BGSW (Bosch) - Fit.Fest'24 University Hackathon

PHASE 1

Storage and Retrieval of Document text & table

College: PSG College of Technology

Team Members:
Aadil Arsh SR   - 21z201
Kavin Dev R   - 21z224
Sanjay Kumaar Eswaran   - 21z248

Table of contents

01   Deliverables

- Ability to retrieve and store entire textual content available in the form of text and tables into Vector DB efficiently.

- Ability to provide relevant and accurate response to user queries.

- Probing questions to get more clarity on user prompts

- Latency of response to user prompt in less than a minute.

- Basic UI application to demonstrate Chat- Q&A.

02 Requirements

- `langchain`
- `huggingface_hub`
- `sentence_transformers`
- `os`
- `langchain-community`
- `TextLoader` from `langchain.document_loaders`
- `CharacterTextSplitter` from `langchain.text_splitter`
- `HuggingFaceEmbeddings` from `langchain.embeddings`
- `faiss-cpu`
- `FAISS` from `langchain.vectorstores`
- `sumy`
- `nltk`
- `PlaintextParser` from `sumy.parsers.plaintext`
- `Tokenizer` from `sumy.nlp.tokenizers`
- `LexRankSummarizer` from `sumy.summarizers.lex_rank`
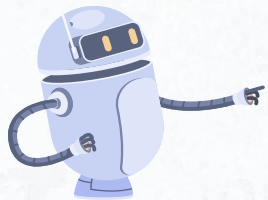- `load_qa_chain` from `langchain.chains.question_answering`

02 Requirements

- **HuggingFaceHub** from **langchain**
- **UnstructuredPDFLoader** from **langchain.document_loaders**
- **VectorstoreIndexCreator** from **langchain.indexes**
- **drive** from **google.collab**
- **RetrievalQA** from **langchain.chains**

03 **Approach**

1. Library Installation:

- Necessary Python libraries like `langchain`, `huggingface_hub`, `sentence_transformers`, `faiss-cpu`, `sumy`, and others that support natural language processing and vector space models are installed.

2. Document Loading:

- Using `langchain.document_loaders.TextLoader`, the program loads documents from a specified text/pdf file.

3. Text Processing:

- It includes a custom function to wrap text while preserving new lines, adjusting for readability.
- `langchain.text_splitter.CharacterTextSplitter` is used to split documents into chunks based on character count, facilitating manageable processing sizes.

Approach
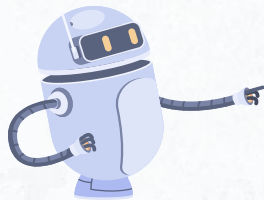
4. Embeddings and Vector Store:

- Utilizes `langchain.embeddings.HuggingFaceEmbeddings` to generate embeddings for the document texts.
- Creates a vector store using `FAISS` (from the `langchain.vectorstores` module), which is a highly efficient similarity search library and clustering of dense vectors.

5. Query Handling and Summarization:

- The document chunks are then queried using vector similarity to find relevant sections.
- Summarization of the found text is done using the `LexRankSummarizer` from the `sumy` library.
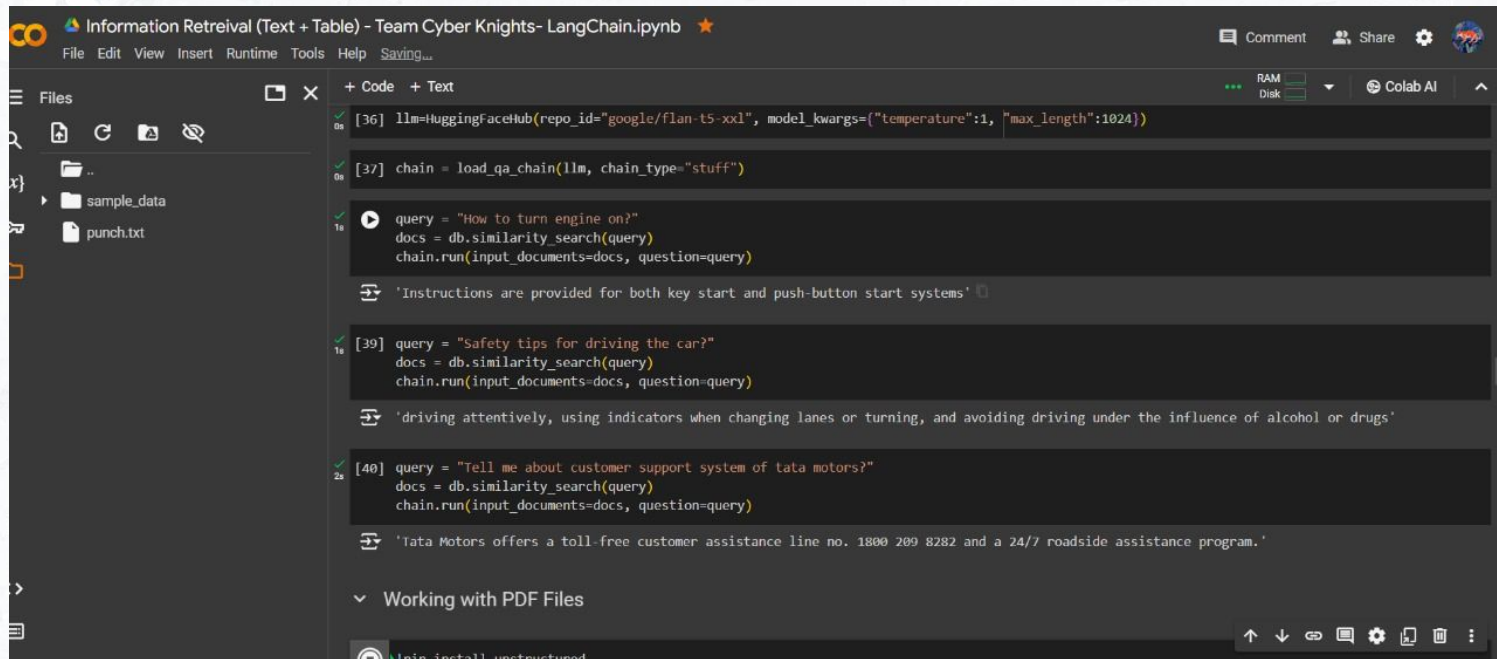
03 Approach
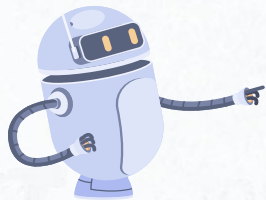
6. Question Answering:

- Integrates a question-answering model from HuggingFace, specifically using `flan-t5-xxl`.
- Implements `langchain.chains.question_answering.load_qa_chain` to process queries and provide answers based on the relevant text extracted from the documents.

7. Additional Dependencies:

- The program also prepares for further extensions by installing additional libraries like `unstructured`, `chromadb`, `Cython`, and `tiktoken`.

04 Output

Information Retreival (Text + Table) - Team Cyber Knights- LangChain.ipynb ⭐

File Edit View Insert Runtime Tools Help Saving...

💬 Comment   🧑‍🤝‍🧑 Share   ⚙️

+ Code + Text

```
[36] llm=HuggingFaceHub(repo_id="google/flan-t5-xxl", model_kwargs={"temperature":1, "max_length":1024})
```

```
[37] chain = load_qa_chain(llm, chain_type="stuff")
```

```
query = "How to turn engine on?"
docs = db.similarity_search(query)
chain.run(input_documents=docs, question=query)
```

'Instructions are provided for both key start and push-button start systems'

```
[39] query = "Safety tips for driving the car?"
docs = db.similarity_search(query)
chain.run(input_documents=docs, question=query)
```

'driving attentively, using indicators when changing lanes or turning, and avoiding driving under the influence of alcohol or drugs'

```
[40] query = "Tell me about customer support system of tata motors?"
docs = db.similarity_search(query)
chain.run(input_documents=docs, question=query)
```

'Tata Motors offers a toll-free customer assistance line no. 1800 209 8282 and a 24/7 roadside assistance program.'

∨ Working with PDF Files

!pip install unstructured
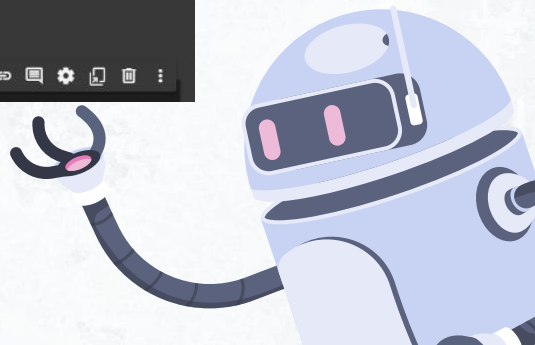
04 Output

```
[37] chain.run('When shoudnt airbag be deployed?')
```

'The amount of visible vehicle damage is not always the correct indicator for airbag deployment. Some collisions can result in visible damage but no airbag SAFETY deployment because the airbags would not have been needed or would not have provided protection even if they had de- ployed. Seat belts, if worn, offer adequate occu- pant protection in such cases.'


```python
from pyngrok import ngrok

# Set up your authtoken from ngrok (you should have already done this step previously)
ngrok.set_auth_token('2FGGrbhYvz2ypBM0ImOEAFVYyW7_4MC8JC3jMw9uiYeHvW57p')

# Setup a tunnel to the streamlit port 8501
public_url = ngrok.connect(8501)
print('Public URL:', public_url)
!streamlit run streamlit_app.py --server.port 8501
```
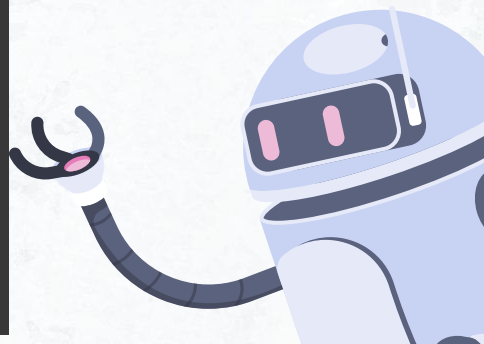
Public URL: NgrokTunnel: "https://18c0-34-81-91-176.ngrok-free.app" -> "http://localhost:8501"

Collecting usage statistics. To deactivate, set browser.gatherUsageStats to false.

  You can now view your Streamlit app in your browser.

  Network URL: http://172.28.0.12:8501
  External URL: http://34.81.91.176:8501

04 Output


```
[39] chain.run('How is abs deployed?')
```

'ABS intervenes when braking, you will feel a pulsing in the brake pedal.'

SAFETY

ANTI-LOCK BRAKING SYSTEM (ABS)

ABS regulates brake pressure in such a way that the wheels do not lock when you brake. This allows you to continue steering the vehicle when braking.

The ABS warning lamp in the instrument cluster lights up when the ignition is switched on. It goes out when the engine is running.

⚠ WARNING

While Braking

- If ABS intervenes: continue to press the brake pedal vigorously until the braking situation is over.
- To make a full brake application: press the brake pedal with full force.

If ABS intervenes when braking, you will feel a pulsing in the brake pedal. The pulsating brake pedal can be an indication of hazardous road conditions, and functions as a reminder to take extra care while driving.

consideration prevailing weather and traffic conditions.

```
[38] chain.run('How should we lock the door of cars?')
```

'The front doors can be locked or un- locked from outside using the key blade. Insert the key in the slot and turn it clock- wise to lock and anticlockwise to unlock the door.'

DOORS

Door Lock And Unlock With Key (if available)

The front doors can be locked or unlocked from outside using the key blade.

Door Locking And Unlocking Using Door Handle Switch (DHS)

To lock/unlock all the doors without operating smart key button/ key blade. Press the door handle switch (DHS) provided on the driver door to lock/unlock all the four doors including Tail gate.
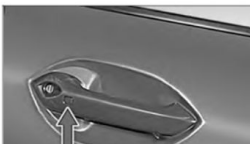
OPENING AND CLOSI

- *Passive entry works only when ig tion is off.*

Rear Door Opening

Door opening handle is provided on side of the window.

04 Output
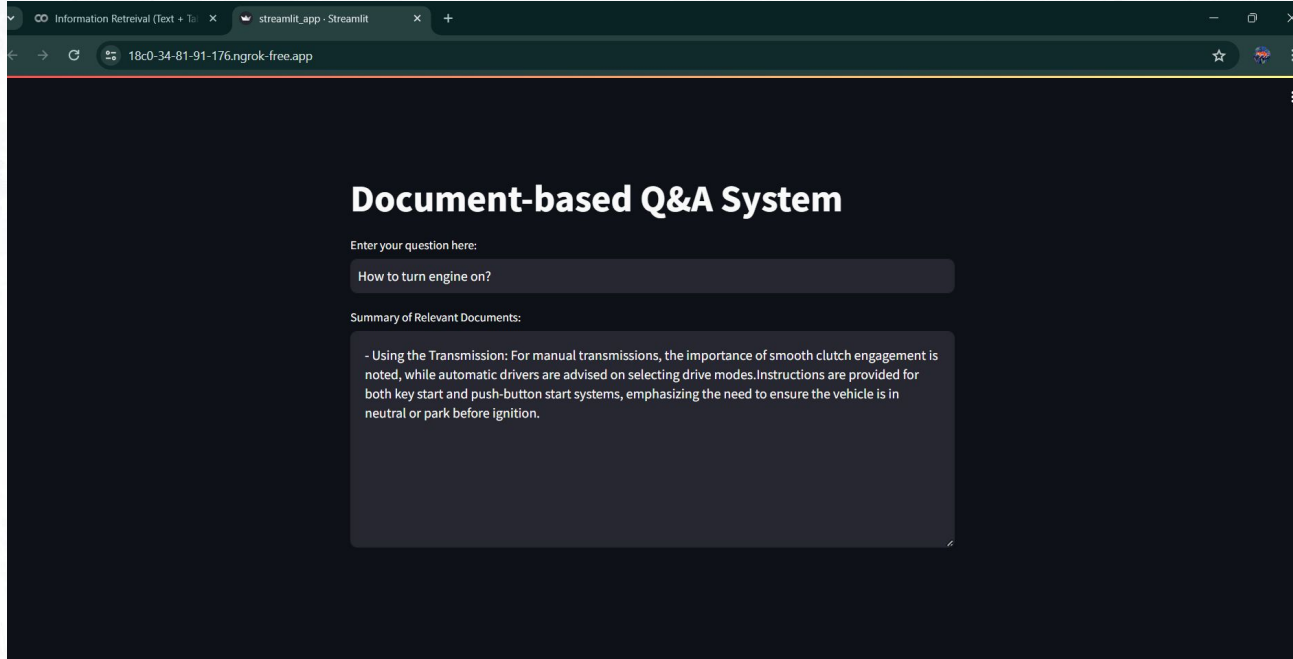
table-team-cyber-knights-langchain

May 21, 2024

0.0.1 Information Retreival (Text + Table) - Team Cyber Knights- LangChain

Two Applications: - Text Documents - Multiple PDF Files

```
[5]: !pip install langchain
     !pip install huggingface_hub
     !pip install sentence_transformers
```

Requirement already satisfied: langchain in /usr/local/lib/python3.10/dist-
packages (0.2.0)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.10/dist-
packages (from langchain) (6.0.1)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in
/usr/local/lib/python3.10/dist-packages (from langchain) (2.0.30)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in
/usr/local/lib/python3.10/dist-packages (from langchain) (3.9.5)
Requirement already satisfied: async-timeout<5.0.0,>=4.0.0 in
/usr/local/lib/python3.10/dist-packages (from langchain) (4.0.3)
Requirement already satisfied: dataclasses-json<0.7,>=0.5.7 in
/usr/local/lib/python3.10/dist-packages (from langchain) (0.6.6)
Requirement already satisfied: langchain-core<0.3.0,>=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from langchain) (0.2.0)
Requirement already satisfied: langchain-text-splitters<0.3.0,>=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from langchain) (0.2.0)
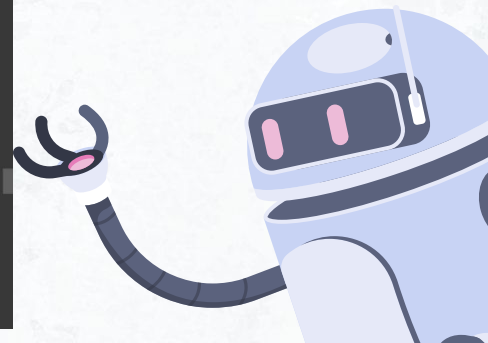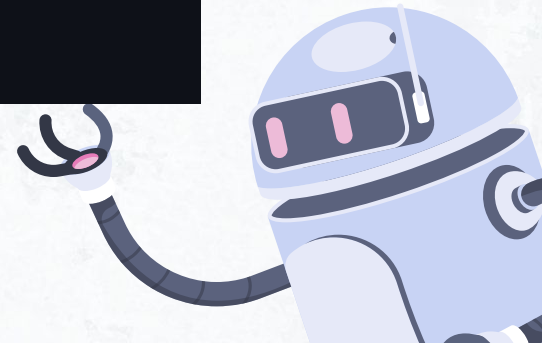Requirement already satisfied: langsmith<0.2.0,>=0.1.17 in
/usr/local/lib/python3.10/dist-packages (from langchain) (0.1.60)
Requirement already satisfied: numpy<2,>=1 in /usr/local/lib/python3.10/dist-
packages (from langchain) (1.25.2)
Requirement already satisfied: pydantic<3,>=1 in /usr/local/lib/python3.10/dist-
packages (from langchain) (2.7.1)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.10/dist-
packages (from langchain) (2.31.0)
Requirement already satisfied: tenacity<9.0.0,>=8.1.0 in
/usr/local/lib/python3.10/dist-packages (from langchain) (8.3.0)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.10/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain)
(1.3.1)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-
packages (from aiohttp<4.0.0,>=3.8.3->langchain) (23.2.0)

Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.10/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain)
(1.4.1)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.10/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain)
(6.0.5)
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.10/dist-
packages (from aiohttp<4.0.0,>=3.8.3->langchain) (1.9.4)
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in
/usr/local/lib/python3.10/dist-packages (from dataclasses-
json<0.7,>=0.5.7->langchain) (3.21.2)
Requirement already satisfied: typing-inspect<1,>=0.4.0 in
/usr/local/lib/python3.10/dist-packages (from dataclasses-
json<0.7,>=0.5.7->langchain) (0.9.0)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in
/usr/local/lib/python3.10/dist-packages (from langchain-
core<0.3.0,>=0.2.0->langchain) (1.33)
Requirement already satisfied: packaging<24.0,>=23.2 in
/usr/local/lib/python3.10/dist-packages (from langchain-
core<0.3.0,>=0.2.0->langchain) (23.2)
Requirement already satisfied: orjson<4.0.0,>=3.9.14 in
/usr/local/lib/python3.10/dist-packages (from
langsmith<0.2.0,>=0.1.17->langchain) (3.10.3)
Requirement already satisfied: annotated-types>=0.4.0 in
/usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1->langchain) (0.6.0)
Requirement already satisfied: pydantic-core==2.18.2 in
/usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1->langchain)
(2.18.2)
Requirement already satisfied: typing-extensions>=4.6.1 in
/usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1->langchain)
(4.11.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2->langchain) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests<3,>=2->langchain) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2->langchain) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2->langchain)
(2024.2.2)
Requirement already satisfied: greenlet!=0.4.17 in
/usr/local/lib/python3.10/dist-packages (from SQLAlchemy<3,>=1.4->langchain)
(3.0.3)
Requirement already satisfied: jsonpointer>=1.9 in
/usr/local/lib/python3.10/dist-packages (from jsonpatch<2.0,>=1.33->langchain-
core<0.3.0,>=0.2.0->langchain) (2.4)
Requirement already satisfied: mypy-extensions>=0.3.0 in
/usr/local/lib/python3.10/dist-packages (from typing-

inspect<1,>=0.4.0->dataclasses-json<0.7,>=0.5.7->langchain) (1.0.0)
Requirement already satisfied: huggingface_hub in
/usr/local/lib/python3.10/dist-packages (0.20.3)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
packages (from huggingface_hub) (3.14.0)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.10/dist-packages (from huggingface_hub) (2023.6.0)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-
packages (from huggingface_hub) (2.31.0)
Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.10/dist-
packages (from huggingface_hub) (4.66.4)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-
packages (from huggingface_hub) (6.0.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.10/dist-packages (from huggingface_hub) (4.11.0)
Requirement already satisfied: packaging>=20.9 in
/usr/local/lib/python3.10/dist-packages (from huggingface_hub) (23.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->huggingface_hub) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests->huggingface_hub) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->huggingface_hub) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->huggingface_hub)
(2024.2.2)
Requirement already satisfied: sentence_transformers in
/usr/local/lib/python3.10/dist-packages (2.7.0)
Requirement already satisfied: transformers<5.0.0,>=4.34.0 in
/usr/local/lib/python3.10/dist-packages (from sentence_transformers) (4.40.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages
(from sentence_transformers) (4.66.4)
Requirement already satisfied: torch>=1.11.0 in /usr/local/lib/python3.10/dist-
packages (from sentence_transformers) (2.2.1+cu121)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages
(from sentence_transformers) (1.25.2)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-
packages (from sentence_transformers) (1.2.2)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages
(from sentence_transformers) (1.11.4)
Requirement already satisfied: huggingface-hub>=0.15.1 in
/usr/local/lib/python3.10/dist-packages (from sentence_transformers) (0.20.3)
Requirement already satisfied: Pillow in /usr/local/lib/python3.10/dist-packages
(from sentence_transformers) (10.3.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub>=0.15.1->sentence_transformers) (3.14.0)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.10/dist-packages (from huggingface-

hub>=0.15.1->sentence_transformers) (2023.6.0)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub>=0.15.1->sentence_transformers) (2.31.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub>=0.15.1->sentence_transformers) (6.0.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.10/dist-packages (from huggingface-
hub>=0.15.1->sentence_transformers) (4.11.0)
Requirement already satisfied: packaging>=20.9 in
/usr/local/lib/python3.10/dist-packages (from huggingface-
hub>=0.15.1->sentence_transformers) (23.2)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages
(from torch>=1.11.0->sentence_transformers) (1.12)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-
packages (from torch>=1.11.0->sentence_transformers) (3.3)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages
(from torch>=1.11.0->sentence_transformers) (3.1.4)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (12.1.105)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (12.1.105)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (12.1.105)
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in
/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (8.9.2.26)
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in
/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (12.1.3.1)
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in
/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (11.0.2.54)
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in
/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (10.3.2.106)
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in
/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (11.4.5.107)
Requirement already satisfied: nvidia-cusparse-cu12==12.1.0.106 in
/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (12.1.0.106)
Requirement already satisfied: nvidia-nccl-cu12==2.19.3 in
/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (2.19.3)
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in

/usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence_transformers) (12.1.105)
Requirement already satisfied: triton==2.2.0 in /usr/local/lib/python3.10/dist-
packages (from torch>=1.11.0->sentence_transformers) (2.2.0)
Requirement already satisfied: nvidia-nvjitlink-cu12 in
/usr/local/lib/python3.10/dist-packages (from nvidia-cusolver-
cu12==11.4.5.107->torch>=1.11.0->sentence_transformers) (12.4.127)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.10/dist-packages (from
transformers<5.0.0,>=4.34.0->sentence_transformers) (2023.12.25)
Requirement already satisfied: tokenizers<0.20,>=0.19 in
/usr/local/lib/python3.10/dist-packages (from
transformers<5.0.0,>=4.34.0->sentence_transformers) (0.19.1)
Requirement already satisfied: safetensors>=0.4.1 in
/usr/local/lib/python3.10/dist-packages (from
transformers<5.0.0,>=4.34.0->sentence_transformers) (0.4.3)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-
packages (from scikit-learn->sentence_transformers) (1.4.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-
learn->sentence_transformers) (3.5.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from
jinja2->torch>=1.11.0->sentence_transformers) (2.1.5)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->huggingface-
hub>=0.15.1->sentence_transformers) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests->huggingface-hub>=0.15.1->sentence_transformers) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->huggingface-
hub>=0.15.1->sentence_transformers) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->huggingface-
hub>=0.15.1->sentence_transformers) (2024.2.2)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-
packages (from sympy->torch>=1.11.0->sentence_transformers) (1.3.0)

0.0.2 Get HUGGINGFACEHUB_API_KEY

```python
import os
os.environ["HUGGINGFACEHUB_API_TOKEN"] = "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"
```

```
pip install langchain-community
```

Requirement already satisfied: langchain-community in
/usr/local/lib/python3.10/dist-packages (0.2.0)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.10/dist-

packages (from langchain-community) (6.0.1)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in
/usr/local/lib/python3.10/dist-packages (from langchain-community) (2.0.30)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in
/usr/local/lib/python3.10/dist-packages (from langchain-community) (3.9.5)
Requirement already satisfied: dataclasses-json<0.7,>=0.5.7 in
/usr/local/lib/python3.10/dist-packages (from langchain-community) (0.6.6)
Requirement already satisfied: langchain<0.3.0,>=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from langchain-community) (0.2.0)
Requirement already satisfied: langchain-core<0.3.0,>=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from langchain-community) (0.2.0)
Requirement already satisfied: langsmith<0.2.0,>=0.1.0 in
/usr/local/lib/python3.10/dist-packages (from langchain-community) (0.1.60)
Requirement already satisfied: numpy<2,>=1 in /usr/local/lib/python3.10/dist-
packages (from langchain-community) (1.25.2)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.10/dist-
packages (from langchain-community) (2.31.0)
Requirement already satisfied: tenacity<9.0.0,>=8.1.0 in
/usr/local/lib/python3.10/dist-packages (from langchain-community) (8.3.0)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.10/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-
community) (1.3.1)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-
packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (23.2.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.10/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-
community) (1.4.1)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.10/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-
community) (6.0.5)
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.10/dist-
packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (1.9.4)
Requirement already satisfied: async-timeout<5.0,>=4.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-
community) (4.0.3)
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in
/usr/local/lib/python3.10/dist-packages (from dataclasses-
json<0.7,>=0.5.7->langchain-community) (3.21.2)
Requirement already satisfied: typing-inspect<1,>=0.4.0 in
/usr/local/lib/python3.10/dist-packages (from dataclasses-
json<0.7,>=0.5.7->langchain-community) (0.9.0)
Requirement already satisfied: langchain-text-splitters<0.3.0,>=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from
langchain<0.3.0,>=0.2.0->langchain-community) (0.2.0)
Requirement already satisfied: pydantic<3,>=1 in /usr/local/lib/python3.10/dist-
packages (from langchain<0.3.0,>=0.2.0->langchain-community) (2.7.1)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in
/usr/local/lib/python3.10/dist-packages (from langchain-

core<0.3.0,>=0.2.0->langchain-community) (1.33)
Requirement already satisfied: packaging<24.0,>=23.2 in
/usr/local/lib/python3.10/dist-packages (from langchain-
core<0.3.0,>=0.2.0->langchain-community) (23.2)
Requirement already satisfied: orjson<4.0.0,>=3.9.14 in
/usr/local/lib/python3.10/dist-packages (from
langsmith<0.2.0,>=0.1.0->langchain-community) (3.10.3)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2->langchain-
community) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests<3,>=2->langchain-community) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2->langchain-
community) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2->langchain-
community) (2024.2.2)
Requirement already satisfied: typing-extensions>=4.6.0 in
/usr/local/lib/python3.10/dist-packages (from SQLAlchemy<3,>=1.4->langchain-
community) (4.11.0)
Requirement already satisfied: greenlet!=0.4.17 in
/usr/local/lib/python3.10/dist-packages (from SQLAlchemy<3,>=1.4->langchain-
community) (3.0.3)
Requirement already satisfied: jsonpointer>=1.9 in
/usr/local/lib/python3.10/dist-packages (from jsonpatch<2.0,>=1.33->langchain-
core<0.3.0,>=0.2.0->langchain-community) (2.4)
Requirement already satisfied: annotated-types>=0.4.0 in
/usr/local/lib/python3.10/dist-packages (from
pydantic<3,>=1->langchain<0.3.0,>=0.2.0->langchain-community) (0.6.0)
Requirement already satisfied: pydantic-core==2.18.2 in
/usr/local/lib/python3.10/dist-packages (from
pydantic<3,>=1->langchain<0.3.0,>=0.2.0->langchain-community) (2.18.2)
Requirement already satisfied: mypy-extensions>=0.3.0 in
/usr/local/lib/python3.10/dist-packages (from typing-
inspect<1,>=0.4.0->dataclasses-json<0.7,>=0.5.7->langchain-community) (1.0.0)

```python
[7]:  # Document Loader
      from langchain.document_loaders import TextLoader
      loader = TextLoader('./punch.txt')
      documents = loader.load()
```

```python
[8]:  documents
```

[8]: [Document(page_content="Introduction to Ownership: The manual welcomes new
     owners to the Tata Motors family, emphasizing the importance of reading the
     manual to ensure safe and enjoyable driving. Owners are encouraged to perform

prescribed maintenance and repairs at authorized Tata Motors dealers to ensure
vehicle reliability and performance.\n\nCustomer Assistance: Tata Motors offers
a toll-free customer assistance line no. 1800 209 8282 and a 24/7 roadside
assistance program. This ensures support is readily available for any inquiries
or emergencies, including breakdowns.\n\nSafety Overview:\n- Seat Belts: All
occupants should wear seat belts at all times. The manual explains how to
properly fasten the belts and highlights the importance of keeping the belts
untwisted for maximum protection.\n- Child Restraint System (CRS): The use of
appropriate child seats is mandatory for children up to 12 years old, and the
manual specifies how to install these seats using the vehicle's seat belts or
ISOFIX anchors.\n- Airbags: Information is provided on the function and location
of airbags. It is critical not to obstruct airbag deployment zones and to always
maintain a proper seating position to maximize airbag effectiveness.\n- Driving
Safety Tips: Advice includes driving attentively, using indicators when changing
lanes or turning, and avoiding driving under the influence of alcohol or drugs.
There's also guidance on handling various driving conditions like driving
through water, on gradients, and on wet or slippery roads.\n\nVehicle
Operation:\n- Starting the Engine: Instructions are provided for both key start
and push-button start systems, emphasizing the need to ensure the vehicle is in
neutral or park before ignition.\n- Using the Transmission: For manual
transmissions, the importance of smooth clutch engagement is noted, while
automatic drivers are advised on selecting drive modes.\n- Climate Control:
Operation of manual and automatic climate systems is explained to maintain
comfort and visibility in all conditions.\n- Infotainment System: Guidelines on
safe usage, including setting up navigation routes and station presets when the
vehicle is stationary.\n\nMaintenance and Emergency Handling:\n- Routine Checks:
Owners are guided on checking and maintaining fluid levels, tire pressure, and
other critical areas to keep the vehicle in optimal condition.\n- Emergency
Procedures: The manual outlines steps for changing tires, jump-starting the car,
and what to do in the event of an accident.\n- Roadside Assistance: The process
to contact and utilize roadside assistance is clearly laid out for scenarios
like vehicle breakdowns or accidents.\n\nConclusion and Additional Resources:
The manual concludes with a reminder to keep the document in the vehicle for
easy reference and to pass it on to any new owner if the vehicle is sold. The
appendix includes a detailed maintenance schedule and technical specifications
for the vehicle.\n", metadata={'source': './punch.txt'})]

```python
import textwrap

def wrap_text_preserve_newlines(text, width=110):
    lines = text.split('\n')
    wrapped_lines = [textwrap.fill(line, width=width) for line in lines]
    wrapped_text = '\n'.join(wrapped_lines)

    return wrapped_text
```

```python
print(wrap_text_preserve_newlines(str(documents[0])))
```

page_content="Introduction to Ownership: The manual welcomes new owners to the Tata Motors family, emphasizing
the importance of reading the manual to ensure safe and enjoyable driving. Owners are encouraged to perform
prescribed maintenance and repairs at authorized Tata Motors dealers to ensure vehicle reliability and
performance.\n\nCustomer Assistance: Tata Motors offers a toll-free customer assistance line no. 1800 209 8282
and a 24/7 roadside assistance program. This ensures support is readily available for any inquiries or
emergencies, including breakdowns.\n\nSafety Overview:\n- Seat Belts: All occupants should wear seat belts at
all times. The manual explains how to properly fasten the belts and highlights the importance of keeping the
belts untwisted for maximum protection.\n- Child Restraint System (CRS): The use of appropriate child seats is
mandatory for children up to 12 years old, and the manual specifies how to install these seats using the
vehicle's seat belts or ISOFIX anchors.\n- Airbags: Information is provided on the function and location of
airbags. It is critical not to obstruct airbag deployment zones and to always maintain a proper seating
position to maximize airbag effectiveness.\n- Driving Safety Tips: Advice includes driving attentively, using
indicators when changing lanes or turning, and avoiding driving under the influence of alcohol or drugs.
There's also guidance on handling various driving conditions like driving through water, on gradients, and on
wet or slippery roads.\n\nVehicle Operation:\n- Starting the Engine: Instructions are provided for both key
start and push-button start systems, emphasizing the need to ensure the vehicle is in neutral or park before
ignition.\n- Using the Transmission: For manual transmissions, the importance of smooth clutch engagement is
noted, while automatic drivers are advised on selecting drive modes.\n- Climate Control: Operation of manual
and automatic climate systems is explained to maintain comfort and visibility in all conditions.\n-
Infotainment System: Guidelines on safe usage, including setting up navigation routes and station presets when
the vehicle is stationary.\n\nMaintenance and Emergency Handling:\n- Routine Checks: Owners are guided on
checking and maintaining fluid levels, tire pressure, and other critical areas to keep the vehicle in optimal
condition.\n- Emergency Procedures: The manual outlines steps for changing tires, jump-starting the car, and
what to do in the event of an accident.\n- Roadside Assistance: The process to contact and utilize roadside

assistance is clearly laid out for scenarios like vehicle breakdowns or
accidents.\n\nConclusion and
Additional Resources: The manual concludes with a reminder to keep the document
in the vehicle for easy
reference and to pass it on to any new owner if the vehicle is sold. The
appendix includes a detailed
maintenance schedule and technical specifications for the vehicle.\n"
metadata={'source': './punch.txt'}

```python
[11]: # Text Splitter
      from langchain.text_splitter import CharacterTextSplitter
      text_splitter = CharacterTextSplitter(chunk_size=1000, chunk_overlap=0)
      docs = text_splitter.split_documents(documents)
```

```python
[12]: len(docs)
```

[12]: 4

```python
[13]: docs[0]
```

[13]: Document(page_content='Introduction to Ownership: The manual welcomes new owners
      to the Tata Motors family, emphasizing the importance of reading the manual to
      ensure safe and enjoyable driving. Owners are encouraged to perform prescribed
      maintenance and repairs at authorized Tata Motors dealers to ensure vehicle
      reliability and performance.\n\nCustomer Assistance: Tata Motors offers a toll-
      free customer assistance line no. 1800 209 8282 and a 24/7 roadside assistance
      program. This ensures support is readily available for any inquiries or
      emergencies, including breakdowns.', metadata={'source': './punch.txt'})

0.0.3 Embeddings

```python
[14]: # Embeddings
      from langchain.embeddings import HuggingFaceEmbeddings

      embeddings = HuggingFaceEmbeddings()
```

/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:88:
UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab
(https://huggingface.co/settings/tokens), set it as secret in your Google Colab
and restart your session.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access
public models or datasets.
  warnings.warn(

```
[15]: !pip install faiss-cpu
```

Requirement already satisfied: faiss-cpu in /usr/local/lib/python3.10/dist-packages (1.8.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from faiss-cpu) (1.25.2)

```
[16]: # Vectorstore: https://python.langchain.com/en/latest/modules/indexes/
      ↪vectorstores.html
      from langchain.vectorstores import FAISS

      db = FAISS.from_documents(docs, embeddings)
```

```
[17]: query = "How to use the car?"
      docs = db.similarity_search(query)
```

```
[18]: pip install sumy
```

Requirement already satisfied: sumy in /usr/local/lib/python3.10/dist-packages (0.11.0)
Requirement already satisfied: docopt<0.7,>=0.6.1 in /usr/local/lib/python3.10/dist-packages (from sumy) (0.6.2)
Requirement already satisfied: breadability>=0.1.20 in /usr/local/lib/python3.10/dist-packages (from sumy) (0.1.20)
Requirement already satisfied: requests>=2.7.0 in /usr/local/lib/python3.10/dist-packages (from sumy) (2.31.0)
Requirement already satisfied: pycountry>=18.2.23 in /usr/local/lib/python3.10/dist-packages (from sumy) (23.12.11)
Requirement already satisfied: nltk>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from sumy) (3.8.1)
Requirement already satisfied: chardet in /usr/local/lib/python3.10/dist-packages (from breadability>=0.1.20->sumy) (5.2.0)
Requirement already satisfied: lxml>=2.0 in /usr/local/lib/python3.10/dist-packages (from breadability>=0.1.20->sumy) (4.9.4)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk>=3.0.2->sumy) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk>=3.0.2->sumy) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk>=3.0.2->sumy) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk>=3.0.2->sumy) (4.66.4)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.7.0->sumy) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.7.0->sumy) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in

```
         /usr/local/lib/python3.10/dist-packages (from requests>=2.7.0->sumy) (2.0.7)
         Requirement already satisfied: certifi>=2017.4.17 in
         /usr/local/lib/python3.10/dist-packages (from requests>=2.7.0->sumy) (2024.2.2)
```

```python
[19]: import nltk
      nltk.download('punkt')
      from sumy.parsers.plaintext import PlaintextParser
      from sumy.nlp.tokenizers import Tokenizer
      from sumy.summarizers.lsa import LsaSummarizer
      from sumy.summarizers.lex_rank import LexRankSummarizer

      def summarize_text(text, num_sentences=2):
          parser = PlaintextParser.from_string(text, Tokenizer("english"))

          # Using LexRank Summarizer
          summarizer = LexRankSummarizer()
          summary = summarizer(parser.document, num_sentences)

          # Convert the summary sentences to a single string
          summary_text = " ".join([str(sentence) for sentence in summary])

          return summary_text

      summary = summarize_text(docs, num_sentences=25)  # Adjust the number of
       ↪sentences in the summary
      print("Summary:", summary)
```

Summary: [Document(page_content='Vehicle Operation:\n- Starting the Engine:
Instructions are provided for both key start and push-button start systems,
emphasizing the need to ensure the vehicle is in neutral or park before
ignition.\n- Using the Transmission: For manual transmissions, the importance of
smooth clutch engagement is noted, while automatic drivers are advised on
selecting drive modes.\n- Climate Control: Operation of manual and automatic
climate systems is explained to maintain comfort and visibility in all
conditions.\n- Infotainment System: Guidelines on safe usage, including setting
up navigation routes and station presets when the vehicle is stationary. ',
metadata={'source': './punch.txt'}), Document(page_content='Maintenance and
Emergency Handling:\n- Routine Checks: Owners are guided on checking and
maintaining fluid levels, tire pressure, and other critical areas to keep the
vehicle in optimal condition.\n- Emergency Procedures: The manual outlines steps
for changing tires, jump-starting the car, and what to do in the event of an
accident.\n- Roadside Assistance: The process to contact and utilize roadside
assistance is clearly laid out for scenarios like vehicle breakdowns or
accidents.\n\nConclusion and Additional Resources: The manual concludes with a
reminder to keep the document in the vehicle for easy reference and to pass it
on to any new owner if the vehicle is sold. The appendix includes a detailed
maintenance schedule and technical specifications for the vehicle. ',
metadata={'source': './punch.txt'}), Document(page_content="Safety Overview:\n-

Seat Belts: All occupants should wear seat belts at all times. The manual
explains how to properly fasten the belts and highlights the importance of
keeping the belts untwisted for maximum protection.\n- Child Restraint System
(CRS): The use of appropriate child seats is mandatory for children up to 12
years old, and the manual specifies how to install these seats using the
vehicle's seat belts or ISOFIX anchors.\n- Airbags: Information is provided on
the function and location of airbags. It is critical not to obstruct airbag
deployment zones and to always maintain a proper seating position to maximize
airbag effectiveness.\n- Driving Safety Tips: Advice includes driving
attentively, using indicators when changing lanes or turning, and avoiding
driving under the influence of alcohol or drugs. There's also guidance on
handling various driving conditions like driving through water, on gradients,
and on wet or slippery roads. ", metadata={'source': './punch.txt'}),
Document(page_content='Introduction to Ownership: The manual welcomes new owners
to the Tata Motors family, emphasizing the importance of reading the manual to
ensure safe and enjoyable driving. Owners are encouraged to perform prescribed
maintenance and repairs at authorized Tata Motors dealers to ensure vehicle
reliability and performance.\n\nCustomer Assistance: Tata Motors offers a toll-
free customer assistance line no. 1800 209 8282 and a 24/7 roadside assistance
program. This ensures support is readily available for any inquiries or
emergencies, including breakdowns. ', metadata={'source': './punch.txt'})]

[nltk_data] Downloading package punkt to /root/nltk_data…
[nltk_data]   Package punkt is already up-to-date!

[20]: `print(wrap_text_preserve_newlines(str(docs[0].page_content)))`

```
Vehicle Operation:
- Starting the Engine: Instructions are provided for both key start and push-
button start systems, emphasizing
the need to ensure the vehicle is in neutral or park before ignition.
- Using the Transmission: For manual transmissions, the importance of smooth
clutch engagement is noted, while
automatic drivers are advised on selecting drive modes.
- Climate Control: Operation of manual and automatic climate systems is
explained to maintain comfort and
visibility in all conditions.
- Infotainment System: Guidelines on safe usage, including setting up navigation
routes and station presets
when the vehicle is stationary.
```

0.0.4 Create QA Chain

[21]:
```python
from langchain.chains.question_answering import load_qa_chain
from langchain import HuggingFaceHub
```

[23]:
```python
llm=HuggingFaceHub(repo_id="google/flan-t5-xxl", model_kwargs={"temperature":1,
 "max_length":1024})
```

```
[47]: chain = load_qa_chain(llm, chain_type="stuff")
```

```
[49]: query = "How to turn engine on?"
      docs = db.similarity_search(query)
      chain.run(input_documents=docs, question=query)
```

[49]: 'Instructions are provided for both key start and push-button start systems'

```
[50]: query = "Safety tips for driving the car?"
      docs = db.similarity_search(query)
      chain.run(input_documents=docs, question=query)
```

[50]: 'driving attentively, using indicators when changing lanes or turning, and
      avoiding driving under the influence of alcohol or drugs'

```
[28]: query = "Tell me about customer support system of tata motors?"
      docs = db.similarity_search(query)
      chain.run(input_documents=docs, question=query)
```

[28]: 'Tata Motors offers a toll-free customer assistance line no. 1800 209 8282 and a
      24/7 roadside assistance program.'

0.0.5 Working with PDF Files

```
[29]: !pip install unstructured
      !pip install chromadb
      !pip install Cython
      !pip install tiktoken
      !pip install unstructured[local-inference]
```

```
Requirement already satisfied: unstructured in /usr/local/lib/python3.10/dist-
packages (0.14.0)
Requirement already satisfied: chardet in /usr/local/lib/python3.10/dist-
packages (from unstructured) (5.2.0)
Requirement already satisfied: filetype in /usr/local/lib/python3.10/dist-
packages (from unstructured) (1.2.0)
Requirement already satisfied: python-magic in /usr/local/lib/python3.10/dist-
packages (from unstructured) (0.4.27)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages
(from unstructured) (4.9.4)
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages
(from unstructured) (3.8.1)
Requirement already satisfied: tabulate in /usr/local/lib/python3.10/dist-
packages (from unstructured) (0.9.0)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-
packages (from unstructured) (2.31.0)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-
```

packages (from unstructured) (4.12.3)
Requirement already satisfied: emoji in /usr/local/lib/python3.10/dist-packages
(from unstructured) (2.12.1)
Requirement already satisfied: dataclasses-json in
/usr/local/lib/python3.10/dist-packages (from unstructured) (0.6.6)
Requirement already satisfied: python-iso639 in /usr/local/lib/python3.10/dist-
packages (from unstructured) (2024.4.27)
Requirement already satisfied: langdetect in /usr/local/lib/python3.10/dist-
packages (from unstructured) (1.0.9)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages
(from unstructured) (1.25.2)
Requirement already satisfied: rapidfuzz in /usr/local/lib/python3.10/dist-
packages (from unstructured) (3.9.1)
Requirement already satisfied: backoff in /usr/local/lib/python3.10/dist-
packages (from unstructured) (2.2.1)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.10/dist-packages (from unstructured) (4.11.0)
Requirement already satisfied: unstructured-client in
/usr/local/lib/python3.10/dist-packages (from unstructured) (0.22.0)
Requirement already satisfied: wrapt in /usr/local/lib/python3.10/dist-packages
(from unstructured) (1.14.1)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-
packages (from beautifulsoup4->unstructured) (2.5)
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in
/usr/local/lib/python3.10/dist-packages (from dataclasses-json->unstructured)
(3.21.2)
Requirement already satisfied: typing-inspect<1,>=0.4.0 in
/usr/local/lib/python3.10/dist-packages (from dataclasses-json->unstructured)
(0.9.0)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages
(from langdetect->unstructured) (1.16.0)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages
(from nltk->unstructured) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages
(from nltk->unstructured) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in
/usr/local/lib/python3.10/dist-packages (from nltk->unstructured) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages
(from nltk->unstructured) (4.66.4)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->unstructured) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests->unstructured) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->unstructured) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->unstructured) (2024.2.2)
Requirement already satisfied: deepdiff>=6.0 in /usr/local/lib/python3.10/dist-

packages (from unstructured-client->unstructured) (7.0.1)
Requirement already satisfied: jsonpath-python>=1.0.6 in
/usr/local/lib/python3.10/dist-packages (from unstructured-client->unstructured)
(1.0.6)
Requirement already satisfied: mypy-extensions>=1.0.0 in
/usr/local/lib/python3.10/dist-packages (from unstructured-client->unstructured)
(1.0.0)
Requirement already satisfied: packaging>=23.1 in
/usr/local/lib/python3.10/dist-packages (from unstructured-client->unstructured)
(23.2)
Requirement already satisfied: pypdf>=4.0 in /usr/local/lib/python3.10/dist-
packages (from unstructured-client->unstructured) (4.2.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from unstructured-client->unstructured)
(2.8.2)
Requirement already satisfied: ordered-set<4.2.0,>=4.1.0 in
/usr/local/lib/python3.10/dist-packages (from deepdiff>=6.0->unstructured-
client->unstructured) (4.1.0)
Requirement already satisfied: chromadb in /usr/local/lib/python3.10/dist-
packages (0.5.0)
Requirement already satisfied: build>=1.0.3 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (1.2.1)
Requirement already satisfied: requests>=2.28 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (2.31.0)
Requirement already satisfied: pydantic>=1.9 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (2.7.1)
Requirement already satisfied: chroma-hnswlib==0.7.3 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (0.7.3)
Requirement already satisfied: fastapi>=0.95.2 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (0.111.0)
Requirement already satisfied: uvicorn[standard]>=0.18.3 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (0.29.0)
Requirement already satisfied: numpy>=1.22.5 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (1.25.2)
Requirement already satisfied: posthog>=2.4.0 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (3.5.0)
Requirement already satisfied: typing-extensions>=4.5.0 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (4.11.0)
Requirement already satisfied: onnxruntime>=1.14.1 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (1.18.0)
Requirement already satisfied: opentelemetry-api>=1.2.0 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (1.24.0)
Requirement already satisfied: opentelemetry-exporter-otlp-proto-grpc>=1.2.0 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (1.24.0)
Requirement already satisfied: opentelemetry-instrumentation-fastapi>=0.41b0 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (0.45b0)
Requirement already satisfied: opentelemetry-sdk>=1.2.0 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (1.24.0)

Requirement already satisfied: tokenizers>=0.13.2 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (0.19.1)
Requirement already satisfied: pypika>=0.48.9 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (0.48.9)
Requirement already satisfied: tqdm>=4.65.0 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (4.66.4)
Requirement already satisfied: overrides>=7.3.1 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (7.7.0)
Requirement already satisfied: importlib-resources in
/usr/local/lib/python3.10/dist-packages (from chromadb) (6.4.0)
Requirement already satisfied: grpcio>=1.58.0 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (1.63.0)
Requirement already satisfied: bcrypt>=4.0.1 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (4.1.3)
Requirement already satisfied: typer>=0.9.0 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (0.12.3)
Requirement already satisfied: kubernetes>=28.1.0 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (29.0.0)
Requirement already satisfied: tenacity>=8.2.3 in
/usr/local/lib/python3.10/dist-packages (from chromadb) (8.3.0)
Requirement already satisfied: PyYAML>=6.0.0 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (6.0.1)
Requirement already satisfied: mmh3>=4.0.1 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (4.1.0)
Requirement already satisfied: orjson>=3.9.12 in /usr/local/lib/python3.10/dist-
packages (from chromadb) (3.10.3)
Requirement already satisfied: packaging>=19.1 in
/usr/local/lib/python3.10/dist-packages (from build>=1.0.3->chromadb) (23.2)
Requirement already satisfied: pyproject_hooks in
/usr/local/lib/python3.10/dist-packages (from build>=1.0.3->chromadb) (1.1.0)
Requirement already satisfied: tomli>=1.1.0 in /usr/local/lib/python3.10/dist-
packages (from build>=1.0.3->chromadb) (2.0.1)
Requirement already satisfied: starlette<0.38.0,>=0.37.2 in
/usr/local/lib/python3.10/dist-packages (from fastapi>=0.95.2->chromadb)
(0.37.2)
Requirement already satisfied: fastapi-cli>=0.0.2 in
/usr/local/lib/python3.10/dist-packages (from fastapi>=0.95.2->chromadb) (0.0.4)
Requirement already satisfied: httpx>=0.23.0 in /usr/local/lib/python3.10/dist-
packages (from fastapi>=0.95.2->chromadb) (0.27.0)
Requirement already satisfied: jinja2>=2.11.2 in /usr/local/lib/python3.10/dist-
packages (from fastapi>=0.95.2->chromadb) (3.1.4)
Requirement already satisfied: python-multipart>=0.0.7 in
/usr/local/lib/python3.10/dist-packages (from fastapi>=0.95.2->chromadb) (0.0.9)
Requirement already satisfied:
ujson!=4.0.2,!=4.1.0,!=4.2.0,!=4.3.0,!=5.0.0,!=5.1.0,>=4.0.1 in
/usr/local/lib/python3.10/dist-packages (from fastapi>=0.95.2->chromadb)
(5.10.0)
Requirement already satisfied: email_validator>=2.0.0 in

/usr/local/lib/python3.10/dist-packages (from fastapi>=0.95.2->chromadb) (2.1.1)
Requirement already satisfied: certifi>=14.05.14 in
/usr/local/lib/python3.10/dist-packages (from kubernetes>=28.1.0->chromadb)
(2024.2.2)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.10/dist-
packages (from kubernetes>=28.1.0->chromadb) (1.16.0)
Requirement already satisfied: python-dateutil>=2.5.3 in
/usr/local/lib/python3.10/dist-packages (from kubernetes>=28.1.0->chromadb)
(2.8.2)
Requirement already satisfied: google-auth>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from kubernetes>=28.1.0->chromadb)
(2.27.0)
Requirement already satisfied: websocket-
client!=0.40.0,!=0.41.*,!=0.42.*,>=0.32.0 in /usr/local/lib/python3.10/dist-
packages (from kubernetes>=28.1.0->chromadb) (1.8.0)
Requirement already satisfied: requests-oauthlib in
/usr/local/lib/python3.10/dist-packages (from kubernetes>=28.1.0->chromadb)
(1.3.1)
Requirement already satisfied: oauthlib>=3.2.2 in
/usr/local/lib/python3.10/dist-packages (from kubernetes>=28.1.0->chromadb)
(3.2.2)
Requirement already satisfied: urllib3>=1.24.2 in
/usr/local/lib/python3.10/dist-packages (from kubernetes>=28.1.0->chromadb)
(2.0.7)
Requirement already satisfied: coloredlogs in /usr/local/lib/python3.10/dist-
packages (from onnxruntime>=1.14.1->chromadb) (15.0.1)
Requirement already satisfied: flatbuffers in /usr/local/lib/python3.10/dist-
packages (from onnxruntime>=1.14.1->chromadb) (24.3.25)
Requirement already satisfied: protobuf in /usr/local/lib/python3.10/dist-
packages (from onnxruntime>=1.14.1->chromadb) (3.20.3)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages
(from onnxruntime>=1.14.1->chromadb) (1.12)
Requirement already satisfied: deprecated>=1.2.6 in
/usr/local/lib/python3.10/dist-packages (from opentelemetry-
api>=1.2.0->chromadb) (1.2.14)
Requirement already satisfied: importlib-metadata<=7.0,>=6.0 in
/usr/local/lib/python3.10/dist-packages (from opentelemetry-
api>=1.2.0->chromadb) (7.0.0)
Requirement already satisfied: googleapis-common-protos~=1.52 in
/usr/local/lib/python3.10/dist-packages (from opentelemetry-exporter-otlp-proto-
grpc>=1.2.0->chromadb) (1.63.0)
Requirement already satisfied: opentelemetry-exporter-otlp-proto-common==1.24.0
in /usr/local/lib/python3.10/dist-packages (from opentelemetry-exporter-otlp-
proto-grpc>=1.2.0->chromadb) (1.24.0)
Requirement already satisfied: opentelemetry-proto==1.24.0 in
/usr/local/lib/python3.10/dist-packages (from opentelemetry-exporter-otlp-proto-
grpc>=1.2.0->chromadb) (1.24.0)
Requirement already satisfied: opentelemetry-instrumentation-asgi==0.45b0 in

/usr/local/lib/python3.10/dist-packages (from opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (0.45b0)
Requirement already satisfied: opentelemetry-instrumentation==0.45b0 in /usr/local/lib/python3.10/dist-packages (from opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (0.45b0)
Requirement already satisfied: opentelemetry-semantic-conventions==0.45b0 in /usr/local/lib/python3.10/dist-packages (from opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (0.45b0)
Requirement already satisfied: opentelemetry-util-http==0.45b0 in /usr/local/lib/python3.10/dist-packages (from opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (0.45b0)
Requirement already satisfied: setuptools>=16.0 in /usr/local/lib/python3.10/dist-packages (from opentelemetry-instrumentation==0.45b0->opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (67.7.2)
Requirement already satisfied: wrapt<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from opentelemetry-instrumentation==0.45b0->opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (1.14.1)
Requirement already satisfied: asgiref~=3.0 in /usr/local/lib/python3.10/dist-packages (from opentelemetry-instrumentation-asgi==0.45b0->opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (3.8.1)
Requirement already satisfied: monotonic>=1.5 in /usr/local/lib/python3.10/dist-packages (from posthog>=2.4.0->chromadb) (1.6)
Requirement already satisfied: backoff>=1.10.0 in /usr/local/lib/python3.10/dist-packages (from posthog>=2.4.0->chromadb) (2.2.1)
Requirement already satisfied: annotated-types>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic>=1.9->chromadb) (0.6.0)
Requirement already satisfied: pydantic-core==2.18.2 in /usr/local/lib/python3.10/dist-packages (from pydantic>=1.9->chromadb) (2.18.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.28->chromadb) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.28->chromadb) (3.7)
Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in /usr/local/lib/python3.10/dist-packages (from tokenizers>=0.13.2->chromadb) (0.20.3)
Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.10/dist-packages (from typer>=0.9.0->chromadb) (8.1.7)
Requirement already satisfied: shellingham>=1.3.0 in /usr/local/lib/python3.10/dist-packages (from typer>=0.9.0->chromadb) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in /usr/local/lib/python3.10/dist-packages (from typer>=0.9.0->chromadb) (13.7.1)
Requirement already satisfied: h11>=0.8 in /usr/local/lib/python3.10/dist-packages (from uvicorn[standard]>=0.18.3->chromadb) (0.14.0)
Requirement already satisfied: httptools>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from uvicorn[standard]>=0.18.3->chromadb) (0.6.1)

Requirement already satisfied: python-dotenv>=0.13 in
/usr/local/lib/python3.10/dist-packages (from
uvicorn[standard]>=0.18.3->chromadb) (1.0.1)
Requirement already satisfied: uvloop!=0.15.0,!=0.15.1,>=0.14.0 in
/usr/local/lib/python3.10/dist-packages (from
uvicorn[standard]>=0.18.3->chromadb) (0.19.0)
Requirement already satisfied: watchfiles>=0.13 in
/usr/local/lib/python3.10/dist-packages (from
uvicorn[standard]>=0.18.3->chromadb) (0.21.0)
Requirement already satisfied: websockets>=10.4 in
/usr/local/lib/python3.10/dist-packages (from
uvicorn[standard]>=0.18.3->chromadb) (12.0)
Requirement already satisfied: dnspython>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from
email_validator>=2.0.0->fastapi>=0.95.2->chromadb) (2.6.1)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from google-
auth>=1.0.1->kubernetes>=28.1.0->chromadb) (5.3.3)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/usr/local/lib/python3.10/dist-packages (from google-
auth>=1.0.1->kubernetes>=28.1.0->chromadb) (0.4.0)
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.10/dist-
packages (from google-auth>=1.0.1->kubernetes>=28.1.0->chromadb) (4.9)
Requirement already satisfied: anyio in /usr/local/lib/python3.10/dist-packages
(from httpx>=0.23.0->fastapi>=0.95.2->chromadb) (3.7.1)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.10/dist-
packages (from httpx>=0.23.0->fastapi>=0.95.2->chromadb) (1.0.5)
Requirement already satisfied: sniffio in /usr/local/lib/python3.10/dist-
packages (from httpx>=0.23.0->fastapi>=0.95.2->chromadb) (1.3.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub<1.0,>=0.16.4->tokenizers>=0.13.2->chromadb)
(3.14.0)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.10/dist-packages (from huggingface-
hub<1.0,>=0.16.4->tokenizers>=0.13.2->chromadb) (2023.6.0)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.10/dist-
packages (from importlib-metadata<=7.0,>=6.0->opentelemetry-
api>=1.2.0->chromadb) (3.18.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from
jinja2>=2.11.2->fastapi>=0.95.2->chromadb) (2.1.5)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.10/dist-packages (from
rich>=10.11.0->typer>=0.9.0->chromadb) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.10/dist-packages (from
rich>=10.11.0->typer>=0.9.0->chromadb) (2.16.1)
Requirement already satisfied: humanfriendly>=9.1 in

/usr/local/lib/python3.10/dist-packages (from
coloredlogs->onnxruntime>=1.14.1->chromadb) (10.0)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-
packages (from sympy->onnxruntime>=1.14.1->chromadb) (1.3.0)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-
packages (from anyio->httpx>=0.23.0->fastapi>=0.95.2->chromadb) (1.2.1)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/dist-
packages (from markdown-it-py>=2.2.0->rich>=10.11.0->typer>=0.9.0->chromadb)
(0.1.2)
Requirement already satisfied: pyasn1<0.7.0,>=0.4.6 in
/usr/local/lib/python3.10/dist-packages (from pyasn1-modules>=0.2.1->google-
auth>=1.0.1->kubernetes>=28.1.0->chromadb) (0.6.0)
Requirement already satisfied: Cython in /usr/local/lib/python3.10/dist-packages
(3.0.10)
Requirement already satisfied: tiktoken in /usr/local/lib/python3.10/dist-
packages (0.7.0)
Requirement already satisfied: regex>=2022.1.18 in
/usr/local/lib/python3.10/dist-packages (from tiktoken) (2023.12.25)
Requirement already satisfied: requests>=2.26.0 in
/usr/local/lib/python3.10/dist-packages (from tiktoken) (2.31.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken)
(3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests>=2.26.0->tiktoken) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken)
(2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken)
(2024.2.2)
Requirement already satisfied: unstructured[local-inference] in
/usr/local/lib/python3.10/dist-packages (0.14.0)
Requirement already satisfied: chardet in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (5.2.0)
Requirement already satisfied: filetype in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (1.2.0)
Requirement already satisfied: python-magic in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (0.4.27)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages
(from unstructured[local-inference]) (4.9.4)
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages
(from unstructured[local-inference]) (3.8.1)
Requirement already satisfied: tabulate in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (0.9.0)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (2.31.0)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-

packages (from unstructured[local-inference]) (4.12.3)
Requirement already satisfied: emoji in /usr/local/lib/python3.10/dist-packages
(from unstructured[local-inference]) (2.12.1)
Requirement already satisfied: dataclasses-json in
/usr/local/lib/python3.10/dist-packages (from unstructured[local-inference])
(0.6.6)
Requirement already satisfied: python-iso639 in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (2024.4.27)
Requirement already satisfied: langdetect in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (1.0.9)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages
(from unstructured[local-inference]) (1.25.2)
Requirement already satisfied: rapidfuzz in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (3.9.1)
Requirement already satisfied: backoff in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (2.2.1)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.10/dist-packages (from unstructured[local-inference])
(4.11.0)
Requirement already satisfied: unstructured-client in
/usr/local/lib/python3.10/dist-packages (from unstructured[local-inference])
(0.22.0)
Requirement already satisfied: wrapt in /usr/local/lib/python3.10/dist-packages
(from unstructured[local-inference]) (1.14.1)
Requirement already satisfied: pikepdf in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (8.15.1)
Requirement already satisfied: pdfminer.six in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (20231228)
Requirement already satisfied: python-pptx<=0.6.23 in
/usr/local/lib/python3.10/dist-packages (from unstructured[local-inference])
(0.6.23)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (3.3)
Requirement already satisfied: pdf2image in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (1.17.0)
Requirement already satisfied: pypandoc in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (1.13)
Requirement already satisfied: onnx in /usr/local/lib/python3.10/dist-packages
(from unstructured[local-inference]) (1.16.0)
Requirement already satisfied: unstructured.pytesseract>=0.3.12 in
/usr/local/lib/python3.10/dist-packages (from unstructured[local-inference])
(0.3.12)
Requirement already satisfied: google-cloud-vision in
/usr/local/lib/python3.10/dist-packages (from unstructured[local-inference])
(3.7.2)
Requirement already satisfied: python-docx in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (1.1.2)
Requirement already satisfied: pypdf in /usr/local/lib/python3.10/dist-packages

(from unstructured[local-inference]) (4.2.0)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages
(from unstructured[local-inference]) (2.0.3)
Requirement already satisfied: pillow-heif in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (0.16.0)
Requirement already satisfied: markdown in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (3.6)
Requirement already satisfied: msg-parser in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (1.2.0)
Requirement already satisfied: unstructured-inference==0.7.31 in
/usr/local/lib/python3.10/dist-packages (from unstructured[local-inference])
(0.7.31)
Requirement already satisfied: openpyxl in /usr/local/lib/python3.10/dist-
packages (from unstructured[local-inference]) (3.1.2)
Requirement already satisfied: xlrd in /usr/local/lib/python3.10/dist-packages
(from unstructured[local-inference]) (2.0.1)
Requirement already satisfied: layoutparser[layoutmodels,tesseract] in
/usr/local/lib/python3.10/dist-packages (from unstructured-
inference==0.7.31->unstructured[local-inference]) (0.3.4)
Requirement already satisfied: python-multipart in
/usr/local/lib/python3.10/dist-packages (from unstructured-
inference==0.7.31->unstructured[local-inference]) (0.0.9)
Requirement already satisfied: huggingface-hub in
/usr/local/lib/python3.10/dist-packages (from unstructured-
inference==0.7.31->unstructured[local-inference]) (0.20.3)
Requirement already satisfied: opencv-python!=4.7.0.68 in
/usr/local/lib/python3.10/dist-packages (from unstructured-
inference==0.7.31->unstructured[local-inference]) (4.8.0.76)
Requirement already satisfied: onnxruntime>=1.17.0 in
/usr/local/lib/python3.10/dist-packages (from unstructured-
inference==0.7.31->unstructured[local-inference]) (1.18.0)
Requirement already satisfied: transformers>=4.25.1 in
/usr/local/lib/python3.10/dist-packages (from unstructured-
inference==0.7.31->unstructured[local-inference]) (4.40.2)
Requirement already satisfied: Pillow>=3.3.2 in /usr/local/lib/python3.10/dist-
packages (from python-pptx<=0.6.23->unstructured[local-inference]) (10.3.0)
Requirement already satisfied: XlsxWriter>=0.5.7 in
/usr/local/lib/python3.10/dist-packages (from python-
pptx<=0.6.23->unstructured[local-inference]) (3.2.0)
Requirement already satisfied: packaging>=21.3 in
/usr/local/lib/python3.10/dist-packages (from
unstructured.pytesseract>=0.3.12->unstructured[local-inference]) (23.2)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-
packages (from beautifulsoup4->unstructured[local-inference]) (2.5)
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in
/usr/local/lib/python3.10/dist-packages (from dataclasses-
json->unstructured[local-inference]) (3.21.2)
Requirement already satisfied: typing-inspect<1,>=0.4.0 in

/usr/local/lib/python3.10/dist-packages (from dataclasses-
json->unstructured[local-inference]) (0.9.0)
Requirement already satisfied: google-api-core[grpc]!=2.0.*,!=2.1.*,!=2.10.*,!=2
.2.*,!=2.3.*,!=2.4.*,!=2.5.*,!=2.6.*,!=2.7.*,!=2.8.*,!=2.9.*,<3.0.0dev,>=1.34.1
in /usr/local/lib/python3.10/dist-packages (from google-cloud-
vision->unstructured[local-inference]) (2.11.1)
Requirement already satisfied: google-auth!=2.24.0,!=2.25.0,<3.0.0dev,>=2.14.1
in /usr/local/lib/python3.10/dist-packages (from google-cloud-
vision->unstructured[local-inference]) (2.27.0)
Requirement already satisfied: proto-plus<2.0.0dev,>=1.22.3 in
/usr/local/lib/python3.10/dist-packages (from google-cloud-
vision->unstructured[local-inference]) (1.23.0)
Requirement already satisfied: protobuf!=3.20.0,!=3.20.1,!=4.21.0,!=4.21.1,!=4.2
1.2,!=4.21.3,!=4.21.4,!=4.21.5,<5.0.0dev,>=3.19.5 in
/usr/local/lib/python3.10/dist-packages (from google-cloud-
vision->unstructured[local-inference]) (3.20.3)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages
(from langdetect->unstructured[local-inference]) (1.16.0)
Requirement already satisfied: olefile>=0.46 in /usr/local/lib/python3.10/dist-
packages (from msg-parser->unstructured[local-inference]) (0.47)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages
(from nltk->unstructured[local-inference]) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages
(from nltk->unstructured[local-inference]) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in
/usr/local/lib/python3.10/dist-packages (from nltk->unstructured[local-
inference]) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages
(from nltk->unstructured[local-inference]) (4.66.4)
Requirement already satisfied: et-xmlfile in /usr/local/lib/python3.10/dist-
packages (from openpyxl->unstructured[local-inference]) (1.1.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas->unstructured[local-
inference]) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas->unstructured[local-inference]) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-
packages (from pandas->unstructured[local-inference]) (2024.1)
Requirement already satisfied: charset-normalizer>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from pdfminer.six->unstructured[local-
inference]) (3.3.2)
Requirement already satisfied: cryptography>=36.0.0 in
/usr/local/lib/python3.10/dist-packages (from pdfminer.six->unstructured[local-
inference]) (42.0.7)
Requirement already satisfied: Deprecated in /usr/local/lib/python3.10/dist-
packages (from pikepdf->unstructured[local-inference]) (1.2.14)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests->unstructured[local-inference]) (3.7)

24

Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->unstructured[local-
inference]) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->unstructured[local-
inference]) (2024.2.2)
Requirement already satisfied: deepdiff>=6.0 in /usr/local/lib/python3.10/dist-
packages (from unstructured-client->unstructured[local-inference]) (7.0.1)
Requirement already satisfied: jsonpath-python>=1.0.6 in
/usr/local/lib/python3.10/dist-packages (from unstructured-
client->unstructured[local-inference]) (1.0.6)
Requirement already satisfied: mypy-extensions>=1.0.0 in
/usr/local/lib/python3.10/dist-packages (from unstructured-
client->unstructured[local-inference]) (1.0.0)
Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.10/dist-
packages (from cryptography>=36.0.0->pdfminer.six->unstructured[local-
inference]) (1.16.0)
Requirement already satisfied: ordered-set<4.2.0,>=4.1.0 in
/usr/local/lib/python3.10/dist-packages (from deepdiff>=6.0->unstructured-
client->unstructured[local-inference]) (4.1.0)
Requirement already satisfied: googleapis-common-protos<2.0.dev0,>=1.56.2 in
/usr/local/lib/python3.10/dist-packages (from google-api-core[grpc]!=2.0.*,!=2.1
.*,!=2.10.*,!=2.2.*,!=2.3.*,!=2.4.*,!=2.5.*,!=2.6.*,!=2.7.*,!=2.8.*,!=2.9.*,<3.0
.0dev,>=1.34.1->google-cloud-vision->unstructured[local-inference]) (1.63.0)
Requirement already satisfied: grpcio<2.0dev,>=1.33.2 in
/usr/local/lib/python3.10/dist-packages (from google-api-core[grpc]!=2.0.*,!=2.1
.*,!=2.10.*,!=2.2.*,!=2.3.*,!=2.4.*,!=2.5.*,!=2.6.*,!=2.7.*,!=2.8.*,!=2.9.*,<3.0
.0dev,>=1.34.1->google-cloud-vision->unstructured[local-inference]) (1.63.0)
Requirement already satisfied: grpcio-status<2.0.dev0,>=1.33.2 in
/usr/local/lib/python3.10/dist-packages (from google-api-core[grpc]!=2.0.*,!=2.1
.*,!=2.10.*,!=2.2.*,!=2.3.*,!=2.4.*,!=2.5.*,!=2.6.*,!=2.7.*,!=2.8.*,!=2.9.*,<3.0
.0dev,>=1.34.1->google-cloud-vision->unstructured[local-inference]) (1.48.2)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from google-
auth!=2.24.0,!=2.25.0,<3.0.0dev,>=2.14.1->google-cloud-
vision->unstructured[local-inference]) (5.3.3)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/usr/local/lib/python3.10/dist-packages (from google-
auth!=2.24.0,!=2.25.0,<3.0.0dev,>=2.14.1->google-cloud-
vision->unstructured[local-inference]) (0.4.0)
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.10/dist-
packages (from google-auth!=2.24.0,!=2.25.0,<3.0.0dev,>=2.14.1->google-cloud-
vision->unstructured[local-inference]) (4.9)
Requirement already satisfied: coloredlogs in /usr/local/lib/python3.10/dist-
packages (from onnxruntime>=1.17.0->unstructured-
inference==0.7.31->unstructured[local-inference]) (15.0.1)
Requirement already satisfied: flatbuffers in /usr/local/lib/python3.10/dist-
packages (from onnxruntime>=1.17.0->unstructured-

inference==0.7.31->unstructured[local-inference]) (24.3.25)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages
(from onnxruntime>=1.17.0->unstructured-inference==0.7.31->unstructured[local-
inference]) (1.12)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
packages (from transformers>=4.25.1->unstructured-
inference==0.7.31->unstructured[local-inference]) (3.14.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-
packages (from transformers>=4.25.1->unstructured-
inference==0.7.31->unstructured[local-inference]) (6.0.1)
Requirement already satisfied: tokenizers<0.20,>=0.19 in
/usr/local/lib/python3.10/dist-packages (from
transformers>=4.25.1->unstructured-inference==0.7.31->unstructured[local-
inference]) (0.19.1)
Requirement already satisfied: safetensors>=0.4.1 in
/usr/local/lib/python3.10/dist-packages (from
transformers>=4.25.1->unstructured-inference==0.7.31->unstructured[local-
inference]) (0.4.3)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.10/dist-packages (from huggingface-hub->unstructured-
inference==0.7.31->unstructured[local-inference]) (2023.6.0)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages
(from layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (1.11.4)
Requirement already satisfied: iopath in /usr/local/lib/python3.10/dist-packages
(from layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (0.1.10)
Requirement already satisfied: pdfplumber in /usr/local/lib/python3.10/dist-
packages (from layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (0.11.0)
Requirement already satisfied: pytesseract in /usr/local/lib/python3.10/dist-
packages (from layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (0.3.10)
Requirement already satisfied: torch in /usr/local/lib/python3.10/dist-packages
(from layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (2.2.1+cu121)
Requirement already satisfied: torchvision in /usr/local/lib/python3.10/dist-
packages (from layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (0.17.1+cu121)
Requirement already satisfied: effdet in /usr/local/lib/python3.10/dist-packages
(from layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (0.4.1)
Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-
packages (from
cffi>=1.12->cryptography>=36.0.0->pdfminer.six->unstructured[local-inference])
(2.22)
Requirement already satisfied: pyasn1<0.7.0,>=0.4.6 in
/usr/local/lib/python3.10/dist-packages (from pyasn1-modules>=0.2.1->google-

auth!=2.24.0,!=2.25.0,<3.0.0dev,>=2.14.1->google-cloud-
vision->unstructured[local-inference]) (0.6.0)
Requirement already satisfied: humanfriendly>=9.1 in
/usr/local/lib/python3.10/dist-packages (from
coloredlogs->onnxruntime>=1.17.0->unstructured-
inference==0.7.31->unstructured[local-inference]) (10.0)
Requirement already satisfied: timm>=0.9.2 in /usr/local/lib/python3.10/dist-
packages (from effdet->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (1.0.3)
Requirement already satisfied: pycocotools>=2.0.2 in
/usr/local/lib/python3.10/dist-packages (from
effdet->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (2.0.7)
Requirement already satisfied: omegaconf>=2.0 in /usr/local/lib/python3.10/dist-
packages (from effdet->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (2.3.0)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages
(from torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (3.1.4)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (12.1.105)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (12.1.105)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (12.1.105)
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in
/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (8.9.2.26)
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in
/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (12.1.3.1)
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in
/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (11.0.2.54)
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in
/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (10.3.2.106)
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in

/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (11.4.5.107)
Requirement already satisfied: nvidia-cusparse-cu12==12.1.0.106 in
/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (12.1.0.106)
Requirement already satisfied: nvidia-nccl-cu12==2.19.3 in
/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (2.19.3)
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from
torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (12.1.105)
Requirement already satisfied: triton==2.2.0 in /usr/local/lib/python3.10/dist-
packages (from torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (2.2.0)
Requirement already satisfied: nvidia-nvjitlink-cu12 in
/usr/local/lib/python3.10/dist-packages (from nvidia-cusolver-
cu12==11.4.5.107->torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (12.4.127)
Requirement already satisfied: portalocker in /usr/local/lib/python3.10/dist-
packages (from iopath->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (2.8.2)
Requirement already satisfied: pypdfium2>=4.18.0 in
/usr/local/lib/python3.10/dist-packages (from
pdfplumber->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (4.30.0)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-
packages (from sympy->onnxruntime>=1.17.0->unstructured-
inference==0.7.31->unstructured[local-inference]) (1.3.0)
Requirement already satisfied: antlr4-python3-runtime==4.9.* in
/usr/local/lib/python3.10/dist-packages (from
omegaconf>=2.0->effdet->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (4.9.3)
Requirement already satisfied: matplotlib>=2.1.0 in
/usr/local/lib/python3.10/dist-packages (from
pycocotools>=2.0.2->effdet->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (3.7.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from
jinja2->torch->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (2.1.5)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.0->pycocotools>=2.
0.2->effdet->layoutparser[layoutmodels,tesseract]->unstructured-
inference==0.7.31->unstructured[local-inference]) (1.2.1)

Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.0->pycocotools>=2.0.2->effdet->layoutparser[layoutmodels,tesseract]->unstructured-inference==0.7.31->unstructured[local-inference]) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.0->pycocotools>=2.0.2->effdet->layoutparser[layoutmodels,tesseract]->unstructured-inference==0.7.31->unstructured[local-inference]) (4.51.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.0->pycocotools>=2.0.2->effdet->layoutparser[layoutmodels,tesseract]->unstructured-inference==0.7.31->unstructured[local-inference]) (1.4.5)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.0->pycocotools>=2.0.2->effdet->layoutparser[layoutmodels,tesseract]->unstructured-inference==0.7.31->unstructured[local-inference]) (3.1.2)

```python
[30]: from langchain.document_loaders import UnstructuredPDFLoader
      from langchain.indexes import VectorstoreIndexCreator
```

```python
[31]: # connect your Google Drive
      import os
      from google.colab import drive
      drive.mount('/content/gdrive', force_remount=True)

      pdf_folder_path = '/content/gdrive/My Drive/bosch/'
      os.listdir(pdf_folder_path)
```

Mounted at /content/gdrive

```
[31]: ['punch.pdf']
```

```python
[32]: loaders = [UnstructuredPDFLoader(os.path.join(pdf_folder_path, fn)) for fn in
       ↪os.listdir(pdf_folder_path)]
      loaders
```

```
[32]: [<langchain_community.document_loaders.pdf.UnstructuredPDFLoader at
       0x7a624a20b070>]
```

```python
[33]: from langchain.text_splitter import CharacterTextSplitter
      from langchain.embeddings import HuggingFaceEmbeddings
      index = VectorstoreIndexCreator(
          embedding=HuggingFaceEmbeddings(),
          text_splitter=CharacterTextSplitter(chunk_size=1000, chunk_overlap=0)).
       ↪from_loaders(loaders)
```

/usr/local/lib/python3.10/dist-packages/langchain/indexes/vectorstore.py:129:
UserWarning: Using InMemoryVectorStore as the default vectorstore.This memory

29

store won't persist data. You should explicitlyspecify a vectorstore when using
VectorstoreIndexCreator
  warnings.warn(
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   /root/nltk_data…
[nltk_data]   Unzipping taggers/averaged_perceptron_tagger.zip.
WARNING:langchain_text_splitters.base:Created a chunk of size 1187, which is
longer than the specified 1000

```python
[34]: from langchain.chains.question_answering import load_qa_chain
      from langchain import HuggingFaceHub
      import os
      os.environ["HUGGINGFACEHUB_API_TOKEN"] = "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"
```

```python
[35]: llm=HuggingFaceHub(repo_id="google/flan-t5-xxl", model_kwargs={"temperature":1,
      ↪"max_length":1024})
```

```python
[52]: from langchain.chains import RetrievalQA
      chain = RetrievalQA.from_chain_type(llm=llm,
                                          chain_type="stuff",
                                          retriever=index.vectorstore.as_retriever(),
                                          input_key="question")
```

```python
[54]: chain.run('When shoudnt airbag be deployed?')
```

```
[54]: 'The amount of visible vehicle damage is not always the correct indicator for
      airbag deployment. Some collisions can result in visible damage but no airbag
      SAFETY deployment because the airbags would not have been needed or would not
      have provided protection even if they had de- ployed. Seat belts, if worn, offer
      adequate occu- pant protection in such cases.'
```

```python
[56]: chain.run('How should we lock the door of cars?')
```

```
[56]: 'The front doors can be locked or un- locked from outside using the key blade.
      Insert the key in the slot and turn it clock- wise to lock and anticlockwise to
      unlock the door.'
```

```python
[57]: chain.run('How is abs deployed?')
```

```
[57]: 'ABS intervenes when braking, you will feel a pulsing in the brake pedal.'
```

1 GUI INTERFACE FOR QUERYING THE DOCUMENT

```
[58]: !pip install streamlit
      !pip install pyngrok
```

Requirement already satisfied: streamlit in /usr/local/lib/python3.10/dist-
packages (1.34.0)
Requirement already satisfied: altair<6,>=4.0 in /usr/local/lib/python3.10/dist-
packages (from streamlit) (4.2.2)
Requirement already satisfied: blinker<2,>=1.0.0 in /usr/lib/python3/dist-
packages (from streamlit) (1.4)
Requirement already satisfied: cachetools<6,>=4.0 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (5.3.3)
Requirement already satisfied: click<9,>=7.0 in /usr/local/lib/python3.10/dist-
packages (from streamlit) (8.1.7)
Requirement already satisfied: numpy<2,>=1.19.3 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (1.25.2)
Requirement already satisfied: packaging<25,>=16.8 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (23.2)
Requirement already satisfied: pandas<3,>=1.3.0 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (2.0.3)
Requirement already satisfied: pillow<11,>=7.1.0 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (10.3.0)
Requirement already satisfied: protobuf<5,>=3.20 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (3.20.3)
Requirement already satisfied: pyarrow>=7.0 in /usr/local/lib/python3.10/dist-
packages (from streamlit) (14.0.2)
Requirement already satisfied: requests<3,>=2.27 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (2.31.0)
Requirement already satisfied: rich<14,>=10.14.0 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (13.7.1)
Requirement already satisfied: tenacity<9,>=8.1.0 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (8.3.0)
Requirement already satisfied: toml<2,>=0.10.1 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (0.10.2)
Requirement already satisfied: typing-extensions<5,>=4.3.0 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (4.11.0)
Requirement already satisfied: gitpython!=3.1.19,<4,>=3.0.7 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (3.1.43)
Requirement already satisfied: pydeck<1,>=0.8.0b4 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (0.9.1)
Requirement already satisfied: tornado<7,>=6.0.3 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (6.3.3)
Requirement already satisfied: watchdog>=2.1.5 in
/usr/local/lib/python3.10/dist-packages (from streamlit) (4.0.0)
Requirement already satisfied: entrypoints in /usr/local/lib/python3.10/dist-
packages (from altair<6,>=4.0->streamlit) (0.4)

Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages
(from altair<6,>=4.0->streamlit) (3.1.4)
Requirement already satisfied: jsonschema>=3.0 in
/usr/local/lib/python3.10/dist-packages (from altair<6,>=4.0->streamlit)
(4.19.2)
Requirement already satisfied: toolz in /usr/local/lib/python3.10/dist-packages
(from altair<6,>=4.0->streamlit) (0.12.1)
Requirement already satisfied: gitdb<5,>=4.0.1 in
/usr/local/lib/python3.10/dist-packages (from
gitpython!=3.1.19,<4,>=3.0.7->streamlit) (4.0.11)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas<3,>=1.3.0->streamlit)
(2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas<3,>=1.3.0->streamlit) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-
packages (from pandas<3,>=1.3.0->streamlit) (2024.1)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.27->streamlit)
(3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests<3,>=2.27->streamlit) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.27->streamlit)
(2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.27->streamlit)
(2024.2.2)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.10/dist-packages (from rich<14,>=10.14.0->streamlit)
(3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.10/dist-packages (from rich<14,>=10.14.0->streamlit)
(2.16.1)
Requirement already satisfied: smmap<6,>=3.0.1 in
/usr/local/lib/python3.10/dist-packages (from
gitdb<5,>=4.0.1->gitpython!=3.1.19,<4,>=3.0.7->streamlit) (5.0.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from jinja2->altair<6,>=4.0->streamlit)
(2.1.5)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.10/dist-
packages (from jsonschema>=3.0->altair<6,>=4.0->streamlit) (23.2.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
/usr/local/lib/python3.10/dist-packages (from
jsonschema>=3.0->altair<6,>=4.0->streamlit) (2023.12.1)
Requirement already satisfied: referencing>=0.28.4 in
/usr/local/lib/python3.10/dist-packages (from
jsonschema>=3.0->altair<6,>=4.0->streamlit) (0.35.1)

Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=3.0->altair<6,>=4.0->streamlit) (0.18.1)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/dist-packages (from markdown-it-py>=2.2.0->rich<14,>=10.14.0->streamlit) (0.1.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas<3,>=1.3.0->streamlit) (1.16.0)
Requirement already satisfied: pyngrok in /usr/local/lib/python3.10/dist-packages (7.1.6)
Requirement already satisfied: PyYAML>=5.1 in /usr/local/lib/python3.10/dist-packages (from pyngrok) (6.0.1)

[59]:
```python
%%writefile streamlit_app.py
import streamlit as st
import os
from langchain_community.document_loaders import TextLoader
from langchain.text_splitter import CharacterTextSplitter
from langchain_community.embeddings import HuggingFaceEmbeddings
from langchain.vectorstores import FAISS
from langchain.chains.question_answering import load_qa_chain
from langchain_community.llms import HuggingFaceEndpoint
import textwrap
import nltk
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.lex_rank import LexRankSummarizer


# Setup environment variable for HuggingFace
os.environ["HUGGINGFACEHUB_API_TOKEN"] = "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"

# Load documents
@st.cache_resource
def load_documents():
    loader = TextLoader('./punch.txt')
    documents = loader.load()
    text_splitter = CharacterTextSplitter(chunk_size=1000, chunk_overlap=0)
    docs = text_splitter.split_documents(documents)
    return docs

# Setup embeddings and vector store
@st.cache_resource
def setup_vector_store(_docs):
    embeddings = HuggingFaceEmbeddings()
    db = FAISS.from_documents(_docs, embeddings)
    return db

# Summarize text
def summarize_text(text, num_sentences=2):
```

```python
    nltk.download('punkt', quiet=True)
    parser = PlaintextParser.from_string(text, Tokenizer("english"))
    summarizer = LexRankSummarizer()
    summary = summarizer(parser.document, num_sentences)
    summary_text = " ".join([str(sentence) for sentence in summary])
    return summary_text


# Load Q&A chain
@st.cache_resource
def load_qa_chain_model():
    llm = HuggingFaceEndpoint(repo_id="google/flan-t5-xxl", temperature=1,␣
 ↪model_kwargs={"max_length": 1024})
    chain = load_qa_chain(llm, chain_type="stuff")
    return chain


# Wrap text to preserve newlines for better readability
def wrap_text_preserve_newlines(text, width=110):
    lines = text.split('\n')
    wrapped_lines = [textwrap.fill(line, width=width) for line in lines]
    wrapped_text = '\n'.join(wrapped_lines)
    return wrapped_text


# Streamlit UI
def main():
    st.title("Document-based Q&A System")

    docs = load_documents()
    db = setup_vector_store(docs)
    qa_chain = load_qa_chain_model()

    query = st.text_input("Enter your question here:", value="How to turn␣
 ↪engine on?")
    if query:
        result_docs = db.similarity_search(query)
        if result_docs:
            # Remove all newline characters before summarization
            flat_texts = [" ".join(doc.page_content.replace('\n', ' ').split())␣
 ↪for doc in result_docs]
            summary = summarize_text(" ".join(flat_texts), num_sentences=5)
            st.text_area("Summary of Relevant Documents:", summary, height=250)
            answer = qa_chain.run(input_documents=result_docs, question=query)
            st.text_area("Answer:", answer, height=150)
        else:
            st.write("No relevant documents found.")


if __name__ == "__main__":
    main()
```

Overwriting streamlit_app.py

```python
from pyngrok import ngrok

# Set up your authtoken from ngrok (you should have already done this step
↪previously)
ngrok.set_auth_token('2FGGrbhYvz2ypBMOImOEAFVYyW7_4MC8JC3jMw9uiYeHvW57p')

# Setup a tunnel to the streamlit port 8501
public_url = ngrok.connect(8501)
print('Public URL:', public_url)
!streamlit run streamlit_app.py --server.port 8501
```