Documentation for Qdrant Code

Important: This Code could run both in 3.11 and 3.14

Steps to Run the Code * pip install -r requirements.txt * python3 rag_qdrant.py

Libraries Required:

```
from langchain_community.document_loaders import PyPDFLoader
from langchain_text_splitters import RecursiveCharacterTextSplitter
from langchain_huggingface import HuggingFaceEmbeddings
from qdrant_client import QdrantClient
from qdrant_client.http.models import VectorParams,Distance,PointStruct
import re
import pprint
```

The above were all Required Libraries

- langchain_community.document_loader import PyPDFLoader helps us to pase the pdf document.
- langchain_text_splitters import RecursiveCharacterTextSplitter helps to chunk the parsed data from documents.
- langchain_hugging_face import HuggingFaceEmbeddings helps us to use the embeding model sentence-transformers/all-MiniLM-L6-v2
- qdrant_client import QdrantClient helps us to create and use the qdrant Vector DB
- qdrant_client.http.models import VectorParams,Distance,PointStruct helps us to configure the qdrant setup the Distance and get Vectorparams and PointStruct a structure of data to insert to Qdrant

Loading the PDF Docuement

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

loader=PyPDFLoader(filepath)
print(loader)

docs=loader.load()
```

Chunking the Docuemnt

```
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
```

The above Code was used to split the loaded document into chunks each with 500 Characters and 100 characters overlap with the Previous chunk to avoid Information loss and add the chunk index using for loop.

Embedding Loading

```
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)
```

Loading the embeding model using the HuggingFace Library **sentence-transformers/all-MiniLM-L6-v2** Used to embed the Chunk which was an Industry Standard

Destructure the Document Chunks into Text,Metadata,Chunk Id and then producing the Vectors using HuggingFace model

Creating the Qdrant Client and Preparing the Data to Insert

```
client=QdrantClient(path="./vectordb/qdrant")
client.recreate_collection(
  collection_name="ragdata",
  vectors_config=VectorParams(
    size=384,
    distance=Distance.COSINE
  )
)
vectors=embedings.embed_documents(texts)
points=[]
for chunk,vector in zip(chunks,vectors):
  points.append(
    PointStruct(
      id=chunk.metadata["chunk_id"],
      vector=vector,
      payload={
        "text":chunk.page_content,
        **chunk.metadata
      }
    )
  )
```

- The Above Code will create a Qdrant Client that will initiate the Qdrant Database
- We create the Collection named **ragdata** and configure the collection with vector_config set the size for parameters to **384** and Store using the **COSINE** Function
- Prepare the All the Points to be inserted to the Qdrant using a PointStruct imported from qdrant_client.http.models including the vectors and metadata

Insert the Data Into the QbridDB

```
client.upsert(
  collection_name="ragdata",
  points=points
)

info=client.get_collection("ragdata")
print(info.points_count)
```

we use the client.upsert to insert the data into the QubridDB and get_collection to have the info and metadata of data.

Query

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

query_embeding=embedings.embed_query(query)
```

```
results=client.query_points(
  collection_name="ragdata",
  query=query_embeding,
  limit=5
)

for index,i in enumerate(results.points):
  print("\n\n")
  print('-'*40)
  print(f"\nResult : {index+1}")
  print(f"\nChunk ID : {i.id}")
  print(f"\nChunk Content :{i.payload['text']}")

  print(f"\nScore : {i.score}")
  print(f"\nPage :{i.payload['page']}")
  print(f"\nSource : {i.payload['source']}")

client.close()
```

The Query was first embeded with the same Model we embeded the chunks from PDF Document.

Then we search the QubridDB using the query and set the limit to control how many relavent results we see.

Sample Outputs

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

Top K=5

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

Top K=5

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

Top K=3

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



---------------------------------------

Result : 1

Chunk ID : 51

Chunk Content :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;

Score : 0.5385089103271771

Page :7

Source : ./example.pdf



---------------------------------------

Result : 2

Chunk ID : 57

Chunk Content :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

Score : 0.506787221224803

Page :8

Source : ./example.pdf



---------------------------------------
```

Figure 1: Qubrid Output

```
---------------------------------------

Result : 3

Chunk ID : 38

Chunk Content :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.

Score : 0.3976157633156742

Page :4

Source : ./example.pdf



---------------------------------------

Result : 4

Chunk ID : 53

Chunk Content :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.

Score : 0.3613273030113159

Page :7

Source : ./example.pdf



---------------------------------------

Result : 5

Chunk ID : 36

Chunk Content :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.

Score : 0.3352837502690068

Page :4

Source : ./example.pdf
```

Figure 2: Qubrid Output

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



---------------------------------------

Result : 1

Chunk ID : 69

Chunk Content :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.

Score : 0.7544924035586212

Page :10

Source : ./example.pdf



---------------------------------------

Result : 2

Chunk ID : 56

Chunk Content :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

Score : 0.7383052704048987

Page :8

Source : ./example.pdf



---------------------------------------
```

Figure 3: Qubrid Output

```
Result : 3

Chunk ID : 75

Chunk Content :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:
a) documented information required by this document; and
   © ISO/IEC 2022 – All rights reserved

Score : 0.7065966622044351

Page :11

Source : ./example.pdf



---------------------------------------

Result : 4

Chunk ID : 119

Chunk Content :laws and regulations and contractual requirements.
5.35 Independent review of informa-
tion security
Control
The organization’s approach to managing information security and
its implementation including people, processes and technologies shall
be reviewed independently at planned intervals, or when significant
changes occur.
5.36 Compliance with policies, rules
and standards for information
security
Control
Compliance with the organization’s information security policy, top -

Score : 0.6591124526267529

Page :18

Source : ./example.pdf



---------------------------------------
```

Figure 4: Qubrid Output

```
--------------------------------------

Result : 5

Chunk ID : 55

Chunk Content :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.

Score : 0.6557645241889571

Page :8

Source : ./example.pdf
```

Figure 5: Qubrid Output

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



---------------------------------------

Result : 1

Chunk ID : 63

Chunk Content :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

Score : 0.7837227773917174

Page :9

Source : ./example.pdf



---------------------------------------

Result : 2

Chunk ID : 68

Chunk Content :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);

Score : 0.76138721445966

Page :10

Source : ./example.pdf



---------------------------------------
```

Figure 6: Qubrid Output

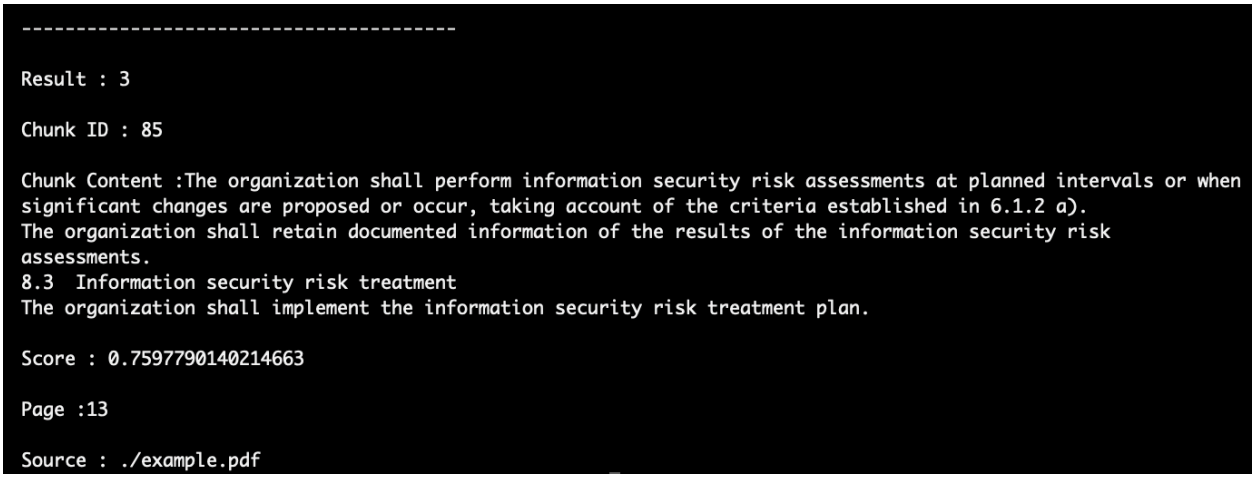

Figure 7: Qubrid Output