Semantic Spotter - RAG Project

Prepared by – Manish Thapa   Date: 24-Nov-2024
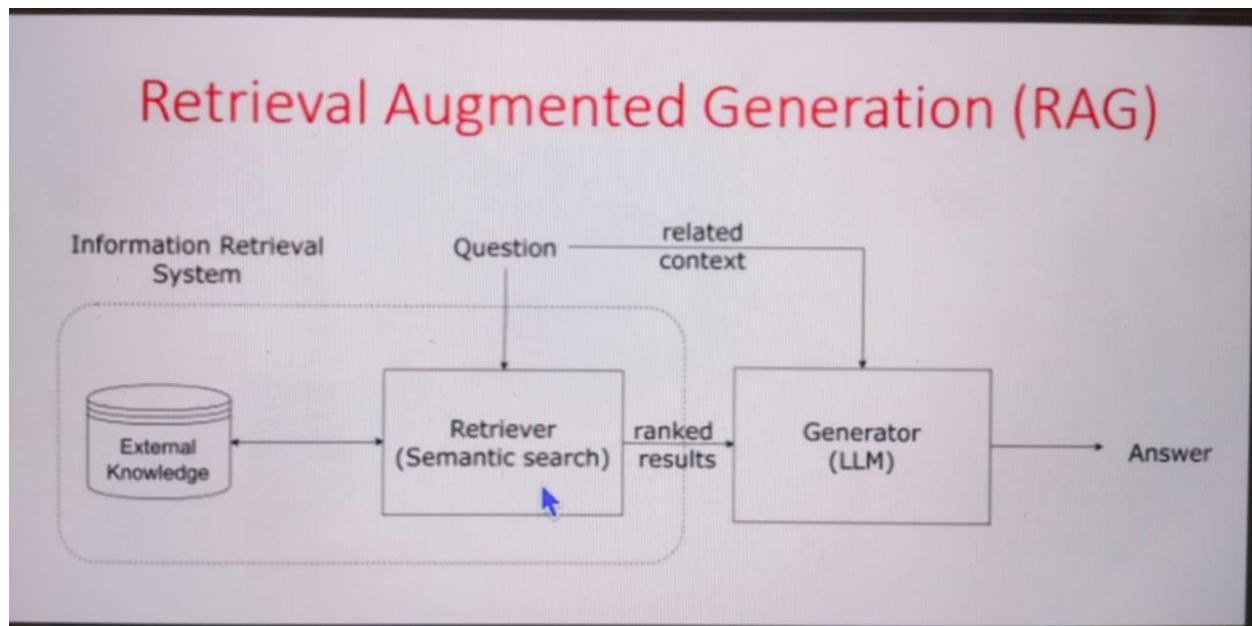
Goal:

Build a robust generative search system capable of effectively and accurately answering questions from various policy documents. You may use LangChain or LlamaIndex to build the generative search application.

Data Source: Policy documents present in PDF form as below:

https://github.com/ManishThapa27/semantic-spotter-project/tree/main/insurance-documents

Collab Notebook PDF: https://github.com/ManishThapa27/semantic-spotter-project/tree/main/Semantic-spotter-project.ipynb%20-%20Colab.pdf

System Design:

RAG (Retrieval-Augmented Generation) is a framework in GenAI that combines the strengths of both retrieval-based and generative models to create more efficient AI assistants. It uses a retrieval mechanism to find relevant information from a large dataset and then generates a response based on that information. This approach improves the accuracy and relevance of responses compared to purely generative models.

LangChain and Llama index can be instrumental in building an efficient AI assistant using RAG. LangChain is a blockchain-based language platform that provides secure and scalable language services. It can be integrated with RAG to enhance its language processing capabilities, allowing the AI assistant to understand and generate more complex and nuanced responses.

On the other hand, Llama index is a large-scale index for efficient language model retrieval. It can be used to quickly retrieve relevant information from a large dataset, which is crucial for the retrieval component of RAG. By leveraging the speed and efficiency of Llama index, the AI assistant can provide faster and more accurate responses to user queries. **We would be using Llama Index for implementing RAG, for efficient searching in insurance policy documents.**

Integrating LangChain and Llama index with RAG can significantly improve the performance and efficiency of AI assistants, making them more capable of understanding and responding to user queries in a more natural and human-like manner.

Implementation:

```python
# Importing necessary libraries
!pip install llama-index


# Document loaders for SimpleDirectoryReader
!pip install docx2txt
!pip install pypdf

# Install OpenAI
!pip install openai

#import openAI
from llama_index.llms.openai import OpenAI
#import ChatMessage
from llama_index.core.llms import ChatMessage
#import os
import os
import openai


from google.colab import drive
drive.mount('/content/drive', force_remount=True)


#Set API key
filepath="/content/drive/MyDrive/GenAI/RAG-project/"
with open(filepath + "openai_api_key.txt","r") as f:
  openai.api_key=''.join(f.readlines())


#import SimpleDirectoryReader
from llama_index.core import SimpleDirectoryReader

# Create object of SimpleDirectoryReader
```

```
reader=SimpleDirectoryReader(input_dir="/content/drive/MyDrive/GenAI/RAG-
DEMO/Policy-Documents/")
```

```
documents=reader.load_data()
#lenth of documents
print(f"Loaded {len(documents)} documents/pages successfully.")
```

Building Query Engine:

```python
# import SimpleNodeParser
from llama_index.core.node_parser import SimpleNodeParser
# import VectorStoreIndex
from llama_index.core import VectorStoreIndex
# import display, HTML
from IPython.display import display, HTML


# Create parser and parse docuemnts into nodes
parser=SimpleNodeParser.from_defaults()
nodes=parser.get_nodes_from_documents(documents)

# build index
index=VectorStoreIndex(nodes)

#construct query engine
query_engine=index.as_query_engine()
```

```python
#query
response=query_engine.query("What provisions may allow for a longer
reinstatement period for an approved leave of absence taken in accordance
with the Uniformed Services Employment and Reemployment Rights Act of 1994
(USERRA)?")
```

```
response.response
```

```
response.response

'The provisions that may allow for a longer reinstatement period for an approved leave of absence taken in accordance with the Uniformed Services
Employment and Reemployment Rights Act of 1994 (USERRA) could include specific clauses in the policy document that outline extended reinstatement
timelines or special considerations for individuals returning from military service.'
```

```
response.metadata
```

```
{'0ceded9c-504d-4879-bc59-d0eb06b6f0d2': {'page_label': '11',
  'file_name': 'HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf',
  'file_path': '/content/drive/MyDrive/GenAI/RAG-DEMO/Policy-Documents/HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf',
  'file_type': 'application/pdf',
  'file_size': 1990500,
  'creation_date': '2024-03-31',
  'last_modified_date': '2024-03-31'},
 '0685a662-ec44-4210-ba8e-5269dfe87c04': {'page_label': '9',
  'file_name': 'HDFC-Life-Group-Poorna-Suraksha-101N137V02-Policy-Document.pdf',
  'file_path': '/content/drive/MyDrive/GenAI/RAG-DEMO/Policy-Documents/HDFC-Life-Group-Poorna-Suraksha-101N137V02-Policy-Document.pdf',
  'file_type': 'application/pdf',
  'file_size': 1371541,
  'creation_date': '2024-03-31',
  'last_modified_date': '2024-03-31'}}
```

```
response.source_nodes
```

```
[NodeWithScore(node=TextNode(id_='0ceded9c-504d-4879-bc59-d0eb06b6f0d2', embedding=None, metadata={'page_label': '11', 'file_name': 'HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf', 'file_path': '/content/drive/MyDrive/GenAI/RAG-DEMO/Policy-Documents/HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf', 'file_type': 'application/pdf', 'file_size': 1990500, 'creation_date': '2024-03-31', 'last_modified_date': '2024-03-31'}, excluded_embed_metadata_keys=['file_name', 'file_type', 'file_size', 'creation_date', 'last_modified_date', 'last_accessed_date'], excluded_llm_metadata_keys=['file_name', 'file_type', 'file_size', 'creation_date', 'last_modified_date', 'last_accessed_date'], relationships={<NodeRelationship.SOURCE: '1'>: RelatedNodeInfo(node_id='c58796a5-51a4-4640-a415-c87538acb693', node_type=<ObjectType.DOCUMENT: '4'>, metadata={'page_label': '11', 'file_name': 'HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf', 'file_path': '/content/drive/MyDrive/GenAI/RAG-DEMO/Policy-Documents/HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf', 'file_type': 'application/pdf', 'file_size': 1990500, 'creation_date': '2024-03-31', 'last_modified_date': '2024-03-31'}, hash='ba37b4e89a59994884d91e2c286f3fc935302ec3d1a8487bd36c0fa248a6df16'), <NodeRelationship.PREVIOUS: '2'>: RelatedNodeInfo(node_id='68e4835f-4630-44ab-be92-0c83df10325e', node_type=<ObjectType.TEXT: '1'>, metadata={'page_label': '11', 'file_name': 'HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf', 'file_path': '/content/drive/MyDrive/GenAI/RAG-DEMO/Policy-Documents/HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf', 'file_type': 'application/pdf', 'file_size': 1990500, 'creation_date': '2024-03-31', 'last_modified_date': '2024-03-31'}, hash='64e566f4b08c76a2bb6d6f4345464c5cbb3cf76dd5008a98a3852ca6421c497b')}, text='D.2.2.  Notwithstanding anything to the contrary contained elsewhere in this Policy, the Company reserves the right to revive \nthe lapsed Policy either on its original terms and conditions or on such other or modified terms and conditions as the \nCompany may specify or to reject the Revival . If needed the Company may refer it to its medical examiner in decid ing \non Revival  of lapse d Policy. Subject to the provisions of Clauses D.2.1 above, the Revival  shall come into effect on the \ndate when the Company specifically communicates it in writing to the Policyholder. \n \nD.2.3  If the Policy is not revived for full Benefits before the Policy  Maturity Date but within five years from the due date for \npayment of the first unpaid Premium and if the Policy has not acquired Guaranteed Surrender Value, then the Policy \nwill terminate. \n \nD.3.  Non-Forfeiture options : PART D  \nPolicy Servicing Related Aspects', start_char_idx=3529, end_char_idx=4449, text_template='{metadata_str}\n\n{content}', metadata_template='{key}: {value}', metadata_seperator='\n'), score=0.7711361909629276),
 NodeWithScore(node=TextNode(id_='0685a662-ec44-4210-ba8e-5269dfe87c04', embedding=None, metadata={'page_label': '9', 'file_name': 'HDFC-Life-Group-Poorna-Suraksha-101N137V02-Policy-Document.pdf', 'file_path': '/content/drive/MyDrive/GenAI/RAG-DEMO/Policy-Documents/HDFC-Life-Group-Poorna-Suraksha-101N137V02-Policy-Document.pdf', 'file_type': 'application/pdf', 'file_size': 1371541, 'creation_date': '2024-03-31', 'last_modified_date': '2024-03-31'}, excluded_embed_metadata_keys=['file_name', 'file_type', 'file_size', 'creation_date', 'last_modified_date', 'last_accessed_date'], excluded_llm_metadata_keys=['file_name', 'file_type', 'file_size', 'creation_date', 'last_modified_date', 'last_accessed_date'], relationships={<NodeRelationship.SOURCE: '1'>: RelatedNodeInfo(node_id='9d424dc8-6e10-4081-abb3-5407654f1dca', node_type=
```

```python
# Extract the score
print(response.source_nodes[0].score)
```

```python
# Response Node Text
response.source_nodes[0].node.text
```

Creating Response Pipeline:

```python
# Query response function
def query_response(user_input):
  response=query_engine.query(user_input)
  file_name=response.source_nodes[0].node.metadata['file_name'] + " Page
No " + response.source_nodes[0].node.metadata['page_label']
  final_response=response.response + "\nCheck further at " + file_name
  return final_response
```

```python
def initialize_conv():
```

```python
    print("Feel free to ask questions related to insurance policies. Enter
exit once you are done!")
    while True:
        user_input=input()
        if user_input.lower() == "exit":
            print("Exiting the program. Bye!!!")
            break
        else:
            response=query_response(user_input)
            display(HTML(f'<p style="font-size:20px">{response}</p>'))
```

```python
initialize_conv()
```

```
initialize_conv()
```

Feel free to ask questions related to insurance policies. Enter exit once you are done!
What provisions may allow for a longer reinstatement period for an approved leave of absence taken in accordance with the Uniformed Services Emplo

The provisions that may allow for a longer reinstatement period for an approved leave of absence taken in accordance with the Uniformed Services Employment and Reemployment Rights Act of 1994 (USERRA) could include clauses that specify the right of the company to revive a lapsed policy on modified terms and conditions, or to reject the revival altogether. Additionally, the reinstatement may come into effect on the date when the company communicates it in writing to the policyholder. Check further at HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf Page No 11

exit
Exiting the program. Bye!!!

Building a Test Pipeline:

```python
questions=["What provisions may allow for a longer reinstatement period
for an approved leave of absence taken in accordance with the Uniformed
Services Employment and Reemployment Rights Act of 1994 (USERRA)?",
            "How is the peroid of time during which a reinstated Member's
insurance was not in force treated for the purpose of determining the
length of continuous coverage under the Group Policy?",
            "What are the requirements for placing in force any Scheduled
benefit that would have been subject to Proof of Good Health has the
member remained continuously insured?"]
```

```python
import pandas as pd

def testing_pipeline(questions):
    test_feedback=[]
    for i in questions:
        print(i)
        print(query_response(i))
        print("\nPlease provide your feedback on the response provided by
bot")
```

```
    user_input=input()
    page=query_response(i).split()[-1]
    test_feedback.append((i,query_response(i),page,user_input))

  feedback_df=pd.DataFrame(test_feedback,columns=["Question","Response","P
age","Good/Bad"])
  return feedback_df
```

```
testing_pipeline(questions)
```



```
testing_pipeline(questions)

What provisions may allow for a longer reinstatement period for an approved leave of absence taken in accordance with the Uniformed Services Emplo
Provisions related to the revival of a lapsed policy may allow for a longer reinstatement period for an approved leave of absence taken in accorda
Check further at HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf Page No 11

Please provide your feedback on the response provided by bot
Yes
How is the peroid of time during which a reinstated Member's insurance was not in force treated for the purpose of determining the length of conti
The period of time during which a reinstated Member's insurance was not in force is typically not considered when determining the length of contir
Check further at HDFC-Life-Group-Term-Life-Policy.pdf Page No 15

Please provide your feedback on the response provided by bot
Good
What are the requirements for placing in force any Scheduled benefit that would have been subject to Proof of Good Health has the member remained
The requirements for placing in force any Scheduled benefit that would have been subject to Proof of Good Health if the member remained continuous
Check further at HDFC-Life-Group-Term-Life-Policy.pdf Page No 15

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

	Question	Response	Page	Good/Bad
0	What provisions may allow for a longer reinsta...	The provisions that may allow for a longer rei...	11	Yes
1	How is the peroid of time during which a reins...	The period of time during which a reinstated M...	15	Good
2	What are the requirements for placing in force...	The requirements for placing in force any Sche...	15	Bad

Building a Custom Prompt Template:

```
response=query_engine.query("What provisions may allow for a longer
reinstatement period for an approved leave of absence taken in accordance
with the Uniformed Services Employment and Reemployment Rights Act of 1994
(USERRA)?")
```

```
retrieved = response.source_nodes[0].node.text +
response.source_nodes[1].node.text

messages=[
        {
            "role":"system",
            "content":"You are AI assistent to user."
        },
        {
            "role":"user",
```

```
            "content": f"""What provisions may allow for a longer
reinstatement period for an approved leave of absence taken
            in accordance with the Uniformed Services Employment and
Reemployment Rights Act of 1994 (USERRA)? Check in '{retrieved}'
            """
        }
    ]
```

```
response2=openai.chat.completions.create(
    model="gpt-3.5-turbo",
    messages=messages
)
response2.choices[0].message.content
```

```
response2=openai.chat.completions.create(
    model="gpt-3.5-turbo",
    messages=messages
)
response2.choices[0].message.content
```

'The provisions that may allow for a longer reinstatement period for an approved leave of absence taken in accordance with the Uniformed Services Employment and Reemployment Rights Act of 1994 (USERRA) are typically outlined in employer policies or collective bargaining agreements. However, based on the information provided, it appears that the reinstatement period may be extended if the policy is not revived for full benefits before the Policy Maturity Date but within five years from the due date for payment of the first unpaid Premium. If the policy has not acquired Guaranteed Surrender Value, then the policy will terminate.\n\nIt's important to review the specific terms and conditions of the policy or agreement in question to determine the exact provisions related to the reinstatement period for approved leaves of absence under USERRA.'

Further Improvement:

This model can be further improved by using better/cleaner dataset or utilizing better data pre-processing techniques.

We can also use custom node and LLMs for better results, shown as below:

```python
#import OpenAIEmbedding
from llama_index.embeddings.openai import OpenAIEmbedding
#import SentenceSplitter
from llama_index.core.node_parser import SentenceSplitter
#import OpenAI
from llama_index.llms.openai import OpenAI
#import Settings
from llama_index.core import Settings

#Initialize the openAI model
Settings.lm=OpenAI(model="gpt-3.5-turbo", temperature=0, max_tokens=256)

#Initialize the embedding model
Settings.embed_model=OpenAIEmbedding()

#Initialize the node_parser with custom node settings
Settings.node_parser=SentenceSplitter(chunk_size=512, chunk_overlap=20)

# Initialize the num_output and context window
Settings.num_output=512
Settings.context_window=3900

#Create a VectorStoreIndex from a list of documents using the service
context
index=VectorStoreIndex.from_documents(documents)

# Initialize a query engine for the index with a specified similiarity
with top-k values
query_engine=index.as_query_engine(similarity_top_k=3)
```

```python
#Query the engine with specific question

response=query_engine.query("""What provisions may allow for a longer
reinstatement period for an approved leave of absence taken
              in accordance with the Uniformed Services Employment and
Reemployment Rights Act of 1994 (USERRA)?""")
```

```python
response.response
```

```
'The provisions that may allow for a longer reinstatement period for an approved leave of absence taken in accordance with the Uniformed Services
Employment and Reemployment Rights Act of 1994 (USERRA) are typically subject to the Board Approved Underwriting Policy of the insurer. These pro
visions may include considering requests from the policyholder to revive lapsed or paid-up policies within a specified timeframe, usually within
five years from the date of lapse or becoming paid-up, and payment of any unpaid premiums with applicable interest as determined by the insure
r.'
```