



HOW TO INSTALL TURBO CHARGERS ON VECTOR RETRIEVERS



INTRODUCING CONTEXTUAL COMPRESSION RETRIEVERS

WHY COMPRESSION?

GIVEN A QUERY,

1) ONLY RETURN THE RELEVANT DOCUMENT

2) PARTS OF DOCUMENT THAT WILL PROVIDE HIGHER QUALITY REPLY

HOW COMPRESSION?

LANGCHAIN USES 3 NEW WRAPPER CLASSES

1) LLM CHAIN EXTRACTORS : INDUCES COMPRESSION

2) LLM CHAIN FILTER : FILTERS BY RELEVANCE

3) EMBEDDING FILTER : FILTERS BY SIMILARITY

4) EMBEDDING REDUNDANT FILTER : ENSURING BEST OUTPUT

CHALLENGE SOLVED :
GAINING HIGHER
QUALITY CONTEXT

PIPELINE WITH COMPRESSION RETRIEVER?

STEPS AND CODE

```
LLM = YOUR_LLM(TEMP=0.0)
```

```
RETRIEVER = DOCUMENT_STORE.AS_RETRIEVER()
```

```
UR_COMPRESSOR = LLMCHAINEXTRACTOR.FROM_LLM(LLM)
```

```
COMPRESSION_RETRIEVER =  
CONTEXTUALCOMPRESSIONRETRIEVER(BASE_COMPRESSOR=UR_COMPRESSOR,  
BASE_RETRIEVER=RETRIEVER)
```

```
UR_FILTER = LLMCHAINFILTER.FROM_LLM(LLM)
```

```
FILTERED_RETRIEVER =  
CONTEXTUALCOMPRESSIONRETRIEVER(BASE_COMPRESSOR=UR_FILTER,  
BASE_RETRIEVER=RETRIEVER)
```

```
EMBED_FILTER = EMBEDDINGSFILTER(EMBEDDINGS=EMBEDDINGS,  
SIMILARITY_THRESHOLD=0.76)
```

```
COMPRESSION_RETRIEVER =  
CONTEXTUALCOMPRESSIONRETRIEVER(BASE_COMPRESSOR=EMBED_FILTER,  
BASE_RETRIEVER=RETRIEVER)
```

LETS HEAD TO COLAB

NOW WE ARE TALKING!!! PRACTICE

**insightbuilder/
python_de_learners_data**



Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning

2 Contributors 0 Issues 27 Stars 10 Forks

[GitHub](#)

python_de_learners_data/compression_retrievers.ipynb at main · insightbuilder/python_de_learners_data

Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning - python_de_learners_data/compression_retrievers.ipynb at main ...

[GitHub](#)

THANKS FOR WATCHING

 **LIKE**

 **SHARE**

 **SUBSCRIBE**