International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Semantic Search - Based Medicine Recommender with LangChain Google Search Agent

Aakash Walavalkar[1], Jay Ajmera[2], Kallind Soni[3]

[1]Data Scientist & Founder, Triumph, AI

[2]GenAI & NLP Practitioner, Student, NMIMS

[3]GenAI & NLP Practitioner, Student, NMIMS

Abstract: ***In the field of medicine, timely and accurate medication suggestions are critical to ensuring that patients receive top - quality care. This paper leverages the functionality of LangChain agents along with advanced semantic search algorithms to address the challenge of finding relevant medications for specific illnesses. Here, we introduce an innovative architecture, the "Semantic Search - Based Medicine Recommender with LangChain Google Search Agent. " Our system makes use of a large dataset of approximately 23, 000 drugs. Our pipeline includes the steps of data scraping and structuring, NLP sentence vectorization, and 1024 - dimensional embeddings using the SBERT model. Because these embeddings are stored in a Neo4j graph database, effective retrieval is possible due to cosine similarity calculations. Through the use of a chatbot as the interface, the right medicines can be suggested in response to a question by users, and simple interaction is feasible with the user. Our experimentation showcases remarkable improvement over the traditional keyword - based search approach in both the accuracy of suggestions as well as precision and recall. The provided outcomes point out how the integration of LangChain agents with semantic search algorithms will enhance the efficiency and robustness of medication recommender systems.***

Keywords: Medicine recommendation system, Medicine Recommender, LangChain Agents, SBERT embeddings, Contextual medicine recommender, Google search agent, LLM Grounding, Neo4j Database, Neo4j Vector Database

1. Introduction

One of the primary and critical tasks that any form of healthcare delivers is the ability to delineate which medications are appropriate for a given ailment. Navigating different classes of medications, each with its own unique indications, contraindications, and adverse effects, can be a challenge for both patients and health care professionals. Traditional approaches to medication suggestion often rely on manual knowledge and keyword - based searching, which, in turn, may neglect the intricate bonds between illnesses and therapies. This can be detrimental in clinical contexts due to the fact that there might be a compromise in treatment - related decisions, which, in turn, may affect the outcome and efficacy of healthcare in general. Proper medication recommendations are one of these aspects that cannot be overstated, as they have a direct impact on patient health and treatment effectiveness. However, most existing medical recommendation systems have a number of limitations. Most systems that rely on keyword - based search or static databases thus overlook the semantic context of user queries. Within this project, a system for intelligent medical recommendation is to be developed in order to overcome the limitations of existing approaches, with the help of LangChain agents and semantic search techniques. Our project, "Semantic Search - Based Medicine Recommender with LangChain Google Search Agent, " aims to provide accurate and contextually appropriate medication recommendations using the help of LangChain agents and semantic search techniques. The plan is that our solution will change the way both the patients and healthcare providers access and use the drug information through the use of sophisticated NLP models like SBERT, efficient storage, and retrieval methods in Neo4j and a comprehensive dataset of 23, 000 medicines. The ultimate goal is to improve patient outcomes and the domain of healthcare informatics by optimizing the accuracy, efficiency, and reliability of medical recommendations.

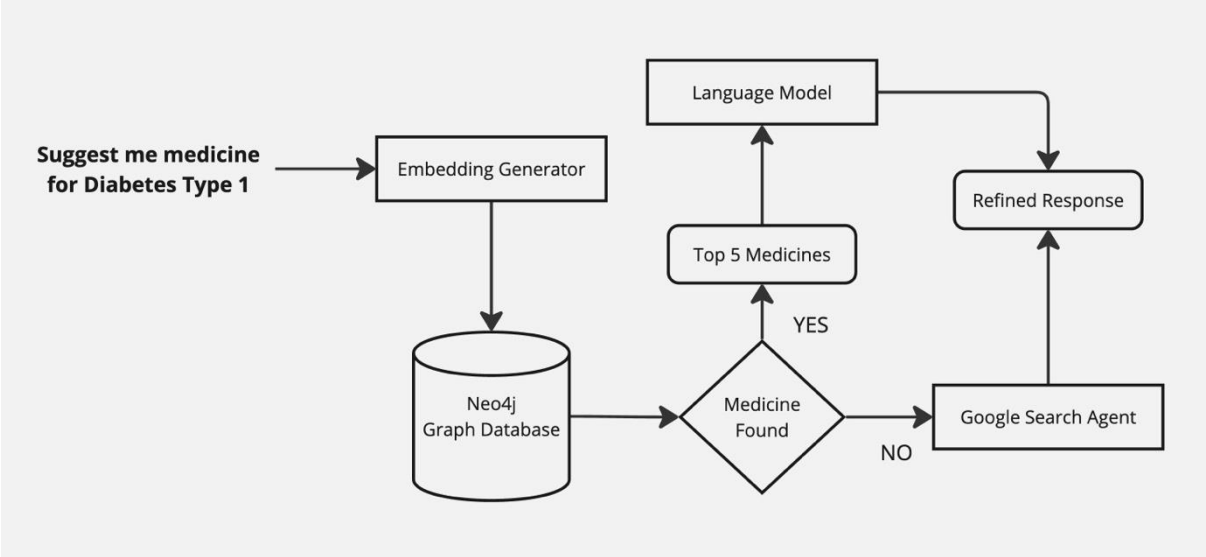
2. Literature Review

For many years, scholars and researchers focused deeply on medicine recommendation systems in an attempt to raise the precision and effectiveness levels of medication prescription. The predominant systems usually deploy characteristics of collaborative filtering techniques or rule - based strategies. While collaborative filtering makes suggestions based on user preferences and past data, rule - based systems rely on pre - set criteria and medical guidelines to make suggestions of medications. The dependence of these techniques on a standardised data format and their inability to adapt precludes them from enjoying the fruits of recently integrated machine learning and deep learning approaches. In order to compensate for these pitfalls, aspects of machine and deep learning approaches have been incorporated into them. For instance, many models like Decision Trees, Random Forests, and Neural Networks have been used to come up with a prediction of the optimal drug that should be recommended depending on patient data. Even with such improvement, most systems still have a hard time understanding the subtle contextual clues in medical searches, which makes their recommendations less dependable.

Comparison

ASPECT	TRADITIONAL SYSTEM	OUR APPROACH
Database Type	Static	Dynamic retrieval with Google Search Agent
Search Method	Keyword - based	Semantic search using SBERT embeddings
Recommendation Technique		Cosine Similarity, Contextual Based
Data Freshness	N/A	Scrapping Year 2024
Understanding Context	Limited	Deep semantic understanding of medical terms
Efficiency	Struggles with large datasets	Efficient storage and retrieval using Neo4j
Flexibility and Robustness	Less adaptable	Enhanced with LangChain agents for dynamic searches

3. System Architecture



The architecture of our medicine recommendation system aims to use modern and sophisticated ways of semantic search, machine learning models, and dynamic data retrieval, delivering correct and context - appropriate medication recommendations. The architecture consists of many interrelated parts that interact to process the user's query, provide embeddings, and eventually recommend the best medicine to use.

Data Scrapping & Cleaning:
The first stage is to generate a comprehensive list of medications, including their uses, descriptions, side effects, and other pertinent information, by scraping data from reputable medical sites. After being scraped, the data is cleaned to get rid of any redundant or inconsistent information. This guarantees that the dataset is correct and prepared for additional processing.

NLP Sentence Creation:
Cleaned data has NLP techniques applied on it in order to produce meaningful language representations of the drug descriptions. This is a very necessary step towards the production of accurate semantic embeddings.

Embeddings generation using SBERT:
Putting the sentences generated for NLP into the SBERT (Sentence - BERT) model results in embeddings in high dimensions. The similarity computation is facilitated because the embeddings capture the semantic information of the medicinal descriptions.

Neo4j Graph Database:
The developed embeddings are stored in a Neo4j graph database. This is a database specifically designed to handle enormously distributed data and retrieves and stores the medical embeddings efficiently and quickly. This database acts as the central repository for our recommendation system.

User Query Processing:
Using the same SBERT model, an embedding is created for a query entered by a user (e. g., "Suggest me medicine for Diabetes Type 1"). To obtain the top 5 most similar medications, the embedding of the query is matched with embeddings stored previously in the Neo4j database.

Medicine Retrieval:
If it finds relevant medicines in the database, the user is provided with a list of the top 5 recommended medicines. If the system does not find any relevant medicines in the database, it performs a real - time Google search using a Google Search Agent powered by LangChain. The information relating to relevant medicines is discovered by processing the search results.

This architecture allows our system to be able to provide recommendations on medicines that are accurate, up - to - date, and contextually relevant, thus improving healthcare experiences.

4. Methodology

We meticulously scraped data for 23, 000 medications from reputed medical resources as a part of our data collection phase to initiate building a strong recommendation system for medications. This large dataset comprised detailed

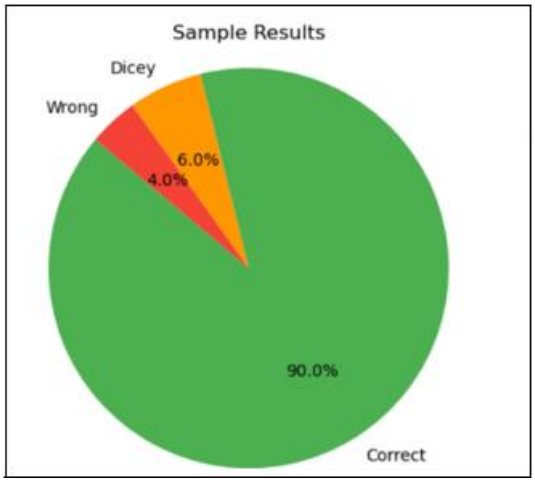
information on the uses, descriptions, adverse effects, and other aspects of every medication. Next, the gathered data needed to be carefully cleaned and organized to ensure its accuracy and usefulness. We used Python programmes to clear the data of inconsistencies, definitively repeated information, and explicitly meaninglessness data, thus transforming the unstructured data into a format suitable for further processing. We then moved on to the NLP sentence construction phase. Here, we received meaningful sentences from the cleaned drug descriptions using natural language processing text. In order to effectively analyze any drug, it was very important to portray the crux in this manner. Using the SBERT (Sentence - BERT) model, we were, text we got, to produce high - dimensional embeddings with these sentences. Every description of medications was converted into a 1024 - dimensional embedding in such a way that the context and semantic meaning of the data is represented. Since it was critical to store these embeddings as efficiently as possible, we chose Neo4j, a graph database designed to handle massively connected data. Due to Neo4j's storage and retrieval ability, we were quite able to manage and query these managed embeddings correctly. To ensure the most relevant recommendations were found during retrieval, we built cosine similarity computations to compare user query embeddings with the stored medicine embeddings. An easy - to - use chatbot user interface makes it easy for users to interact with our system. The chatbot uses the same SBERT model to process user queries such as \"Suggest me medicine for Diabetes Type 1\" and creates an embedding based on the results. This is followed by comparing such an embedding with the embeddings that Neo4j has stored. The user is then presented with the top recommendations for which good medicines have been found. If not, the system does a live search through a Google Search Agent integrated, powered by LangChain, fully integrated into the system. The answer is then refined to ensure accuracy and relevance in context. Our work in a nutshell integrates data scraping and cleaning, natural language processing, embedding conception, effective storage, and user - friendly interfaces to afford the end user a state - of - the - art medicine recommendation system both accurate and responsive to the needs of the user.

5. Implementation

We deployed the following technologies in constructing the medicine recommendation system: Neo4j, to enable efficient storage and querying of data; Python, for scripting and data manipulation; SBERT (Sentence - BERT), for developing 1024 - dimensional semantic embeddings; and LangChain, to enable real - time search functionalities through a Google Search Agent. The initial implementation phase is data collection, in which we collected detailed information on about 23, 000 medications from reliable sources. To maintain high quality standards in the data, Python scripts were used to clean and structure the data. We then created meaningful natural language statements for every medication. An SBERT model converts these statements into suitable embeddings. This data is held within the Neo4j graph database, which supports fast and accurate similarity searches. The chatbot user interface processes the inquiries a user makes. It then converts the user's inquiry into embeddings and looks up the previously stored data using a similarity search. In case no relevant medications are found, LangChain performs a real - time search to provide the latest information. We had issues dealing with large volumes of data and maintaining the quality of the embeddings generated in every stage of the process. This was overcome by using efficient data structures, optimizing data cleaning steps, and enhancing the efficiency of the SBERT model. Additionally, combining Neo4j with LangChain to ensure that data and search are in real - time coordination required a detailed approach, which we developed by iteratively testing and adjusting.

6. Conclusion

In conclusion, our study has successfully created a complex pharmaceutical recommendation system that makes use of state - of - the - art technology like LangChain for dynamic information retrieval, Neo4j for effective data storage, and SBERT for semantic embedding. The user experience stands to be substantially improved due to the system making correct interpretations of user inquiries, ultimately resulting in pertinent and current suggestions of medication. This has been a major achievement through robust applications, namely high - dimensional embeddings in capturing semantic context, integration mechanisms of NLP techniques, and development towards building seamless user - interface systems for intuitive interaction. Further work on this project will improve and take it in new ways: adding some more medical diseases into our dataset, increasing user feedback to improve recommendations, and using more powerful AI models to increase the system's real - time search capabilities. With the same benefit and impact, multilingual support can be studied in order to help the system be used by a larger audience.



7. Results & Evaluation

Total Sample used were 50 out of which 45 showed correct results, 3 gave partially correct outputs (medicine to be consumed only after medical consultation) and 2 failed to return correct response

Question	Recommendation	Ground Truth
Suggest medicine for kidney stones	ALKASTON ORAL SOLUTION	TRUE
Suggest medicines for diabetes mellitus type 2	AMARYL M 2mg Tablet 20's	TRUE
Which medicine should be used to cure fungal infection on skin	AMROLFUNG Cream 30gm	TRUE
Suggest medicine to apply on wound	DRESIN OINTMENT	TRUE
Which medicine to use for controlling high blood pressure	CARDIOLOL H TABLET:	TRUE

References

[1] Medicine Recommendation using Machine Learning: https: //ijsrset. com/paper/8152. pdf
[2] Medicine Recommendation System: https: //www.irjet. net/archives/V7/i3/IRJET - V7I3310. pdf
[3] Drug Recommendation System in Medical Emergencies using Machine Learning: https: //ieeexplore. ieee. org/document/10099607
[4] Medicine Recommendation System Based On Patient Reviews: https: //www.ijstr. org/final - print/feb2020/Medicine - Recommendation - System - Based - On - Patient - Reviews. pdf

Appendices

Project Repository: https: //github. com/Kallind/Semantic - Search - Based - Medicine - Recommender - with - LangChain - Google - Search - Agent