

# Symantic Spotter with LlamaIndex

## Project Objective

Build a project in the insurance domain, similar to the project you saw in the 'Retrieval Augmented Generation' session.

The goal of the project will be to build a robust generative search system capable of effectively and accurately answering questions from various policy documents. We are given set of life insurance documents. Our purpose is to build a RAG application for efficient searching in the documents with Llama Index.

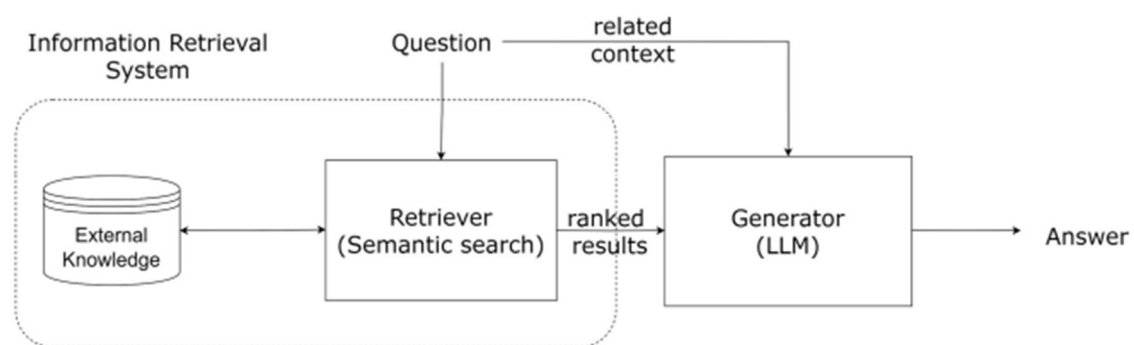
## Solution Strategy

Build a solution which should solve the following requirements using LlamaIndex:

- Users would responses from insurance policy knowledge base.
- If user want to perform a query system must be able to response to query accurately.

## Recap for RAG

## Retrieval Augmented Generation (RAG)



## Goal

Solving the above two requirements well in and would ensure that the accuracy of the overall model is good.

## Data Used

HDFC various Insurance policy documents stored in single folder.

## Tools used

LlamaIndex, ChatGPT has been used due to its powerful query engine, fast data processing using data loaders and directory readers as well as easier and faster implementation using fewer lines of code.

## Model Used

OpenAI from Llama\_Index with model="gpt-3.5-turbo"

## Why LlamaIndex ?

LlamaIndex is an innovative data framework specially designed to support LLM-based RAG framework application development. It offers an advanced framework that empowers developers to integrate diverse data sources with large language models

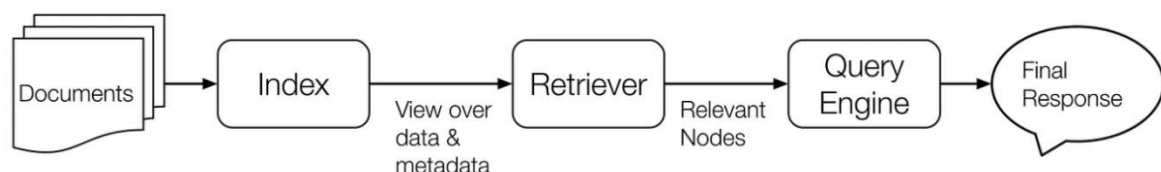
LlamaIndex includes a variety of file formats, such as PDFs and PowerPoints, as well as applications like Notion and Slack and even databases like Postgres and MongoDB.

The framework brings an array of connectors that assist in data ingestion, facilitating a seamless interaction with LLMs. Moreover, LlamaIndex boasts an efficient data retrieval and query interface.

LlamaIndex enables developers to input any LLM prompt and, in return, receive an output that is both context-rich and knowledge-augmentation.

Key Feature of LlamaIndex:

- Data connectors allow ingestion from various data sources and formats.
- It can synthesize data from multiple documents or heterogeneous data sources.
- It provides numerous integrations with vector stores, ChatGPT plugins, tracing tools, LangChain, and more.



## Documents and Nodes

Documents in LlamaIndex may be different from your traditional perception of documents. Document and Node objects are core abstractions within LlamaIndex.

A **Document** is a generic container around any data source - for instance, a PDF, an API output, or retrieved data from a database. They can be constructed manually or created automatically via data loaders. By default, a Document stores text along with some other attributes. Some of these are

metadata - a dictionary of annotations that can be appended to the text (basically, additional info about the document)

relationships - a dictionary containing relationships to other Documents/Nodes.

A **Node** represents a "chunk" of a source Document, whether that is a text chunk, an image, or other. Like Documents, they contain metadata and relationship information with other nodes.

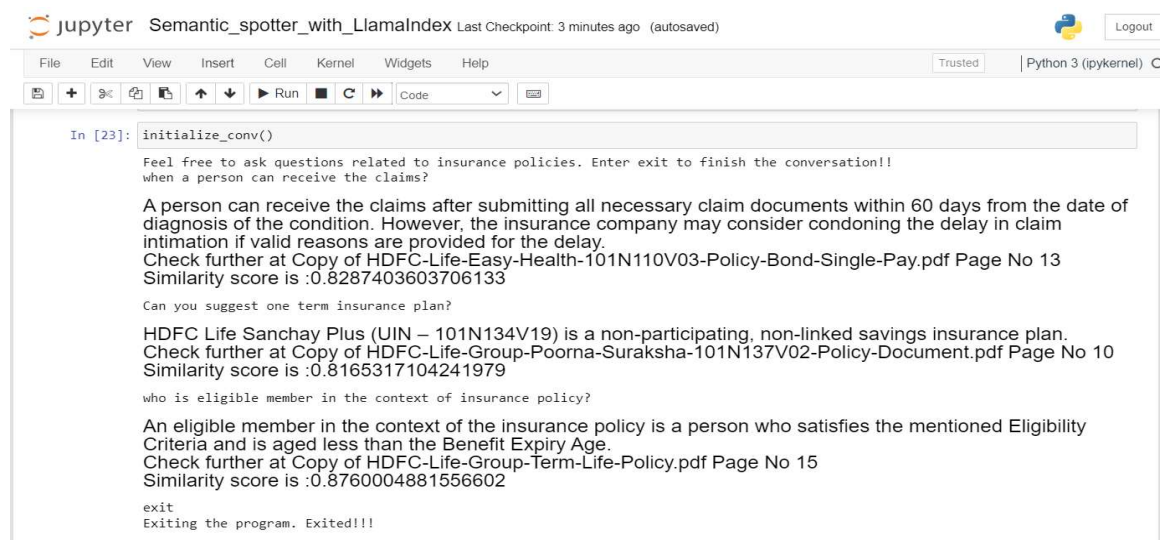
Nodes are a first-class citizen in LlamaIndex. You can choose to define Nodes and all its attributes directly. You may also choose to "parse" source Documents into Nodes through our NodeParser classes. By default, every Node derived from a Document will inherit the same metadata from that Document (e.g. a "file\_name" filed in the Document is propagated to every Node).

Both Documents and Nodes have unique identifies called ID. These can be set automatically or manually. ID is generally used to identify, update, and define relationships between documents (or nodes).

## Generative Search Response from Insurance documents:

Below are attached custom query generative search results.

### Conversational Message – Single Query:



The screenshot shows a Jupyter Notebook titled "Semantic\_spotter\_with\_LlamaIndex". The code cell contains a function `initialize_conv()` that initializes a chatbot. The chatbot's response is displayed below the code. The chatbot's response is a conversational message that provides information about insurance policies and claims. The chatbot's response is as follows:

```
In [23]: initialize_conv()

Feel free to ask questions related to insurance policies. Enter exit to finish the conversation!!
when a person can receive the claims?

A person can receive the claims after submitting all necessary claim documents within 60 days from the date of
diagnosis of the condition. However, the insurance company may consider condoning the delay in claim
intimation if valid reasons are provided for the delay.
Check further at Copy of HDFC-Life-Easy-Health-101N110V03-Policy-Bond-Single-Pay.pdf Page No 13
Similarity score is :0.8287403603706133

Can you suggest one term insurance plan?

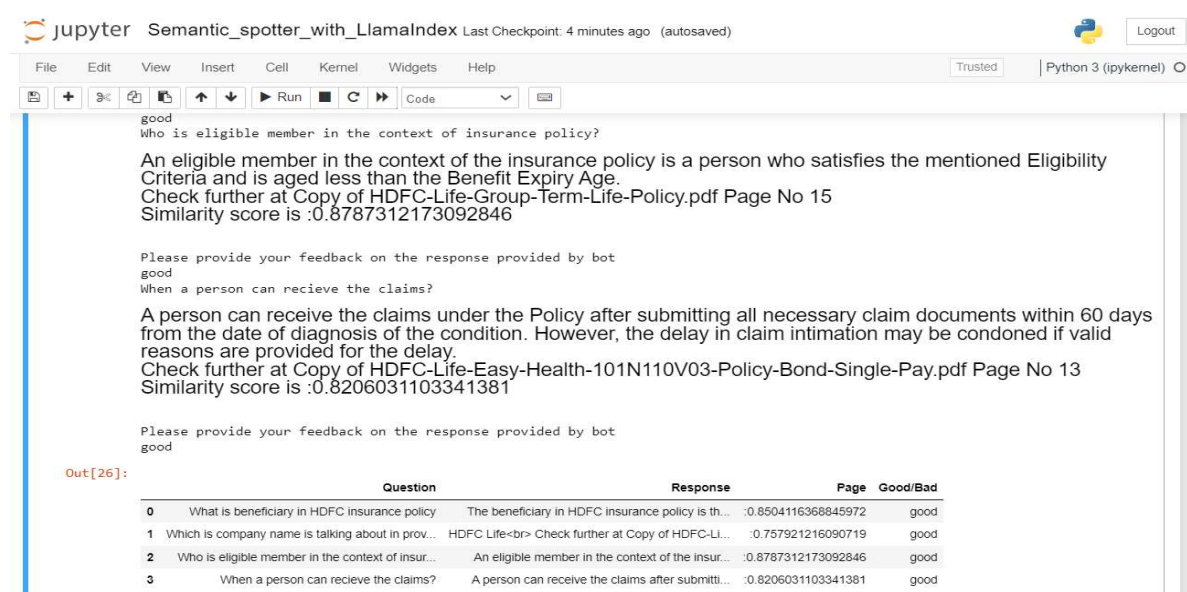
HDFC Life Sanchay Plus (UIN – 101N134V19) is a non-participating, non-linked savings insurance plan.
Check further at Copy of HDFC-Life-Group-Poorna-Suraksha-101N137V02-Policy-Document.pdf Page No 10
Similarity score is :0.8165317104241979

who is eligible member in the context of insurance policy?

An eligible member in the context of the insurance policy is a person who satisfies the mentioned Eligibility
Criteria and is aged less than the Benefit Expiry Age.
Check further at Copy of HDFC-Life-Group-Term-Life-Policy.pdf Page No 15
Similarity score is :0.8760004881556602

exit
Exiting the program. Exited!!!
```

### Testing pipeline – Multiple Queries:



The screenshot shows a Jupyter Notebook titled "Semantic\_spotter\_with\_LlamaIndex". The code cell contains a function `good` that tests the chatbot's response. The chatbot's response is displayed below the code. The chatbot's response is a conversational message that provides information about insurance policies and claims. The chatbot's response is as follows:

```
good
Who is eligible member in the context of insurance policy?

An eligible member in the context of the insurance policy is a person who satisfies the mentioned Eligibility
Criteria and is aged less than the Benefit Expiry Age.
Check further at Copy of HDFC-Life-Group-Term-Life-Policy.pdf Page No 15
Similarity score is :0.8787312173092846

Please provide your feedback on the response provided by bot
good
When a person can receive the claims?

A person can receive the claims under the Policy after submitting all necessary claim documents within 60 days
from the date of diagnosis of the condition. However, the delay in claim intimation may be condoned if valid
reasons are provided for the delay.
Check further at Copy of HDFC-Life-Easy-Health-101N110V03-Policy-Bond-Single-Pay.pdf Page No 13
Similarity score is :0.8206031103341381

Please provide your feedback on the response provided by bot
good
```

Out[26]:

	Question	Response	Page	Good/Bad
0	What is beneficiary in HDFC insurance policy?	The beneficiary in HDFC insurance policy is th...	:0.8504116368845972	good
1	Which is company name is talking about in prov...	HDFC Life  Check further at Copy of HDFC-Li...	:0.757921216090719	good
2	Who is eligible member in the context of insur...	An eligible member in the context of the insur...	:0.8787312173092846	good
3	When a person can receive the claims?	A person can receive the claims after submitti...	:0.8206031103341381	good

## **Code**

Jupyter Notebook developed for Semantic Spotter.

## **Challenges Faced**

Tried to get better score with different techniques.

Faced challenges to implement caching mechanism.

## **Future Work**

Based on the current POC, it can be understood where to improve further for better results.

Overall score can be more efficient and better by cleaning up data

Some more tools can be explored to improve the POC further and for more better accuracy and fast retrieval.