

Lab 4: OpenSearch Integration

CSI403 - Full Stack Development

Faculty of Information Technology

Sripatum University

Weight: 3.75%

Objectives

- ✓ Connect to OpenSearch from Python
- ✓ Implement hybrid search
- ✓ Index and search documents

Repository: <https://github.com/amornpan/Generic-RAG>

Task 1: Connect to OpenSearch

```
from opensearchpy import OpenSearch

client = OpenSearch(
    hosts=[{"host": "localhost", "port": 9200}],
    use_ssl=False
)
print(client.info())
```

Task 2: Create Index

```
index_body = {  
    "settings": {"index": {"knn": True}},  
    "mappings": {  
        "properties": {  
            "content": {"type": "text"},  
            "content_vector": {  
                "type": "knn_vector",  
                "dimension": 1024  
            }  
        }  
    }  
}  
client.indices.create(index="documents", body=index_body)
```

Task 3: Index Documents

```
doc = {  
    "content": "Sample\u2022document\u2022text",  
    "content_vector": [0.1, 0.2, ...] # 1024 dims  
}  
client.index(index="documents", body=doc)
```

Note: Vector dimension must match
your embedding model output (1024 for bge-m3).

Task 4: Implement Hybrid Search

Implement hybrid search combining:

Vector Search (KNN)

- Semantic similarity
- Dense vectors
- Weight: 0.7

BM25 (Keyword)

- Lexical matching
- Term frequency
- Weight: 0.3

Deliverables

Item	Check
Connection working	<input type="checkbox"/>
Index created	<input type="checkbox"/>
Documents indexed	<input type="checkbox"/>
Search implemented	<input type="checkbox"/>

Deadline: Sunday 23:59