Documentation for Chroma

Retrieval Augmented Generation using ChromaDB

1 Environment Requirements

- This code will run only on **Python 3.11**
- This code is **not compatible** with the latest version of Python 3.14

2 Steps to Run the Code

```
git clone https://github.com/adityasaicareer/
   celestial_systems_chroma.git
cd celestial_systems_chroma
pip install -r requirements.txt
python rag_chroma.py
```

3 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
```

3.1 Library Description

- **PyPDFLoader** parses PDF documents to extract text content.
- **RecursiveCharacterTextSplitter** splits documents into manageable chunks.
- **HuggingFaceEmbeddings** uses the `sentence-transformers/all-MiniLM-L6-v2` model to generate embeddings.

- **chromadb** is used to create and manage a local vector database.
- **Settings** configures ChromaDB persistence and storage.

4 Loading the PDF

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

loader = PyPDFLoader(filepath)
print(loader)

docs = loader.load()
```

The above code demonstrates loading a PDF document using the LangChain `PyPDFLoader`.

5 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
```

Each document is split into chunks of 500 characters with an overlap of 100 characters to preserve contextual continuity.

6 Embedding the Chunks

```
embeddings = 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 = embeddings.embed_documents(texts)
```

The embeddings are generated using the industry-standard MiniLM model.

7 ChromaDB Setup

```
client = chromadb.Client(
    Settings(persist_directory="./vectordb/chroma")
)

collection = client.create_collection(name="my_collection")

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

- A ChromaDB client is initialized
- A collection is created
- Document embeddings, metadata, and IDs are stored using `upsert`

8 Querying the Vector Database

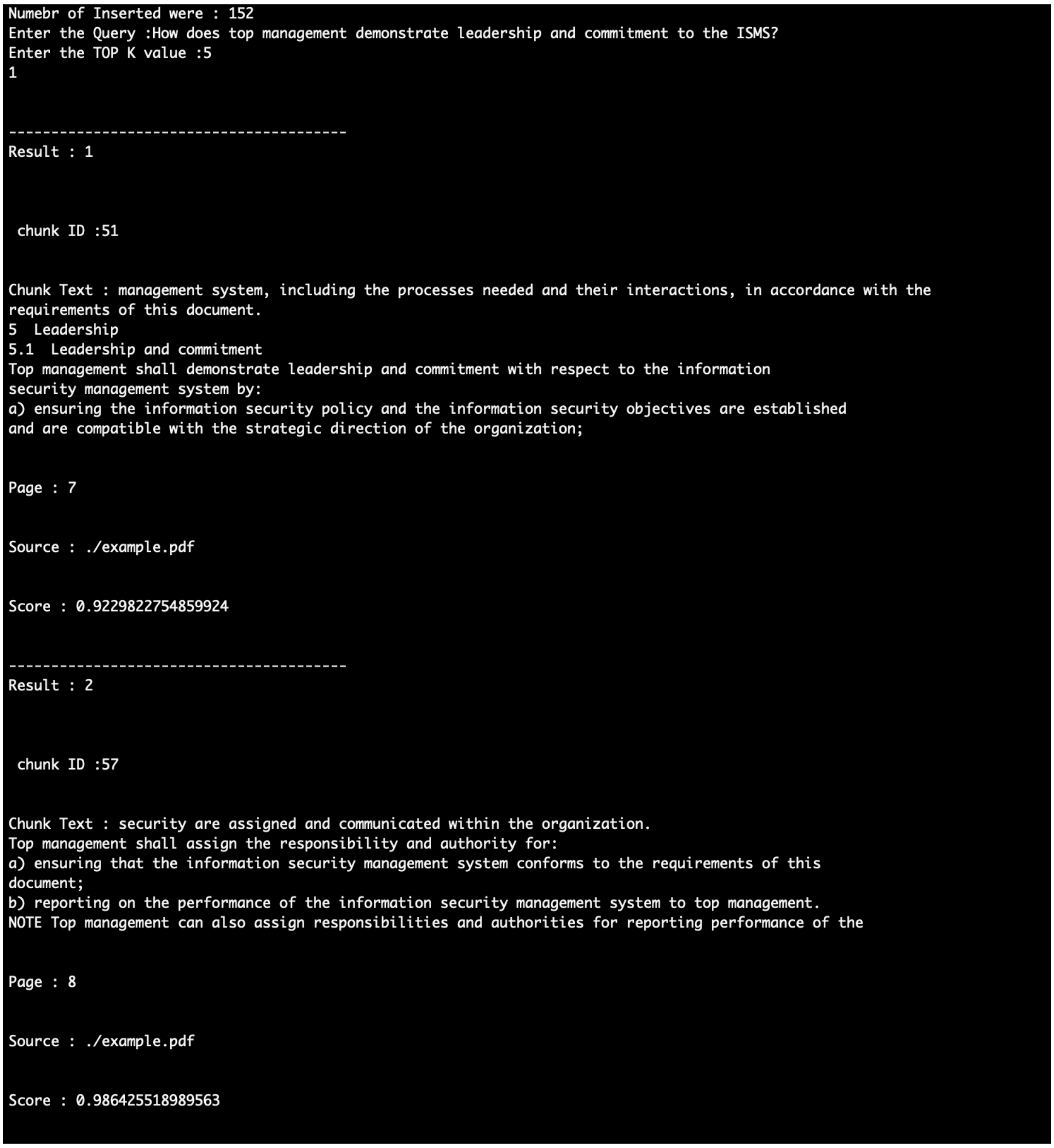
```
query = "How␣does␣top␣management␣demonstrate␣leadership␣and␣
   commitment␣to␣the␣ISMS?"

query_vector = embeddings.embed_query(query)

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

The query is embedded and matched against stored vectors to retrieve the most relevant chunks.

9 Outputs

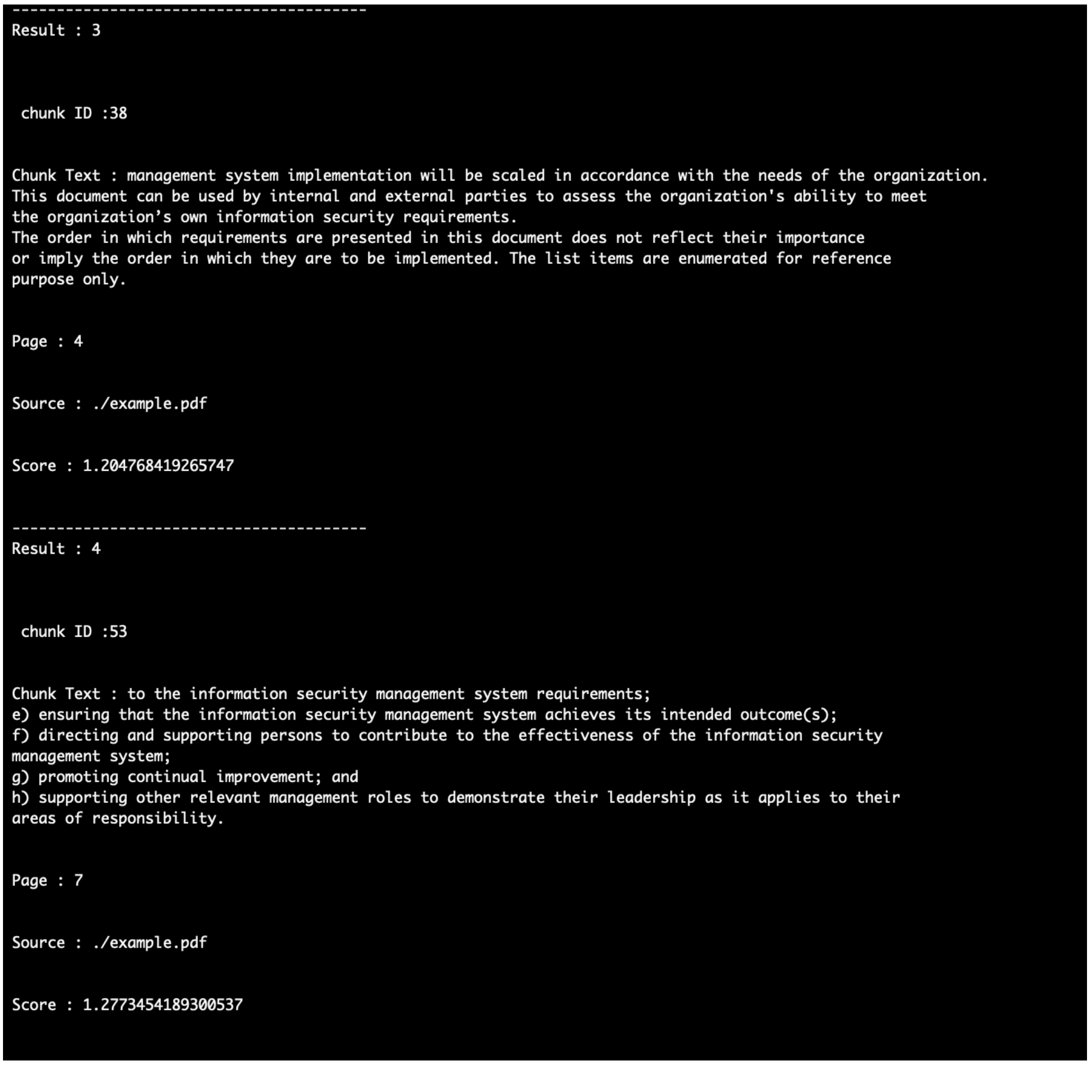
9.1 Query 1

How does top management demonstrate leadership and commitment to the ISMS?
Top K = 5



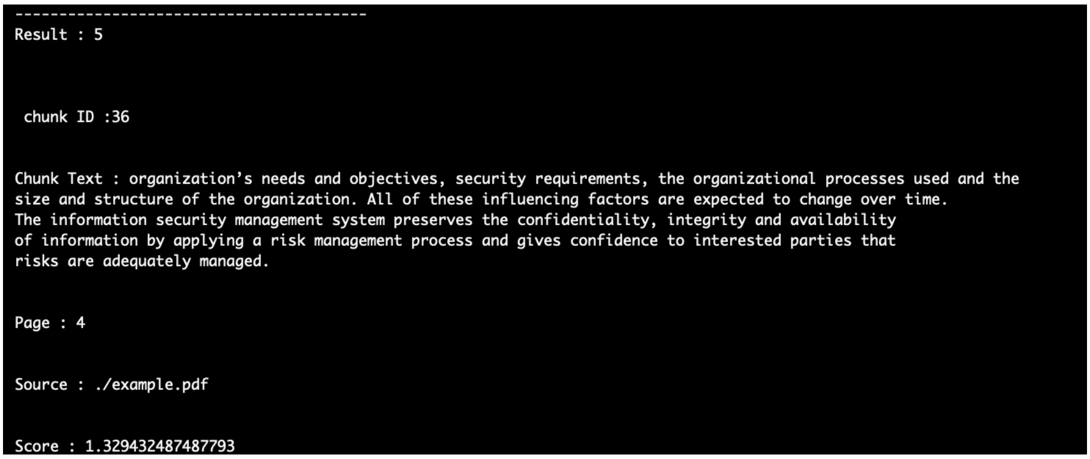
```
Numebr of Inserted were : 152
Enter the Query :How does top management demonstrate leadership and commitment to the ISMS?
Enter the TOP K value :5
1


---------------------------------------
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: Query 1 Result - Output 1



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

Figure 2: Query 1 Result - Output 2

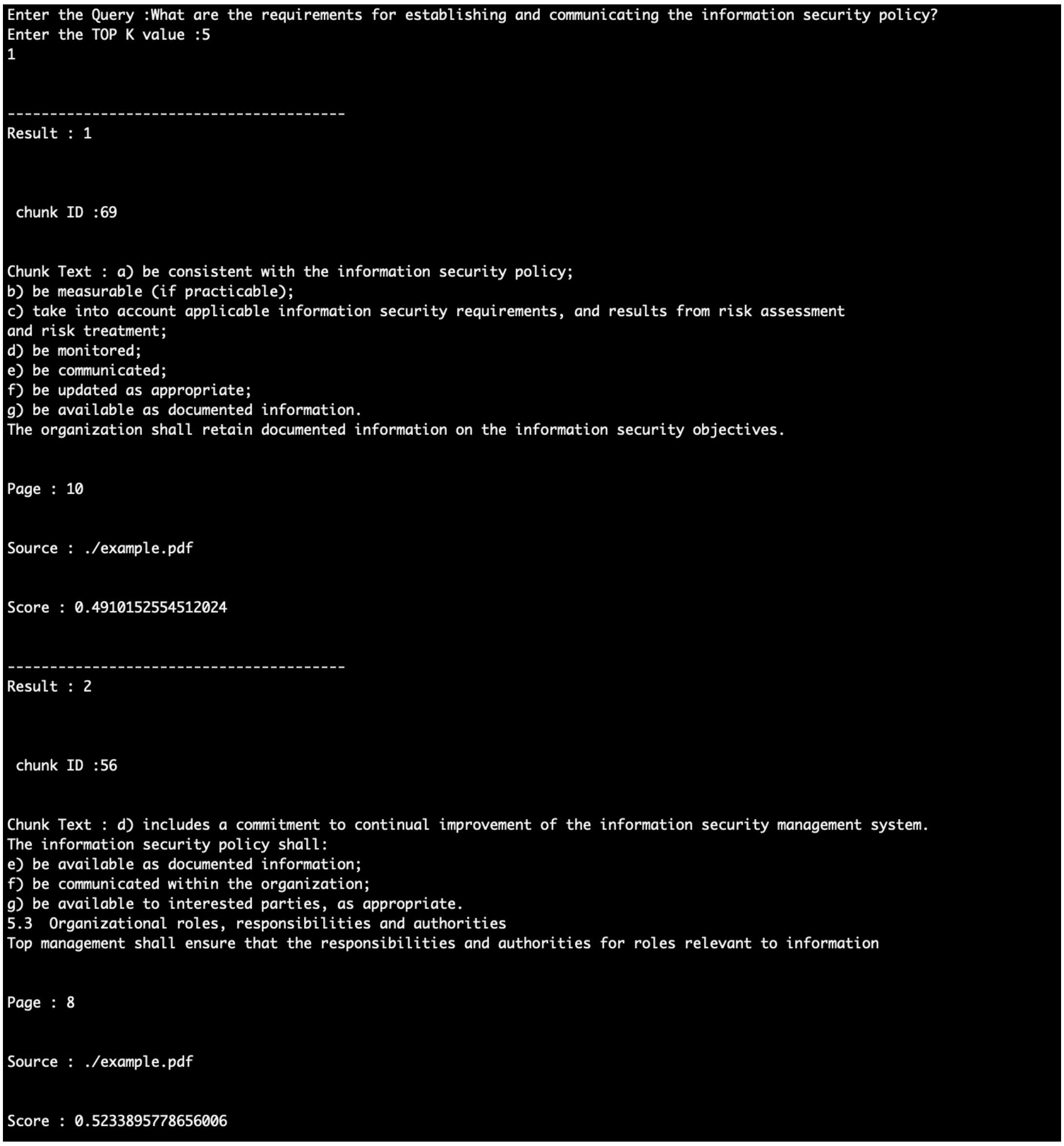


```
----------------------------------------
Result : 5



 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
```

Figure 3: Query 1 Result - Output 3

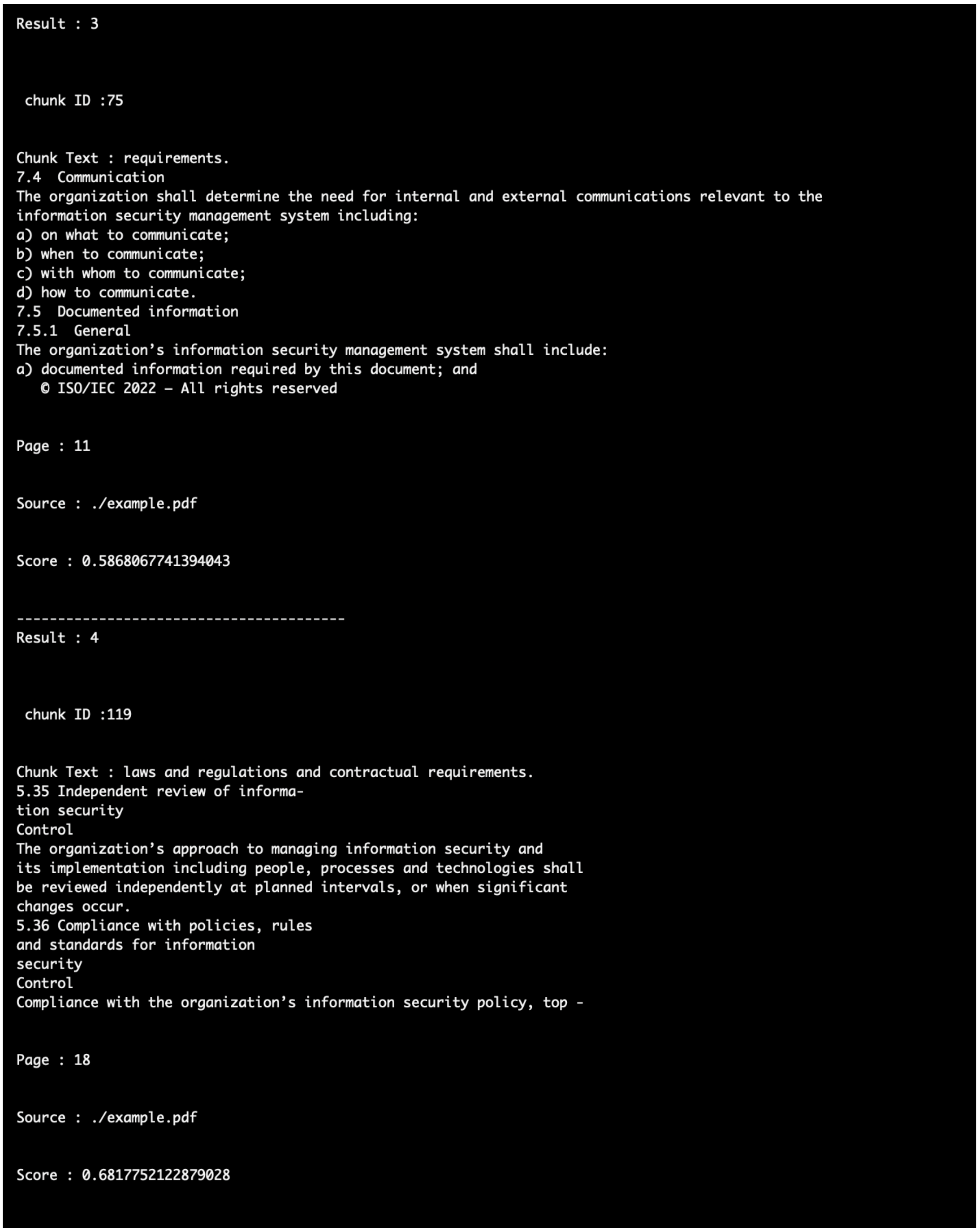
9.2 Query 2

What are the requirements for establishing and communicating the information security policy?
Top K = 5



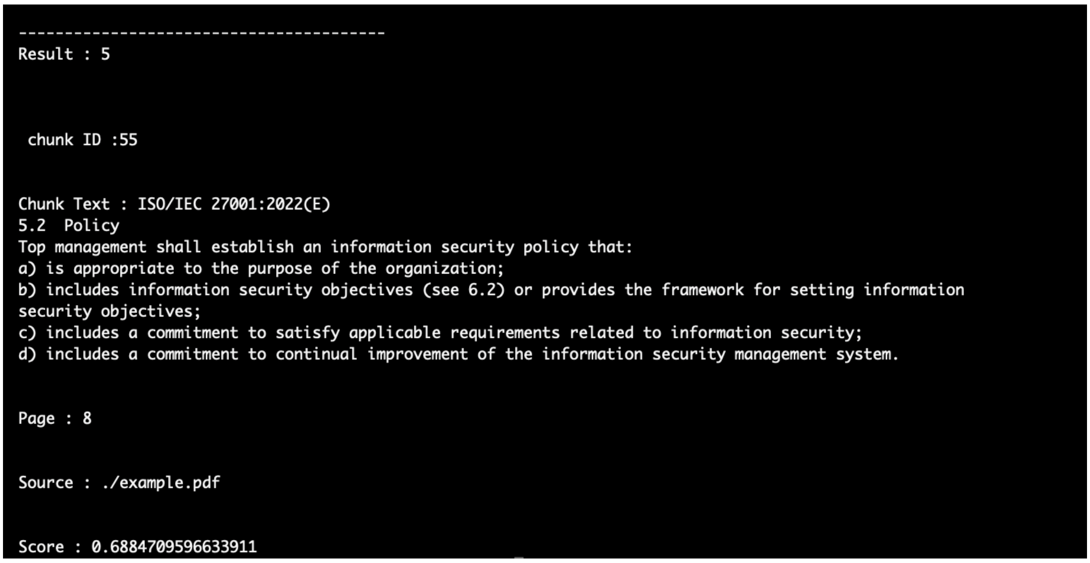
```
Enter the Query :What are the requirements for establishing and communicating the information security policy?
Enter the TOP K value :5
1


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

Figure 4: Query 2 Result - Output 1



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


Page : 11


Source : ./example.pdf


Score : 0.5868067741394043


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



 chunk ID :119


Chunk Text : 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 -


Page : 18


Source : ./example.pdf


Score : 0.6817752122879028
```

Figure 5: Query 2 Result - Output 2

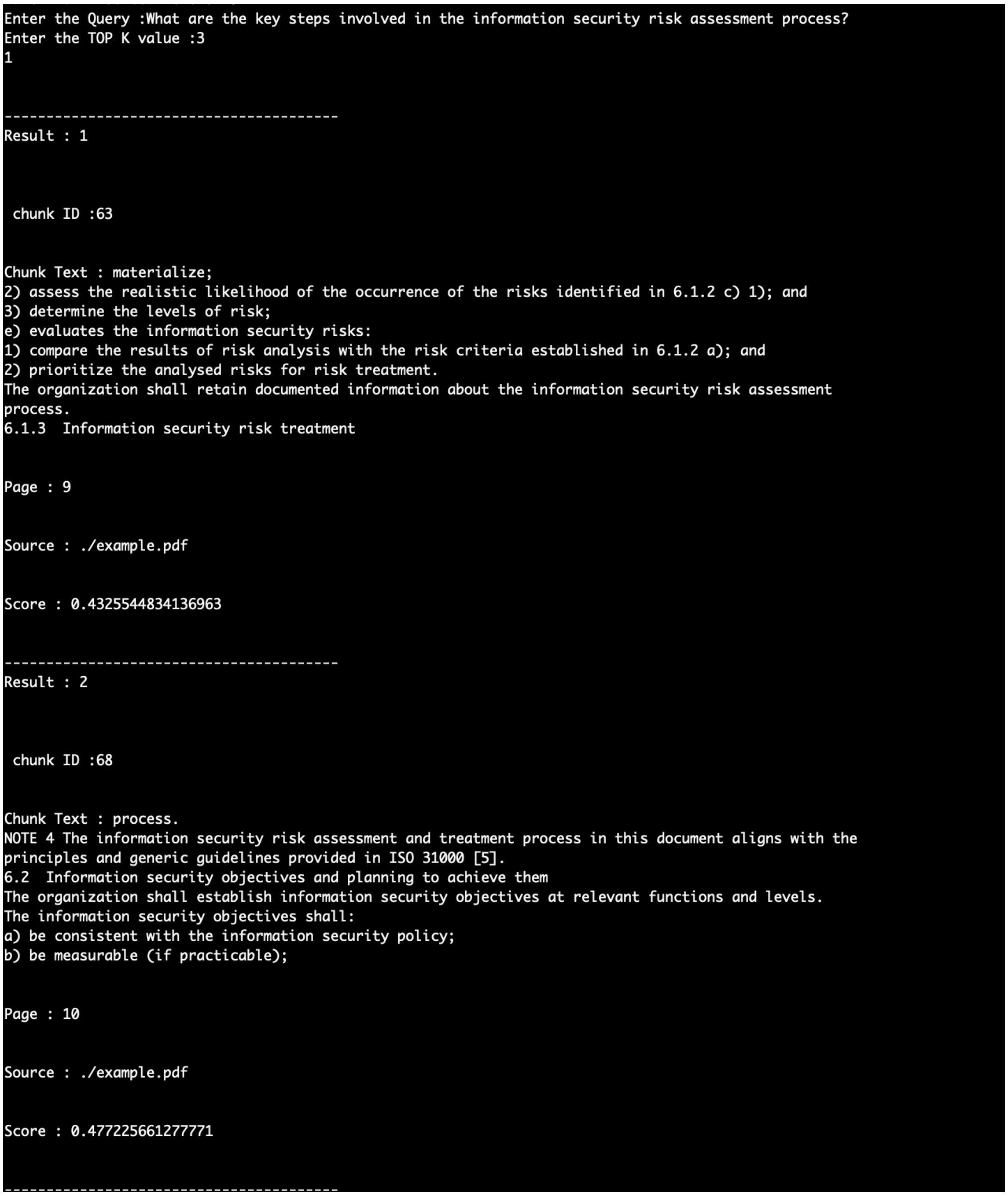


```
---------------------------------------
Result : 5



 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
```

Figure 6: Query 2 Result - Output 3

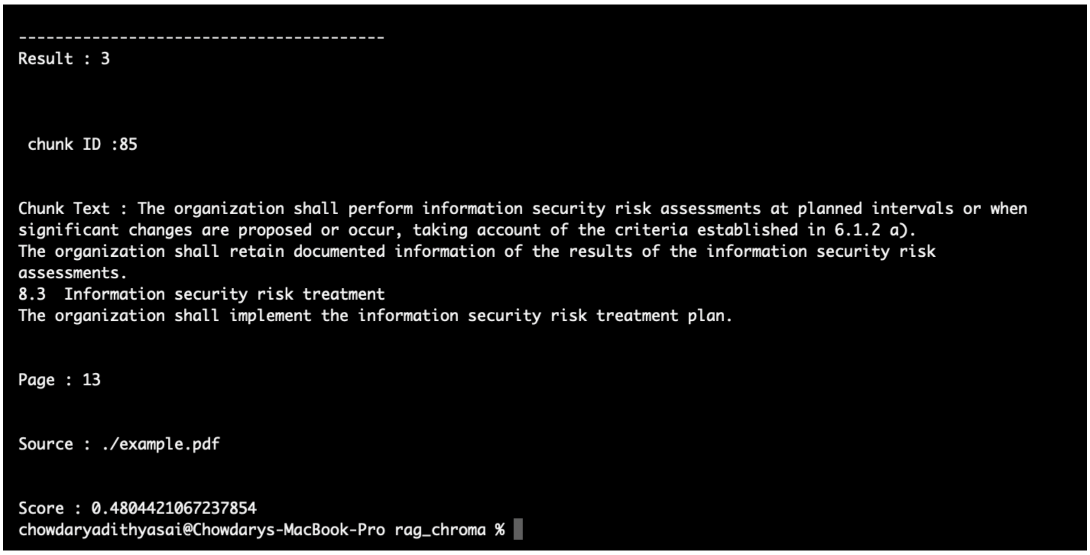
9.3 Query 3

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


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

Figure 7: Query 3 Result - Output 1



```
----------------------------------------
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
The organization shall implement the information security risk treatment plan.


Page : 13


Source : ./example.pdf


Score : 0.4804421067237854
chowdaryadithyasai@Chowdarys-MacBook-Pro rag_chroma %
```

Figure 8: Query 3 Result - Output 2