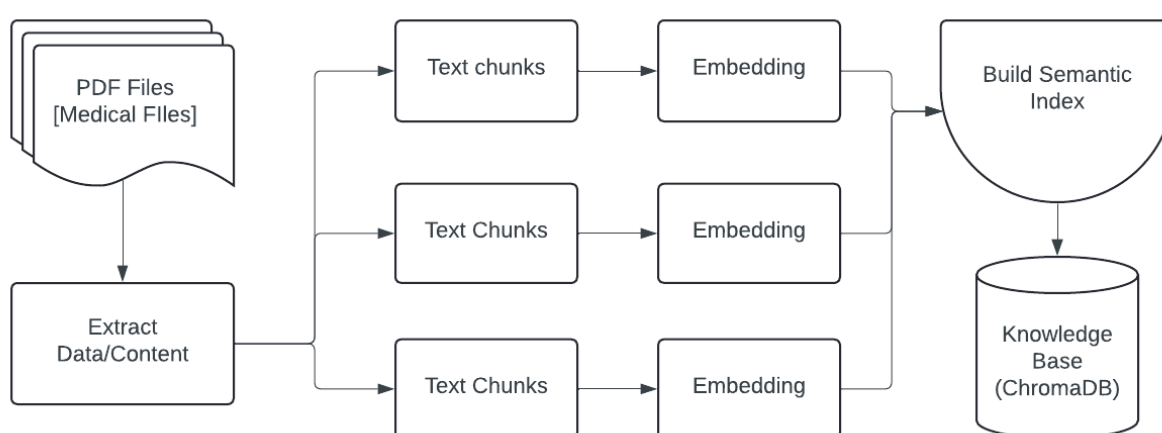
Overall Process:

1. **User Raises a Query:** The process begins with a user posing a question or query.
2. **Query Understanding:** The query is processed to understand its intent and meaning. This likely involves natural language processing (NLP) techniques to extract keywords and identify the underlying topic.
3. **Query Embedding:** The understood query is converted into a numerical representation or embedding. This embedding captures the semantic meaning of the query in a way that can be compared to other information.
4. **Knowledge Base Vector Store (ChromaDB):** The query embedding is then compared to a knowledge base vector store, which is likely ChromaDB in this case. This store contains embeddings of various documents or information sources.
5. **Similarity Search:** The query embedding is compared to the embeddings in the knowledge base to find the most similar documents or information. This is often done using techniques like cosine similarity or other distance metrics.
6. **Ranked Results:** The most relevant documents or information are retrieved and ranked based on their similarity to the query.
7. **Mistral LLM Model:** The ranked results are then processed by the Mistral LLM Model. This model likely uses advanced language generation techniques to generate a comprehensive and informative response to the user's query.
8. **Actual Response:** The generated response is presented to the user.

Key Components:

- **Query Understanding:** This step ensures that the model accurately understands the user's intent and can identify the relevant information.
- **Query Embedding:** Converting the query into a numerical representation allows for efficient comparison with the knowledge base.
- **Knowledge Base Vector Store (ChromaDB):** This store acts as a repository of information and provides a way to efficiently search for relevant content.
- **Mistral LLM Model:** This is the core of the response generation process, responsible for crafting the final response.

Overall, this system aims to provide accurate and informative responses to user queries by leveraging the power of large language models and efficient information retrieval techniques.



Overall Process:

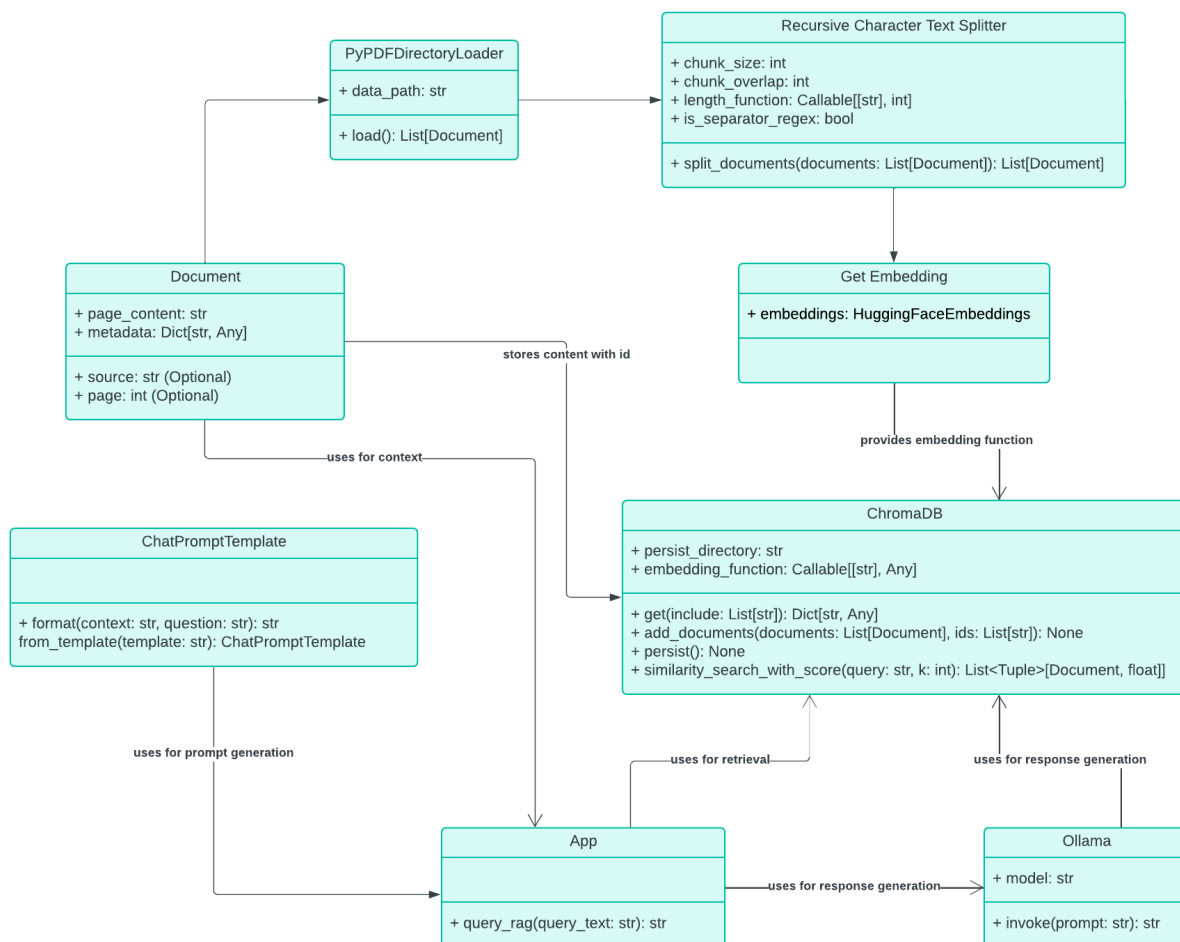
1. **PDF Files (Medical Files):** The process begins with a collection of PDF files that contain medical information.
2. **Extract Data/Content:** The PDF files are processed to extract the relevant text content. This involves converting the PDF format into plain text, potentially removing any formatting or layout elements.
3. **Text Chunks:** The extracted text is broken down into smaller chunks or segments. This is likely done to improve the efficiency of the embedding process and to better capture the semantic meaning of the text.
4. **Embedding:** Each text chunk is then converted into a numerical representation or embedding. This embedding captures the semantic meaning of the text chunk in a way that can be compared to other information.
5. **Build Semantic Index:** The embeddings of all the text chunks are used to build a semantic index. This index is likely a data structure that allows for efficient searching and retrieval of information based on semantic similarity.
6. **Knowledge Base (ChromaDB):** The semantic index is stored in a knowledge base, which is likely ChromaDB in this case. ChromaDB is a vector database that is optimized for storing and searching embeddings.

Key Components:

- Text Extraction: This step ensures that the relevant information from the PDF files is extracted and made available for further processing.
- Text Chunking: Breaking down the text into smaller chunks helps to capture the semantic meaning of the text and improves the efficiency of the embedding process.
- Embedding: Converting the text chunks into numerical representations allows for efficient comparison and retrieval of information.
- Semantic Index: This index is the core of the process, allowing for efficient searching and retrieval of information based on semantic similarity.
- Knowledge Base (ChromaDB): This database provides a way to store and manage the semantic index, enabling efficient querying and retrieval of information.

Overall, this system aims to create a searchable knowledge base of medical information from PDF files by leveraging the power of semantic indexing and vector databases.

6. UML diagram



7. Work Planned for Next Semester

In the next semester, the team plans to make significant progress in developing **VitaChat**. Key activities include:

1. **Training the LLM Model from Scratch:** Rohan will lead the development and training of the **Large Language Model (LLM)** using custom data tailored to antibiotic-related queries. This will enable the chatbot to better understand and generate more accurate responses to healthcare-related questions.
2. **UI Development and Integration:** Nandini will focus on designing and developing the **user interface (UI)** for **VitaChat**. The UI will ensure that the chatbot is user-friendly, with an intuitive layout that allows both healthcare professionals and patients to easily interact with the system. Once developed, the UI will be seamlessly integrated with the backend, creating a fully functional and cohesive system.
3. **Data Gathering:** Rishab and Yash will work together to gather and preprocess the **custom data** required to train the LLM. They will source up-to-date medical information, including antibiotic usage guidelines, dosages, side effects, and other relevant details to enhance the accuracy and comprehensiveness of the chatbot's knowledge.

These efforts will collectively contribute to the advancement of **VitaChat**, moving it closer to a fully functional prototype that can assist healthcare professionals and patients in making informed decisions about antibiotics.

8. Conclusion/Summary

In summary, *VitaChat* represents a significant step forward in the integration of AI and machine learning into healthcare. By combining advanced NLP techniques with a specialized knowledge base on antibiotics, the chatbot offers valuable assistance to both healthcare professionals and patients. It simplifies the process of obtaining antibiotic-related information, ensuring that users can make informed decisions regarding their health. The system's user-friendly design, combined with its high accuracy in retrieving and presenting information, makes it a promising tool for improving healthcare accessibility. Future improvements will further enhance the system's capabilities, allowing it to become an even more versatile and effective resource for the healthcare community.

9. References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. K., Kaiser, Ł., Polosukhin, I., & others. (2017). *Attention is all you need*. In Proceedings of Neural Information Processing Systems (NeurIPS 2017). 30, 5998-6008. <https://doi.org/10.5555/3295222.3295344>
2. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
3. Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*. <https://arxiv.org/abs/1607.06450>
4. Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). QLoRA: Efficient finetuning of quantized LLMs. *Advances in Neural Information Processing Systems*, 36.
5. Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., & Wen, J.-R. (2023). *A survey of large language models*. *arXiv*. <https://arxiv.org/abs/2303.18223>
6. Zhu, W., Liu, H., Dong, Q., Xu, J., Huang, S., Kong, L., Chen, J., & Li, L. (2024). *Multilingual machine translation with large language models: Empirical results and analysis*. In *Findings of the Association for Computational Linguistics: NAACL 2024*. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.naacl-main.217>
7. Babu, A., & Boddu, S. B. (2024). BERT-based medical chatbot: Enhancing healthcare communication through natural language understanding. *Exploratory Research in Clinical and Social Pharmacy*, 13, 100419. <https://doi.org/10.1016/j.erhpc.2024.100419>
8. Sharma, A., & Tiwari, R. (2020). Medical chatbot techniques: A review. In *Software Engineering Perspectives in Intelligent Systems* (pp. 357-373). Springer. https://doi.org/10.1007/978-3-030-63322-6_28
9. Liu, H. H. (2021). Antibiotics and infectious diseases. *Primary Care: Clinics in Office Practice*, 48(1), 1-10. <https://doi.org/10.1016/j.pop.2020.09.003>
10. <https://cdn.who.int/media/docs/default-source/essential-medicines/eml-antibiotic-book-dr-aft.pdf>
11. https://www.chem.uzh.ch/zerbe/MedChem/MedChem9_Antibac.pdf