Documentation for Chroma

This code will run only on Python3.11

This code is not compatable with latest version of python 3.14

Steps to run the Code

- pip install -r requirements.txt
- python3 rag_chroma.py

Libraries Required

```
from langchain_community.document_loaders import PyPDFLoader
from langchain_text_splitters import RecursiveCharacterTextSplitter
from langchain_huggingface import HuggingFaceEmbeddings
import chromadb
from chromadb.config import Settings
import re
import pprint
```

The Above Libraries were used in the code

- **langchain_community.document_loaders import PyPDFLoader** helps us toparse the PDF documents required for the context
- **langchain_text_splitters import RecursiveCharacterTextSplitter**. This will Help us to create Chunks from the parsed PDF document data
- **langchain_huggingface import HuggingFaceEmbeddings** helps us to use the **sentence-transformers/all-MiniLM-L6-v2** an Industry Standard to create the embedings
- **import chromadb** This helps to inport the chromaDB library to create a local vectorDB
- **chromadb.config import Settings** this will help us to manage the settings of the chromadb like the location of local DB

Loading the PDF

```
filepath="./example.pdf"

loader=PyPDFLoader(filepath)
print(loader)

docs=loader.load()
```

The Above code Demonstrates The loading of the PDF document using the Langchain Library PyPDFLoader

Chunking the Document

```
text_splitter=RecursiveCharacterTextSplitter(chunk_size=500,chunk_overlap=100)

chunks=text_splitter.split_documents(docs)

for idx,chunk in enumerate(chunks):
  chunk.metadata["chunk_id"]=idx
```

From the Above code we use the RecursivecharacterTextSplitter for the chunking of the documents and each chunk has 500 characters as the information was largely in paragraph format and overlap of 100 to preserve the edge information

Embeding of the Chunks

```
embedings=HuggingFaceEmbeddings(model_name="sentence-transformers/all-MiniLM-L6-v2")

texts=[chunk.page_content for chunk in chunks]
metadata=[chunk.metadata for chunk in chunks]
ids=[str(chunk.metadata["chunk_id"]) for chunk in chunks]

vectors=embedings.embed_documents(texts)
```

The Above code Snippets shows that we use the **sentence-transformers/all-MiniLM-L6-v2** transformer to create the embeding

we Destructure the chunk data into the texts, metadatas and ids for the better management of the data

and we generate the vectors using embeding.embed_documents(texts) function

Chroma DB

```
client=chromadb.Client(Settings(persist_directory="./vectordb/chroma"))
collection=client.create_collection(name="my_collection")

vectors=embedings.embed_documents(texts)

collection.upsert(documents=texts,embeddings=vectors,metadatas=metadata,ids=ids)
```

- In the Above Code Snippet we create the client for the chromaDB which will act as an instance for chromadb
- We create the Collection where the vectors will be stored
- Now we will insert all the data like texts,vectors and metadatas **collection.upsert(documents=texts,embeddings=**

Query

```
query="How does top management demonstrate leadership and commitment to the ISMS?"

query_vector=embedings.embed_query(query)

results=collection.query(
  query_embeddings=[query_vector],
  n_results=5
)
```

The above code snippet will represents the embeding of the query using embed_query and then use collection.query to search the database

```
Numebr of Inserted were : 152
Enter the Query :How does top management demonstrate leadership and commitment to the ISMS?
Enter the TOP K value :5
1


---------------------------------------
Result : 1



 chunk ID :51


Chunk Text : management system, including the processes needed and their interactions, in accordance with the
requirements of this document.
5  Leadership
5.1  Leadership and commitment
Top management shall demonstrate leadership and commitment with respect to the information
security management system by:
a) ensuring the information security policy and the information security objectives are established
and are compatible with the strategic direction of the organization;


Page : 7


Source : ./example.pdf


Score : 0.9229822754859924


---------------------------------------
Result : 2



 chunk ID :57


Chunk Text : security are assigned and communicated within the organization.
Top management shall assign the responsibility and authority for:
a) ensuring that the information security management system conforms to the requirements of this
document;
b) reporting on the performance of the information security management system to top management.
NOTE Top management can also assign responsibilities and authorities for reporting performance of the


Page : 8


Source : ./example.pdf


Score : 0.986425518989563
```

Figure 1: Chroma Output

Outputs

Sample Outputs

Query 1: How does top management demonstrate leadership and commitment to the ISMS?

TOP K=5

```
---------------------------------------
Result : 3


 chunk ID :38


Chunk Text : management system implementation will be scaled in accordance with the needs of the organization.
This document can be used by internal and external parties to assess the organization's ability to meet
the organization’s own information security requirements.
The order in which requirements are presented in this document does not reflect their importance
or imply the order in which they are to be implemented. The list items are enumerated for reference
purpose only.


Page : 4


Source : ./example.pdf


Score : 1.204768419265747


---------------------------------------
Result : 4


 chunk ID :53


Chunk Text : to the information security management system requirements;
e) ensuring that the information security management system achieves its intended outcome(s);
f) directing and supporting persons to contribute to the effectiveness of the information security
management system;
g) promoting continual improvement; and
h) supporting other relevant management roles to demonstrate their leadership as it applies to their
areas of responsibility.


Page : 7


Source : ./example.pdf


Score : 1.2773454189300537


---------------------------------------
Result : 5


 chunk ID :36


Chunk Text : organization’s needs and objectives, security requirements, the organizational processes used and the
size and structure of the organization. All of these influencing factors are expected to change over time.
The information security management system preserves the confidentiality, integrity and availability
of information by applying a risk management process and gives confidence to interested parties that
risks are adequately managed.


Page : 4


Source : ./example.pdf


Score : 1.329432487487793
```

Query 2: What are the requirements for establishing and communicating the information security policy?

TOP K=5

```
Enter the Query :What are the requirements for establishing and communicating the information security policy?
Enter the TOP K value :5
1


---------------------------------------
Result : 1



 chunk ID :69


Chunk Text : a) be consistent with the information security policy;
b) be measurable (if practicable);
c) take into account applicable information security requirements, and results from risk assessment
and risk treatment;
d) be monitored;
e) be communicated;
f) be updated as appropriate;
g) be available as documented information.
The organization shall retain documented information on the information security objectives.


Page : 10


Source : ./example.pdf


Score : 0.4910152554512024


---------------------------------------
Result : 2



 chunk ID :56


Chunk Text : d) includes a commitment to continual improvement of the information security management system.
The information security policy shall:
e) be available as documented information;
f) be communicated within the organization;
g) be available to interested parties, as appropriate.
5.3  Organizational roles, responsibilities and authorities
Top management shall ensure that the responsibilities and authorities for roles relevant to information


Page : 8


Source : ./example.pdf


Score : 0.5233895778656006
```

```
 Result : 3



  chunk ID :75


 Chunk Text : requirements.
 7.4  Communication
 The organization shall determine the need for internal and external communications relevant to the
 information security management system including:
 a) on what to communicate;
 b) when to communicate;
 c) with whom to communicate;
 d) how to communicate.
 7.5  Documented information
 7.5.1  General
 The organization's information security management system shall include:
```

```
--------------------------------------
Result : 5



 chunk ID :55


Chunk Text : ISO/IEC 27001:2022(E)
5.2  Policy
Top management shall establish an information security policy that:
a) is appropriate to the purpose of the organization;
b) includes information security objectives (see 6.2) or provides the framework for setting information
security objectives;
c) includes a commitment to satisfy applicable requirements related to information security;
d) includes a commitment to continual improvement of the information security management system.


Page : 8


Source : ./example.pdf


Score : 0.6884709596633911
```

Query 3: What are the key steps involved in the information security risk assessment process?

TOP K=3

```
Enter the Query :What are the key steps involved in the information security risk assessment process?
Enter the TOP K value :3
1


---------------------------------------
Result : 1



 chunk ID :63


Chunk Text : materialize;
2) assess the realistic likelihood of the occurrence of the risks identified in 6.1.2 c) 1); and
3) determine the levels of risk;
e) evaluates the information security risks:
1) compare the results of risk analysis with the risk criteria established in 6.1.2 a); and
2) prioritize the analysed risks for risk treatment.
The organization shall retain documented information about the information security risk assessment
process.
6.1.3  Information security risk treatment


Page : 9


Source : ./example.pdf


Score : 0.4325544834136963


---------------------------------------
Result : 2



 chunk ID :68


Chunk Text : process.
NOTE 4 The information security risk assessment and treatment process in this document aligns with the
principles and generic guidelines provided in ISO 31000 [5].
6.2  Information security objectives and planning to achieve them
The organization shall establish information security objectives at relevant functions and levels.
The information security objectives shall:
a) be consistent with the information security policy;
b) be measurable (if practicable);


Page : 10


Source : ./example.pdf


Score : 0.477225661277771


---------------------------------------
 ---------------------------------------
 Result : 3



  chunk ID :85


 Chunk Text : The organization shall perform information security risk assessments at planned intervals or when
 significant changes are proposed or occur, taking account of the criteria established in 6.1.2 a).
 The organization shall retain documented information of the results of the information security risk
 assessments.
 8.3  Information security risk treatment
```