Generative Search with OpenAI and Chroma – Ayurbook Project

Author: Mamatha K

Domain: AI | Generative Search | RAG | NLP | Ayurveda Knowledge Systems

1. Introduction

Traditional information retrieval systems rely heavily on keyword-based search, which struggles to capture semantic meaning, especially in domain-specific corpora such as Ayurveda. This project addresses that limitation by implementing a Generative Search system using OpenAI's language models and ChromaDB. The system enables users to ask natural language questions about Ayurvedic concepts and receive context-aware, human-like answers generated from relevant text sources. The project demonstrates how Retrieval-Augmented Generation (RAG) can be applied to classical text collections to preserve traditional knowledge while making it accessible through modern AI tools.

2. Objectives

- To design a Generative Search pipeline that can retrieve and generate answers from unstructured Ayurvedic text data.
- To combine OpenAI LLMs with ChromaDB to enable semantic retrieval and generative summarization.
- To process and embed Ayurvedic books into a searchable vector database for context retrieval.
- To evaluate the system for relevance, contextual accuracy, and coherence in generated answers.
- To demonstrate a scalable, reusable RAG framework applicable to other specialized knowledge domains.

3. Data Sources

Dataset: Ayurbook Corpus, derived from classical Ayurvedic literature and modern interpretations.

Data Preprocessing included text cleaning, chunking, normalization, and metadata tagging.

4. System Design

The overall architecture is based on the Retrieval-Augmented Generation (RAG) paradigm, which combines information retrieval with generative modeling to produce contextually relevant responses.

Pipeline: User Query → Embedding → Vector Search (ChromaDB) → Context Retrieval → OpenAI LLM → Generated Answer

5. Implementation

The project uses Python 3.10+, LangChain, ChromaDB, and OpenAI APIs for embedding and generation. Text data was processed into smaller chunks and stored in a local vector store. User queries are embedded, matched against this database, and passed along with context to the OpenAI model for final answer generation.

6. Challenges Faced

Challenge	Solution
Text Complexity	Normalized Sanskrit transliterations and cleaned datasets.
Large File Sizes	Implemented chunking with overlap to preserve context.
Embedding Latency	Batched and cached embeddings locally.
Token Limit Exceeded	Implemented token-count checks using tiktoken.
Domain-Specific Knowledge	Provided contextual grounding via retrieved text chunks.

7. Lessons Learned

- RAG improves factual accuracy compared to direct generation.
- Chunking with overlap preserves context and improves coherence.
- Metadata enriches retrieval results.
- Manual validation is essential for specialized domains.
- Combining retrieval and generation yields robust performance.

8. Results and Observations

The system generated accurate, context-grounded responses to queries related to Ayurvedic topics. Retrieval precision improved with optimized chunk sizes between 800–1200 characters.

9. Future Enhancements

- Integrate local embedding models such as Ollama or HuggingFace.
- Expand corpus with additional Ayurvedic texts (Charaka Samhita, Sushruta Samhita).
- Introduce semantic evaluation metrics (ROUGE, BLEU, cosine similarity).
- Deploy as a web dashboard using Streamlit or FastAPI.
- Explore multimodal retrieval (text + image data).

10. Conclusion

The Ayurbook Generative Search project demonstrates a successful application of Retrieval-Augmented Generation in traditional medicine. By integrating semantic retrieval (ChromaDB) with generative AI (OpenAI), the system bridges ancient knowledge with modern technology, enhancing accessibility and enabling AI-assisted knowledge discovery.

Screenshots

(Gen layer-RAG):

```
# Generate the response
query = "What are the top three reasons for heart diseases"
response = generate_response(query, top_3_RAG)
```

```
# Print the response

print("\n".join(response))
```

```
Heart diseases in Ayurveda are often attributed to imbalances in the doshas and other factors. The top three reasons for heart diseases according

1. **Imbalanced Doshas**: When the doshas Vata, Pitta, and Kapha are imbalanced, it can lead to various heart issues. For example, an excess of Pi

2. **Poor Digestion**: Weak digestion can result in the accumulation of toxins (ama) in the body. These toxins can affect the heart and its functi

3. **Emotional Stress**: In Ayurveda, emotional well-being is closely connected to physical health. Chronic stress, anxiety, and emotional disturb

**Citations:**
- *Ayurveda Prakasha - A Text of Indian Alchemy 3*, Page 56
- *Ancient Science Ayurveda*, Page 23
```

[224] ✓ 3s

```
query = "How is pitta imbalance treated naturally?"
response = generate_response(query, top_3_RAG)
print("\n".join(response))
```

```
Pitta imbalance can be treated naturally in Ayurveda through various methods. Some common ways to treat pitta imbalance naturally include:

1. **Dietary Changes**: Avoiding hot, spicy, and oily foods that can aggravate pitta dosha. Opt for cooling foods like cucumbers, mint, and coria

2. **Herbal Remedies**: Using herbs like coriander, fennel, and aloe vera to help balance pitta dosha.

3. **Lifestyle Modifications**: Practicing calming activities like meditation and yoga, avoiding excessive heat or stress, and maintaining a regu

4. **Ayurvedic Therapies**: Panchakarma treatments like Abhyanga (oil massage) and Shirodhara (oil pouring) can help balance pitta dosha.

For more detailed information on treating pitta imbalance naturally, you can refer to the Ayurvedic documents provided in the dataset.

Now, let's look at the citations for more specific details:

1. In 'Ayurveda Prakasha - A Text of Indian Alchemy', please refer to the section on pitta dosha treatment on page 35 for in-depth information on
```

[225] ✓ 2s

```
query = "Which herbs are used for improving digestion in Ayurveda?"
response = generate_response(query, top_3_RAG)
print("\n".join(response))
```

```
The herbs commonly used in Ayurveda for improving digestion include:

1. **Ginger (Zingiber officinale):** Helps in stimulating the digestive fire and improving digestion.
2. **Turmeric (Curcuma longa):** Aids in reducing inflammation in the digestive system and supporting digestion.
3. **Fennel (Foeniculum vulgare):** Helps in relieving gas, bloating, and indigestion.
4. **Coriander (Coriandrum sativum):** Supports digestion and helps in reducing digestive discomfort.
5. **Mint (Mentha):** Known for its carminative properties, which aid in digestion.

You can find more detailed information on these herbs and their specific benefits for digestion in the Ayurvedic document "Ayurveda Prakasha - A

---

Citations:
- Document: Ayurveda Prakasha - A Text of Indian Alchemy
- Page numbers: Refer to the relevant sections in the document as per the metadata information.
```

(Search Layer)

```
cache_results
```

```
{'ids': [['What are the top three reasons for heart diseases']],
 'embeddings': None,
 'documents': [['What are the top three reasons for heart diseases']],
 'uris': None,
 'included': ['metadatas', 'documents', 'distances'],
 'data': None,
 'metadatas': [[{'documents1': '[Downloaded free from http://www.ancientscienceoflife.org on Friday, December 19, 2014, IP: 117.236.172.188] ||
Click here to download free Android application for this journal Book Review defects and but not the tastes. The eight defects of cooked the
other hand, even nectarine drugs if used in improper rice are aṣṭānnna (rice preparation using rice whose quantity could become a poison.
qualities go against the season), paicchilyānna (over cooked, sticky rice), kwathitānna (only litle quantity of rice taken The method of sūpa
(dehusked legumes cooked) preparation for cooking and boiled), śuṣkānna (dried cooked rice), has been explained. Horse gram (Dolichos bilorus),
black dagdhānna (burnt rice), virūpānna (cooking rice without gram (Vigna mungo), cow peas (Vigna unguiculata), and proper size and shape), and
anaṛtujānna (rice against the chickpea (Cicer arietinum) etc., which are cleaned of season or stale cooked rice). Further the author explains
physical impurities using grain sieve (śūrpa), or de-husked the method of cooking of rice. Rice which is unpounded, green gram (Vigna radiata)
should be taken in one part, dry and older should be taken and should be washed in and equal quantity of water has to be mixed and boiled. hot
water. A vessel should be taken and should be added When legume is well prepared, suitable quantity of rock with three parts of water. The
vessel should be placed on salt has to be added. Turmeric is used for colouring and ire. When water slightly boils up, the washed rice should
asafoetida is used for its aroma. This has to be churned be put in the vessel. When rice starts boiling, it should be thoroughly using a spoon
adding water. Later, this mixture stirred using a spoon repeatedly. At the same time, milk, is seasoned with camphor, and various aromatic
lowers. buter milk or water should be sprinkled. Then the vessel This dal preparation appetizing, subsides pita and lighter should be taken out
from ire for some time and again kept for digestion. for boiling till it is cooked and smooth. This rice is good for longevity and health.
Different kinds of butter milk preparations have been dealt with. A few interesting combinations include buter In the same chapter, the author
explains the method of milk being mixed with certain drugs like ajamoda (Carum preparation of nonvegetarian biriyani. For this, three parts
```

```
cache_results
```

```
{'ids': [['How is pitta imbalance treated naturally?']],
 'embeddings': None,
 'documents': [['How is pitta imbalance treated naturally?']],
 'uris': None,
 'included': ['metadatas', 'documents', 'distances'],
 'data': None,
 'metadatas': [[{'included1': 'e',
    'metadatas0': "{'Page_No.': 'Page 3', 'Book_Specialization_Name': 'The_Paka_Darpanam_The_text_on_Indian_coo'}",
    'ids1': '3',
    'included0': 'm',
    'metadatas4': "{'Book_Specialization_Name': 'The_Paka_Darpanam_The_text_on_Indian_coo', 'Page_No.': 'Page 2'}",
    'distances0': '0.40709444880485535',
    'metadatas1': "{'Book_Specialization_Name': 'OverviewonAyurvedPrakash', 'Page_No.': 'Page 4'}",
    'included3': 'a',
    'metadatas2': "{'Book_Specialization_Name': 'OverviewonAyurvedPrakash', 'Page_No.': 'Page 8'}",
    'distances3': '0.4574417173862457',
    'distances2': '0.45107778906822205',
    'ids0': '11',
    'documents2': 'Ayurveda Prakasha - A Text of Indian Alchemy 35 and practical utility from other texts e.g. quoted Author had adopted easy,
practically Suddha Suta Lakshana as reduction to 1/8th after applicable methods only e.g. he suggests using samskara, acquisition of semisolid
nature of Loha churna in place of Patra for processing Parada after Swedana Samskara, and opines to collect it with the help of chumbaka.
Dvandvamelaka Aushadhi, physical effects of As the author accepts scarcity of Kanta therefore sequential increment in quantity of Grasa in he
```

11. References

1. OpenAI API Documentation – https://platform.openai.com/docs
2. LangChain Framework – https://python.langchain.com
3. ChromaDB Documentation – https://docs.trychroma.com
4. Python tiktoken Library – https://github.com/openai/tiktoken
5. Classical Ayurvedic Literature (Public Domain Sources)