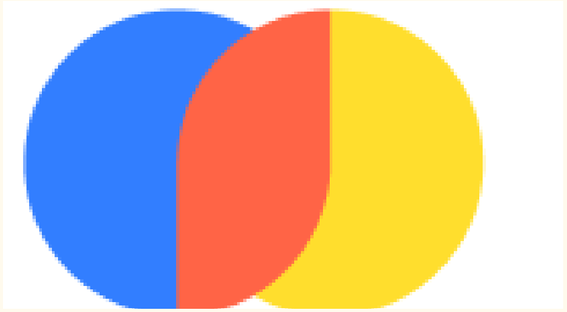
EFFICIENTLY STORE EMBEDDINGS IN CHROMADB WITH TOWHEE



STORE & QUERY EMBEDDING LOCALLY WITH CHROMADB

KEY IS IN UNDERSTANDING ENTITIES

- PIPE MODULE CAN BE USED FOR CREATING THE EMBEDDINGS
- OUTPUT DATAQUEUE IS TYPECASTED TO DATA COLLECTION
- DATACOLLECTION CONTAINS THE ENTITIES THAT ARE CONVERTED TO PYTHON LISTS
- THE LISTS ARE DIRECTLY LOADED INTO CHROMADB

```
IMPORT CHROMADB
CLIENT = CHROMADB.CLIENT()
COLLECTION = CLIENT.CREATE_COLLECTION("TOWHEE_TEST")
INSERT_PIPE = (PIPE.INPUT('DF')
.FLAT_MAP('DF', ('ID','ARTICLE_HEAD','VEC'), LAMBDA DF: DF.VALUES.TOLIST())
.OUTPUT('ID','ARTICLE_HEAD','VEC'))
```

TOWHEE NODES IN ACTION

```
PIPE_W_DATA = INSERT_PIPE(DF[:1000])
PIPE_W_DATALIST = DATACOLLECTION(PIPE_W_DATA).TO_LIST()

IDS = [STR(ELEM.ID) FOR ELEM IN PIPE_W_DATALIST]
HEAD_TEXTS = [{"TEXT":ELEM.ARTICLE_HEAD} FOR ELEM IN PIPE_W_DATALIST]
EMBEDS = [ELEM.VEC FOR ELEM IN PIPE_W_DATALIST]
```

```
COLLECTION.ADD(IDS=IDS,
               METADATAS=HEAD_TEXTS,
               EMBEDDINGS=EMBEDS)
```

```
SEARCH_PIPE =PIPE.INPUT('QUERY')
.MAP('QUERY', 'VEC', OPS.TEXT_EMBEDDING.DPR(MODEL_NAME="FACEBOOK/DPR-CTX_ENCODER-SINGLE-NQ-
BASE"))
.MAP('VEC', 'METADATAS', LAMBDA X : COLLECTION.QUERY(QUERY_EMBEDDINGS=[X],N_RESULTS=2))
.OUTPUT('QUERY', 'METADATAS')
```

PRACTICE NOTEBOOK!!!

READING THE CODE IS NEVER ENOUGH TO UNDERSTAND

HTTPS://GITHUB.COM/INSIGHTBUILDER

THANKS FOR WATCHING

HTTPS://GITHUB.COM/INSIGHTBUILDER