



Disease Diagnosis.

Mandate 1

Brief Description

This project addresses the challenge of identifying and understanding early symptoms of diseases, a common struggle for many individuals.

Early diagnosis is crucial for effective treatment, and our aim is to assist patients in diagnosing diseases in the early stages. Not everyone possesses the medical knowledge to recognize the significance of their symptoms, which can result in delayed or inadequate medical care. Leveraging Natural Language Processing (NLP), our project seeks to create an “Disease Diagnosis Support System”.

This system will assist users in comprehending their symptoms, assess the potential severity, and provide information from trusted sources to guide users in seeking appropriate medical attention.

Example:

Imagine waking up with a headache and a sore throat, uncertain if it's a common cold or something more serious. Our project serves as a virtual health advisor, listening to your symptoms and offering reliable information to guide you on whether you need rest or should consult a doctor.

Objective

The significance of early diagnosis cannot be overstated, profoundly impacting medical outcomes. Early detection allows for prompt intervention, diminishing complications, and amplifying treatment efficacy.

Despite this importance, many individuals struggle to interpret their symptoms accurately, leading to delayed or insufficient medical care. The project aims to empower individuals by providing accessible and reliable information, fostering informed decision-making for better healthcare outcomes.

Key Points:

Crucial Role of Early Diagnosis:

Early diagnosis shapes the trajectory of medical conditions.

Timely Intervention:

Early detection allows for timely intervention, diminishing complications and amplifying treatment efficacy.

Problem Faced by Patients:

Many struggle with accurately interpreting their symptoms.

Consequences of Lack of Interpretation:

The inability to interpret symptoms leads to delayed or inadequate medical care.

Informed Decision-Making:

Empowered individuals are better equipped to make informed decisions about their health.

Contributing to Better Healthcare Outcomes:

Ultimately, the project contributes to superior healthcare outcomes by emphasizing early diagnosis and fostering informed decision-making.

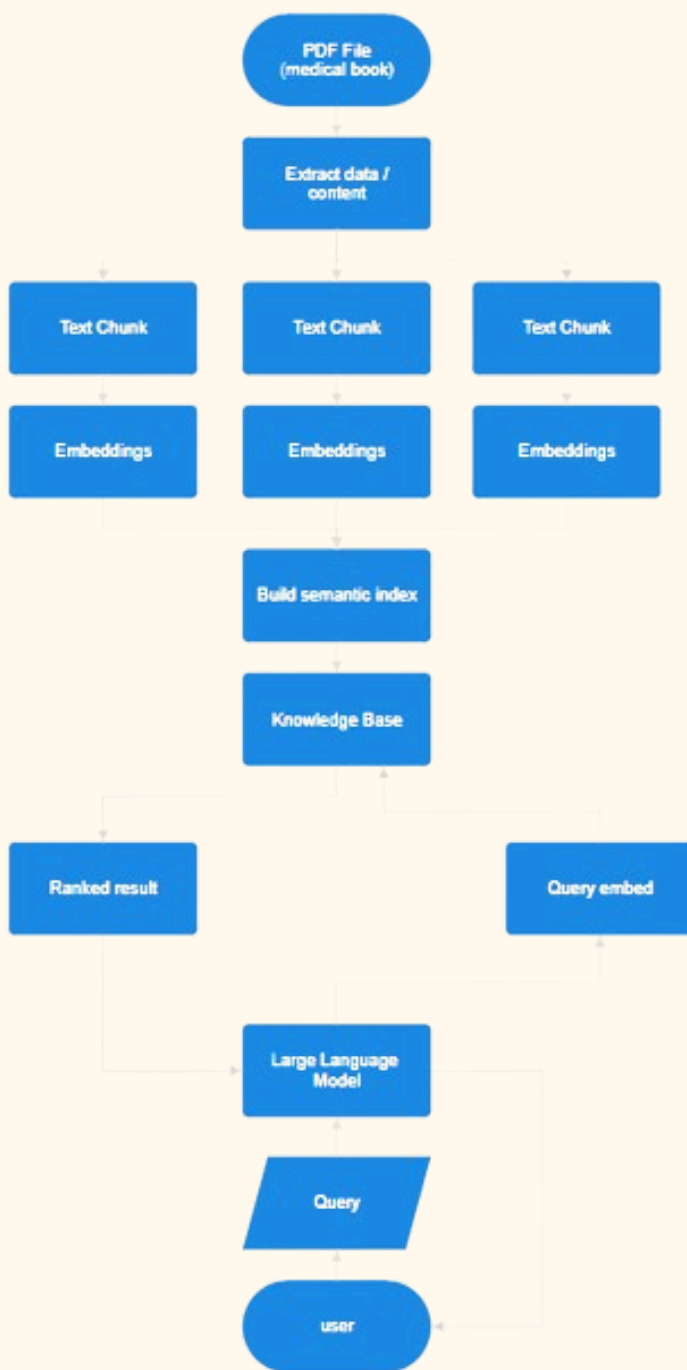
Tech Stack

1. **Langchain**
2. **SpacyEmbeddings**
3. **OpenAIEmbeddings**
4. **Chroma**
5. **PyPDF**
6. **React**

Workflow

1. Data Collection:
 - a. Gather a large dataset of medical documents in PDF format. These documents should contain information on various diseases, symptoms, treatments, etc. Ensure that the sources are reliable and trustworthy.
2. Data Extraction:
 - a. Implement a PDF parsing module to extract text data from the PDF documents.
 - b. Clean and preprocess the extracted text to remove noise and irrelevant information.
3. Text Chunking:
 - a. Segment the preprocessed text into smaller meaningful units or "chunks". This could be done using techniques such as sentence segmentation or keyword extraction.
4. Vector Embedding:
 - a. Utilize pre-trained word embeddings or train custom embeddings on the text chunks to represent them as numerical vectors. This step is essential for capturing the semantic meaning of the text.

5. Semantic Indexing:
 - a. Build a semantic index using the vector embeddings. This index should efficiently map queries to relevant text chunks based on semantic similarity.
6. Knowledge Base Creation:
 - a. Organize the indexed text chunks into a knowledge base structure. Each entry in the knowledge base should contain relevant information about diseases, symptoms, and treatments.
7. User Query Processing:
 - a. Receive user queries in English language. Embed the user query and try to get relevant chunk from the knowledge base
8. Reranking:
 - a. Implement a reranking mechanism to prioritize and refine the retrieved text chunks.
9. Language Model Evaluation:
 - a. Experiment with multiple language models to find the one that provides the most accurate and relevant responses to user queries.



Future Scope

1. Prescription/Report summarization.
2. Medicine suggestion.
3. Disease severity classification.

Mandate 2

We are focusing on Data Collection by gathering diverse medical documents, Data Extraction by parsing and preprocessing PDF text, Text Chunking by segmenting the preprocessed text, and Vector Embedding by converting text chunks into numerical representations for semantic analysis.

Mandate 3

We'll construct the complete system and experiment with various embedding models and language models. Based on performance, we'll conduct further fine-tuning. Tasks include Semantic Indexing, Knowledge Base Creation, User Query Processing, Reranking, and Language Model Evaluation.

Mandate 4

Developing the UI(Front-End) section of the project.

Integrating the model to feed UI about the report analysis & representing it in a simplified manner.

Implement Word Sense Disambiguation techniques to interpret the intended meaning of words in context, reducing ambiguity in health-related content.

Team & Roles.

Somesh Awasthi (MT2023172) -

- Implement the PDF parsing module to extract text data from PDF documents, alongside performing all the necessary preprocessing.
- Generate embeddings for text chunks to represent them as numerical vectors, ensuring proper integration of embeddings to capture semantic meaning in the text..
- Integrate word embeddings and semantic similarity measures.
- Train and validate Large learning models using relevant datasets.

Karnadevsinh Zala (MT2023188) -

- Integration of the model with frontend UI & managing APIs.
- Evaluate the performance of language models and select the most suitable one for the system.
- Retrieve and process user queries. Implement embedding techniques to retrieve relevant chunks from the knowledge base.
- Set up and manage databases for storing medical ontologies, knowledge graphs, and user data securely.