A SIMPLE GUIDE TO

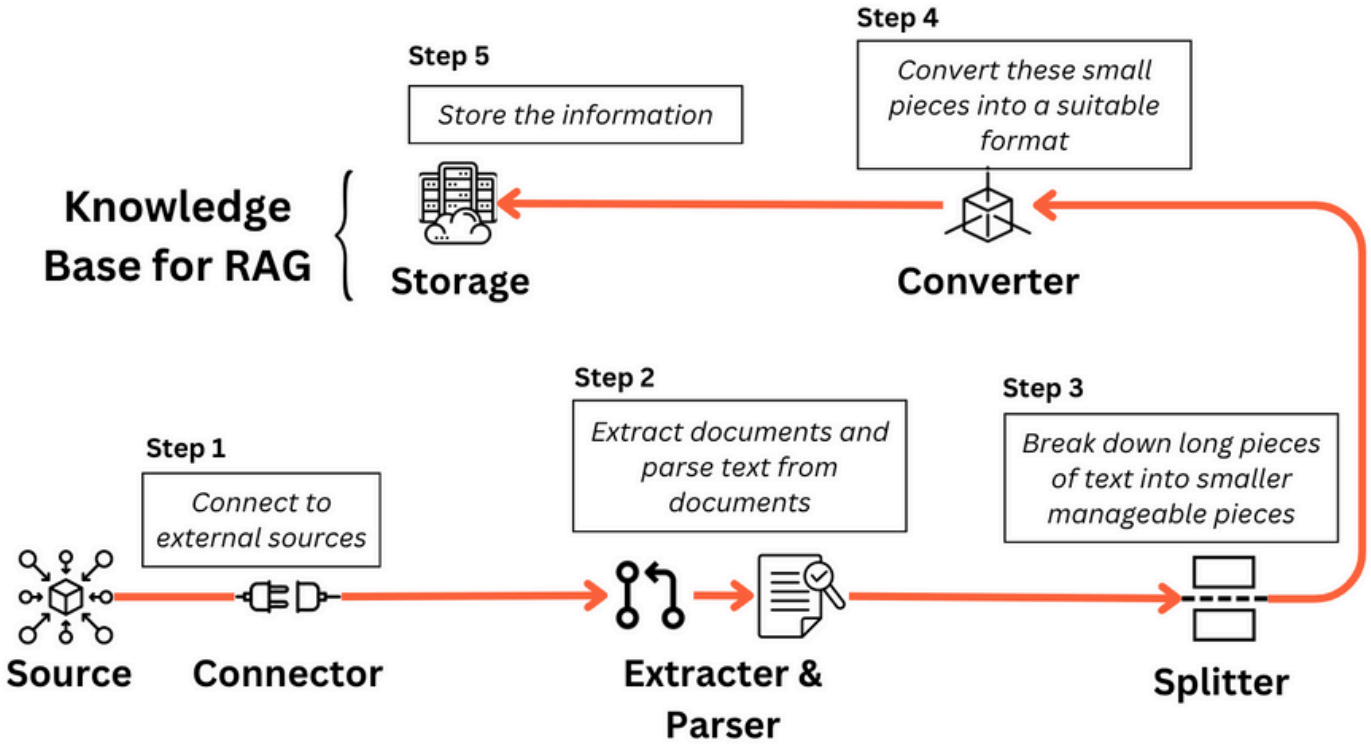
Retrieval Augmented Generation

Github Repository

- Indexing Pipeline
- Generation Pipeline
- RAGAS Evaluation
- Benchmarking

MANNING

Indexing Pipeline : Creating a knowledge base for RAG based applications

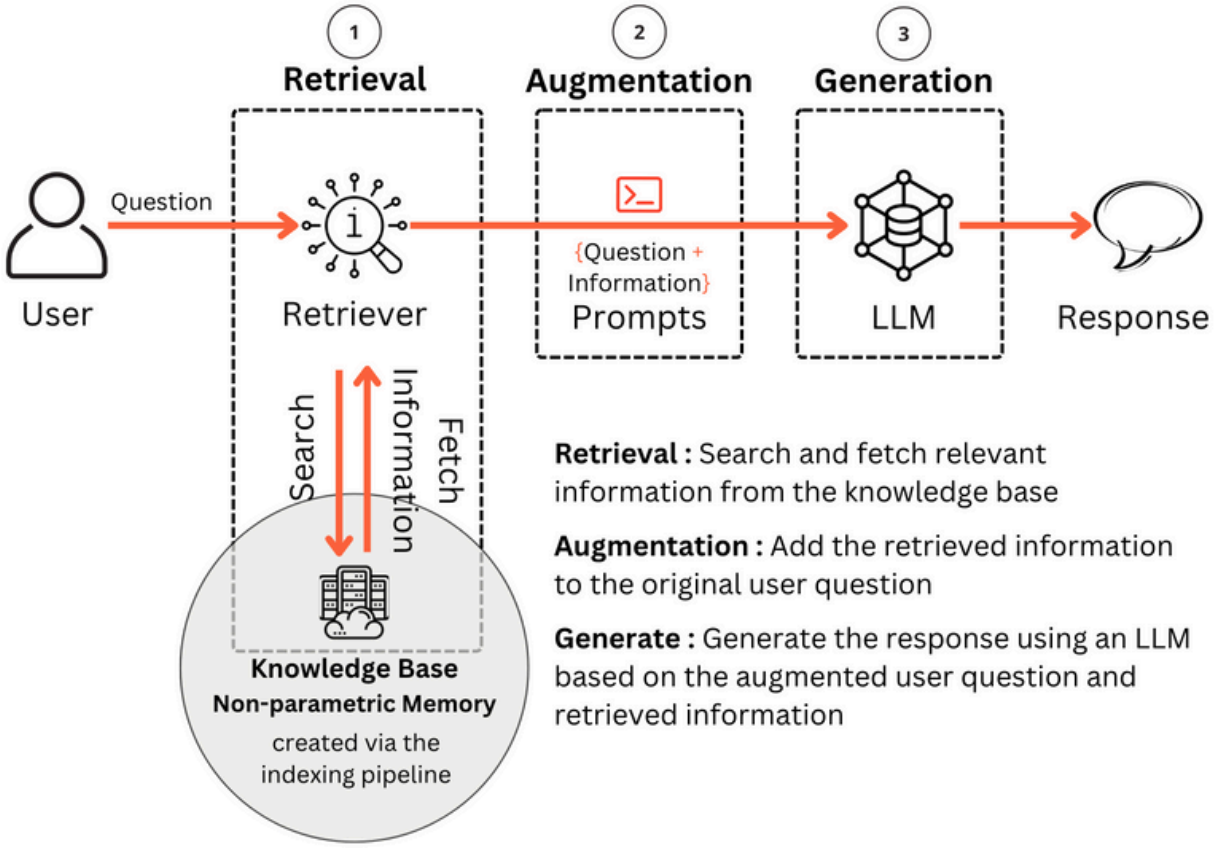


Indexing Pipeline covering the steps to create the Knowledge Base for RAG

Notebook Summary

A knowledge base is created for the **2023 Cricket World Cup** based on the **Wikipedia Article** on the topic. We use **AsyncHtmlLoader** and **Html2TextTransformer** to load the article, chunk the text using **RecursiveCharacterTextSplitter**, use **text-embedding-3-large** from **OpenAI** to convert chunks into vectors and use **FAISS** as the vector index to store the embeddings.

Generation Pipeline: Real time interaction for contextual responses

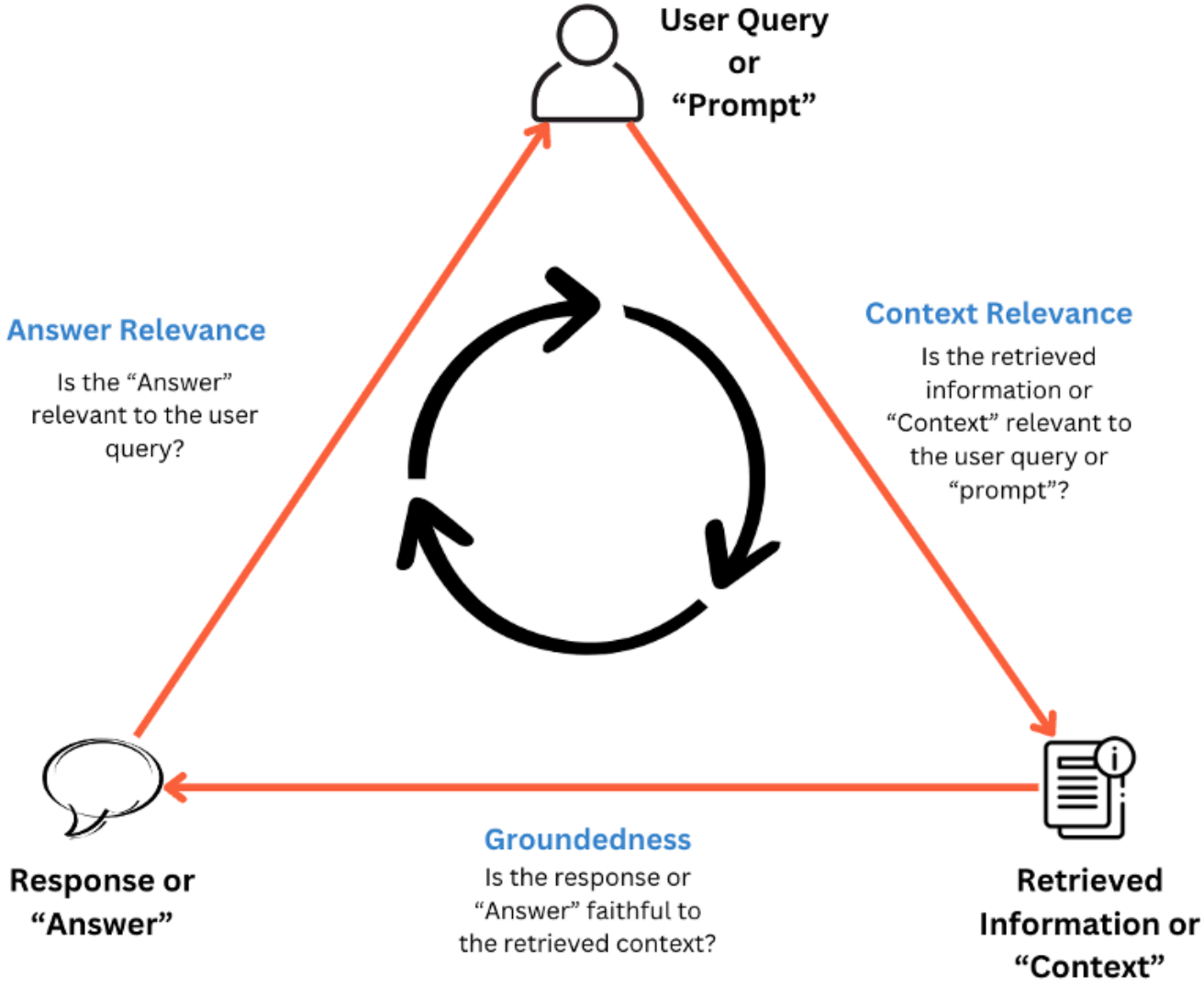


Generation Pipeline Overview with its three components

Notebook Summary

We use the knowledge base on the **wikipedia article** on **2023 Cricket World Cup**. We load the **FAISS** index and use the **similarity search** function to retrieve chunks. We then Augment the user query with the retrieved chunk and use **GPT 4o** model from OpenAI to generate the response.
This notebook also includes functions that can be used to generate answers for different queries that a user may want to ask.

RAG Evaluation : Checking accuracy, relevance and faithfulness



RAG Triad proposed by TruLens

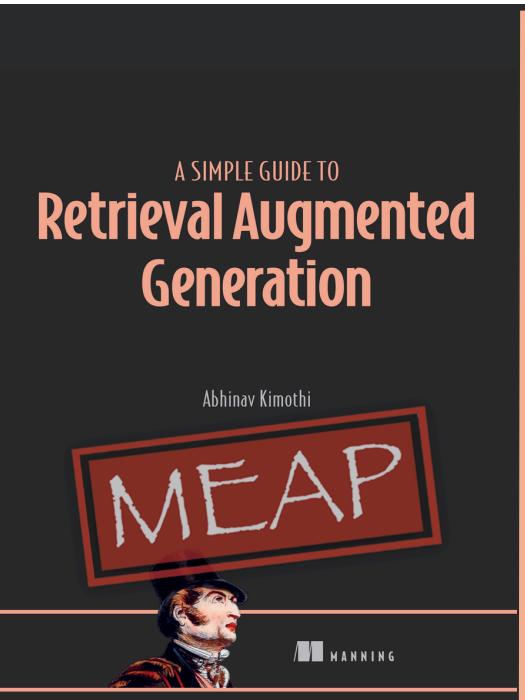
Notebook Summary

We evaluate the RAG pipeline created in chapters 3 and 4 using the **RAGAS** framework.

Additionally, this chapter includes a notebook that uses **LangChain Benchmarks** to benchmark our RAG pipeline on **LangChain QnA docs**.

Hello!
I'm Abhinav...
A data science and AI professional with over 15 years in the industry. Passionate about AI advancements, I constantly explore emerging technologies to push the boundaries and create positive impacts in the world.

A Simple Guide to Retrieval Augmented Generation is now available for Early Access



Learning Goals

- Develop a solid understanding of RAG fundamentals, the components of a RAG enabled system and its practical applications.
- Gain knowledge about developing a RAG enabled system with details about the indexing pipeline and the generation pipeline.
- Gain deep insights into the evaluation of RAG enabled systems and modularised evaluation strategies
- Familiarise yourself with advanced RAG strategies and the evolving landscape of GraphRAG, AgenticRAG & more

Why Join MEAP?

- Immediate access to the book's current draft and all future updates
- A chance to provide feedback and shape the final content
- Exclusive discounts and early-bird offers

Subscribe Now to get 50% off

Avail Early Access Discounts to Chapter 1-3

Raw & Unedited

Ch 1 : LLMs & the need for RAG
Ch 2 : RAG enabled systems & their design
Ch 3 : Indexing Pipeline - Creating a knowledge base for RAG based applications

Complete book coming soon

View Code

Join MEAP