

Project- FinBot

Project Background

On a daily basis there are many news articles that come about on the internet based on the stock market. A user may find it time consuming to sit and read all the articles, hence gets overwhelmed. Enter FinBot.

It helps to ingest all the articles you like to gain knowledge about by questioning the bot. The bot then goes through the articles, gathers the information and provides the answer.

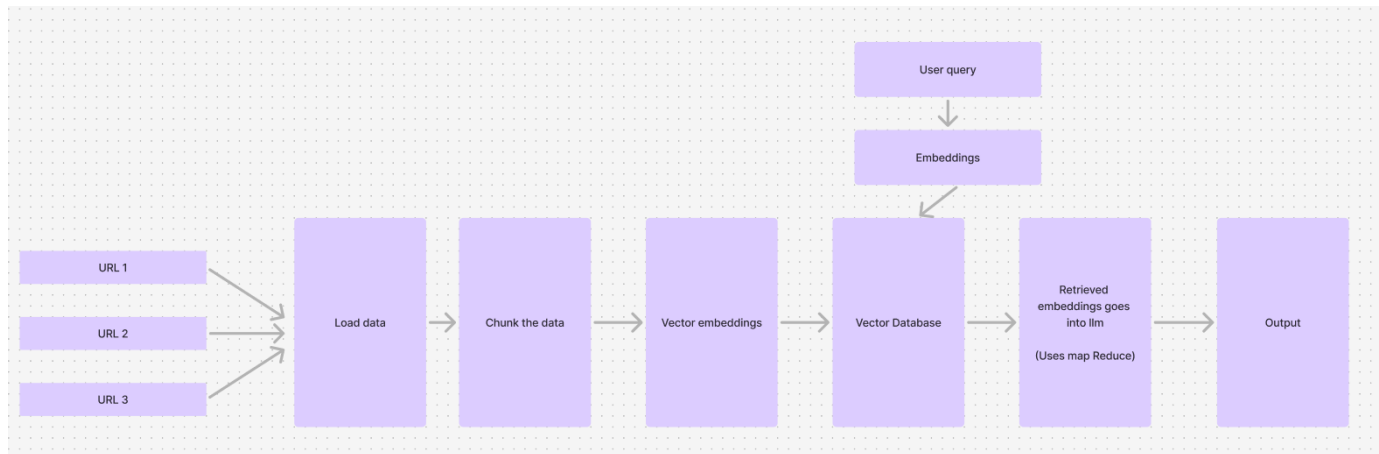
Problem Statement

Given 5 urls, various websites- can be Hindu, ET, moneycontrol. Feed the data from the website to a Database and extract the information related to the user query. Works on the basic principal of RAG system.

Approach:

1. Load the data from all the URLs
2. Chunk the data using text splitters
3. Convert the chunks into vector embeddings
4. Store the embeddings to a vector DB
5. When queried, the LLM provides an answer

System Design:



Implementation:

This project is built on Langchain libraries.

Step 1: Loading the data from the urls:

- ➔ Using the UnstructuredURLLoader from langchain
It scraps the data from the urls and loads it into a langchain object

Step 2: Splitting the data into chunks:

- ➔ Using RecursiveCharacterTextSplitter from langchain
it recursively splits further while preserving meaning of the sentences, adding the chunk overlap also helps to maintain the previous context of the chunk.

Step 3: Create vector embeddings of these texts and store it into FAISS:

- ➔ Using OpenAIEmbeddings to create the vector embeddings
- ➔ Using Faiss to store the vector indexed embeddings.

Facebook AI Similarity Search, Handles millions/billions of embeddings quickly and best for fast in-memory search for RAG.

Step 4: Retrieve similar embeddings for a given question and call LLM to retrieve it

- ➔ Using RetrievalQAWithSourcesChain from langchain
 - Retrieves documents based on the user's query .
 - Feeds retrieved documents into the LLM to generate an answer.
 - Returns both the answer and the sources for transparency.

Screenshots of output:

```
1 query = "Can you tell me the reason for stock market crash?"
2 langchain.debug=True
3
4 q_a_chain({"question":query}, return_only_outputs=True)

✓ 3.5s
```

```
[chain/start] [chain:RetrievalQAWithSourcesChain] Entering Chain run with input:
{
  "question": "Can you tell me the reason for stock market crash?"
}
[chain/start] [chain:RetrievalQAWithSourcesChain > chain:MapReduceDocumentsChain] Entering Chain run with input:
[inputs]
[chain/start] [chain:RetrievalQAWithSourcesChain > chain:MapReduceDocumentsChain > chain:LLMChain] Entering Chain run with input:
{
  "input_list": [
    {
      "context": "Written by Hitesh Vyas Mumbai | Updated: February 28, 2025 16:13 IST\n5 min read\nDecline in Indian equities mirrored ]
      "question": "Can you tell me the reason for stock market crash?"
    },
    {
      "context": "Foreign portfolio investors (FPIs) have been at the forefront of the selling spree, pulling out a whopping Rs 2.13 lakh cr
      "question": "Can you tell me the reason for stock market crash?"
    },
    {
      "context": "Story continues below this ad\nAsian markets were the most affected by tariff news on Friday which led to a sharp correct
      "question": "Can you tell me the reason for stock market crash?"
    },
    {
      "context": ""Stock markets dislike uncertainty, and uncertainty has been on the rise ever since Trump was elected the US president. The
      "question": "Can you tell me the reason for stock market crash?"
    }
  ]
}
[LLM/start] [chain:RetrievalQAWithSourcesChain > chain:MapReduceDocumentsChain > chain:LLMChain > llm:OpenAI] Entering LLM run with input:
{
  "input_list": [
    {
      "context": "Written by Hitesh Vyas Mumbai | Updated: February 28, 2025 16:13 IST\n5 min read\nDecline in Indian equities mirrored ]
      "question": "Can you tell me the reason for stock market crash?"
    },
    {
      "context": "Foreign portfolio investors (FPIs) have been at the forefront of the selling spree, pulling out a whopping Rs 2.13 lakh cr
      "question": "Can you tell me the reason for stock market crash?"
    },
    {
      "context": "Story continues below this ad\nAsian markets were the most affected by tariff news on Friday which led to a sharp correct
      "question": "Can you tell me the reason for stock market crash?"
    },
    {
      "context": ""Stock markets dislike uncertainty, and uncertainty has been on the rise ever since Trump was elected the US president. The
      "question": "Can you tell me the reason for stock market crash?"
    }
  ]
}
{
  "answer": "The stock market has crashed due to the uncertainty and impact of tariffs by the US president, as well as large foreign portfolio
  'sources': 'https://indianexpress.com/article/business/market/stock-market-crash-today-sensex-nifty-live-updates-9860683/'
}
```

✓ 4.0s

```
[chain/start] [chain:RetrievalQAWithSourcesChain] Entering Chain run with input:
{
  "question": "What are the stocks that gained over 10% in BSE?"
}
[chain/start] [chain:RetrievalQAWithSourcesChain > chain:MapReduceDocumentsChain] Entering Chain run with input:
]
[chain/start] [chain:RetrievalQAWithSourcesChain > chain:MapReduceDocumentsChain > chain:LLMChain] Entering Chain run with input:
{
  "input_list": [
    {
      "context": "Follow us\n\nShare\n\nFont Size\n\nAbcSmall\n\nAbcMedium\n\nAbcLarge\n\nSave\n\nPrint\n\nComment\n\nSynopsis\n\nIn the Nifi",
      "question": "What are the stocks that gained over 10% in BSE?"
    },
    {
      "context": "Benchmarks\n\nNifty22,124.70-420.35\n\nFEATURED FUNDS\n\n★★★★★\n\nHSBC Large Cap Fund Direct-Growth\n\n5Y Return\n\n15.1",
      "question": "What are the stocks that gained over 10% in BSE?"
    },
    {
      "context": "Top Trending Stocks: SBI Share Price, Axis Bank Share Price, HDFC Bank Share Price, Infosys Share Price, Wipro Share Price",
      "question": "What are the stocks that gained over 10% in BSE?"
    },
    {
      "context": "Top Trending Stocks: SBI Share Price, Axis Bank Share Price, HDFC Bank Share Price, Infosys Share Price, Wipro Share Price",
      "question": "What are the stocks that gained over 10% in BSE?"
    }
  ]
}
[llm/start] [chain:RetrievalQAWithSourcesChain > chain:MapReduceDocumentsChain > chain:LLMChain > llm:OpenAI] Entering LLM run with input:
{'answer': ' The high-performing stocks that gained over 10% during the session on BSE were Medico Remedies, Home First Finance Co, and Transpe',
'sources': 'https://economictimes.indiatimes.com/markets/stocks/stock-watch/sensex-falls-but-these-stocks-gained-over-10-on-bse/articleshow/11
```

3s

```
[chain/start] [chain:RetrievalQAWithSourcesChain] Entering Chain run with input:
{
  "question": "Can you summarize on the stocks that are hitting 52 week high"
}
[chain/start] [chain:RetrievalQAWithSourcesChain > chain:MapReduceDocumentsChain] Entering Chain run with input:
[inputs]
[chain/start] [chain:RetrievalQAWithSourcesChain > chain:MapReduceDocumentsChain > chain:LLMChain] Entering Chain run with input:
{
  "input_list": [
    {
      "context": "Meanwhile, stocks such as Photon Capital, Hira Automobiles, Aris Intl, Dhanlaxmi Cotex and Triumph International Finance ]
      "question": "Can you summarize on the stocks that are hitting 52 week high"
    },
    {
      "context": "Benchmarks\n\nNifty22,124.70-420.35\n\nNFEATED FUNDS\n\n★★★★★\n\nHSBC Large Cap Fund Direct-Growth\n\n5Y Return\n\n15.1
      "question": "Can you summarize on the stocks that are hitting 52 week high"
    },
    {
      "context": "ETMarkets.comLast Updated: Feb 28, 2025, 05:17:00 PM IST\n\nRate Story\n\nFollow us\n\nShare\n\nFont Size\n\nAbcSmall\n\nAbc
      "question": "Can you summarize on the stocks that are hitting 52 week high"
    },
    {
      "context": "Follow us\n\nShare\n\nFont Size\n\nAbcSmall\n\nAbcMedium\n\nAbcLarge\n\nSave\n\nPrint\n\nComment\n\nSynopsis\n\nIn the Nift
      "question": "Can you summarize on the stocks that are hitting 52 week high"
    }
  ]
}
[LLM/start] [chain:RetrievalQAWithSourcesChain > chain:MapReduceDocumentsChain > chain:LLMChain > llm:OpenAI] Entering LLM run with input:
{
  "answer": " Stocks hitting their fresh 52-week high in today's trade include Photon Capital, Hira Automobiles, Aris Intl, Dhanlaxmi Cotex, Tr
  'sources': "https://economictimes.indiatimes.com/markets/stocks/stock-watch/sensex-falls-but-these-stocks-gained-over-10-on-bse/articleshow/1
```

My Learnings:

- ➔ Learnt about the libraries- UnstructuredURLLoader
- ➔ Learnt about Faiss and how to save the db files and extract and run it
- ➔ Got more experience in RAG model but with urls this time instead pdfs.