

Lecture 37 # Building RAG System

1. Chunking of Data

2. Embeddings of Data

Client Psychology :-

- Add some personal touch in service / product
- customized touch

many almost all model \Rightarrow trained on fixed dataset
(generalized data)

Solution

RAG

\rightarrow streamline

(will allow external data + power of LLM)

- accuracy & precision \rightarrow can get result in format you want
like json, xml

① Multi-Modal (Text, Speech, Image)

② Localizator

③ Multi-Lingual

④ contextual understanding

⑤ Personalized Touch

\Rightarrow where customer involves

⑥ currently 50-60% business makes RAG

\Rightarrow Develop the interest

\Rightarrow How it change business... choose the pin point

Tech knowledge + Tech Personality

way to LLM

\rightarrow either through user-interface (dependent on web)

\rightarrow Through RAG

Develop Critical Theory ^{Exposure} _{loading}

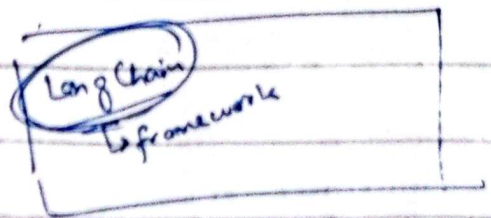
The hottest new programming language is English

previous lecture revision

... knowledge source (example bank policy, medical records)
can question answer with your data.

knowledge source + LLM attached \Rightarrow response in Natural language

- custom knowledge base
- LLM with your own data



- data can be of multiple format
(like pdf, csv, txt, etc. -)
- loading your data through LangChain loaders

Step 1: loading the data (loaders - loaders in LangChain)

Step 2: How to use?
generate embeddings
(certain limitation with them)

To make our data pre-processing simpler

divide the text \rightarrow chunks with
(for example each chunk
2 page)

- for getting desired output so
only relevant part

- so model can process
- LLM can maintain large
context

Splitter

Tool used to divide large text documents into smaller, more manageable chunks.



میں نے اسے بڑے بڑے chunk کے (size) کے حصے میں

Technique of overlapping

یعنی اگر آپ کے پاس کئی صفحات ہیں تو ان کے overlap رکھیں گے۔

in some scenario we not need overlap

Types of Splitters

- ① Recursive Splitter
- ② HTML Splitter
- ③ Markdown splitter
- ④ code splitter
- ⑤ Token Splitter
- ⑥ Character splitter
- ⑦ Semantic Splitter

یعنی بات بات آپ کے ٹکڑوں میں رکھیں گے۔

He is good.

He is nice.

(meaningfully close together)

Practical demonstration

Documents → chunks
(loading) (splitting)

longchain package

• importing the package for splitter

• from longchain.textsplitter import

in CharacterTextSplitter

• by default separator "paragraph"

Recursive Text Splitter

↳ can pass list of separator

example ["\n", "\n\n", " ", ""]

- chunk-size

- chunk-overlap (to solving context)

- separators

Example

↳ in holy Quran separate embeddings for each ayah

ایک بڑا جگہ ڈیفائن کرنا ہے



separator بتا دے میں مگر اگر ان میں سے کوئی آتا ہے تو

میں بڑا کر دیتا ہے درجہ جگہ سے مطابقت رکھتا

- 2

✓ maximum chunk size depends on embeddings model.

• splitter example #1

• splitter example #2

- ⇒ Just giving the direction
- ⇒ worth of product ideas
- ⇒ Selling skills
- ⇒ personality communication building
- focus on prenegotiation [knowledge + Experience]
- selling, communication is much important.

✓ Concept of Web Scrappers