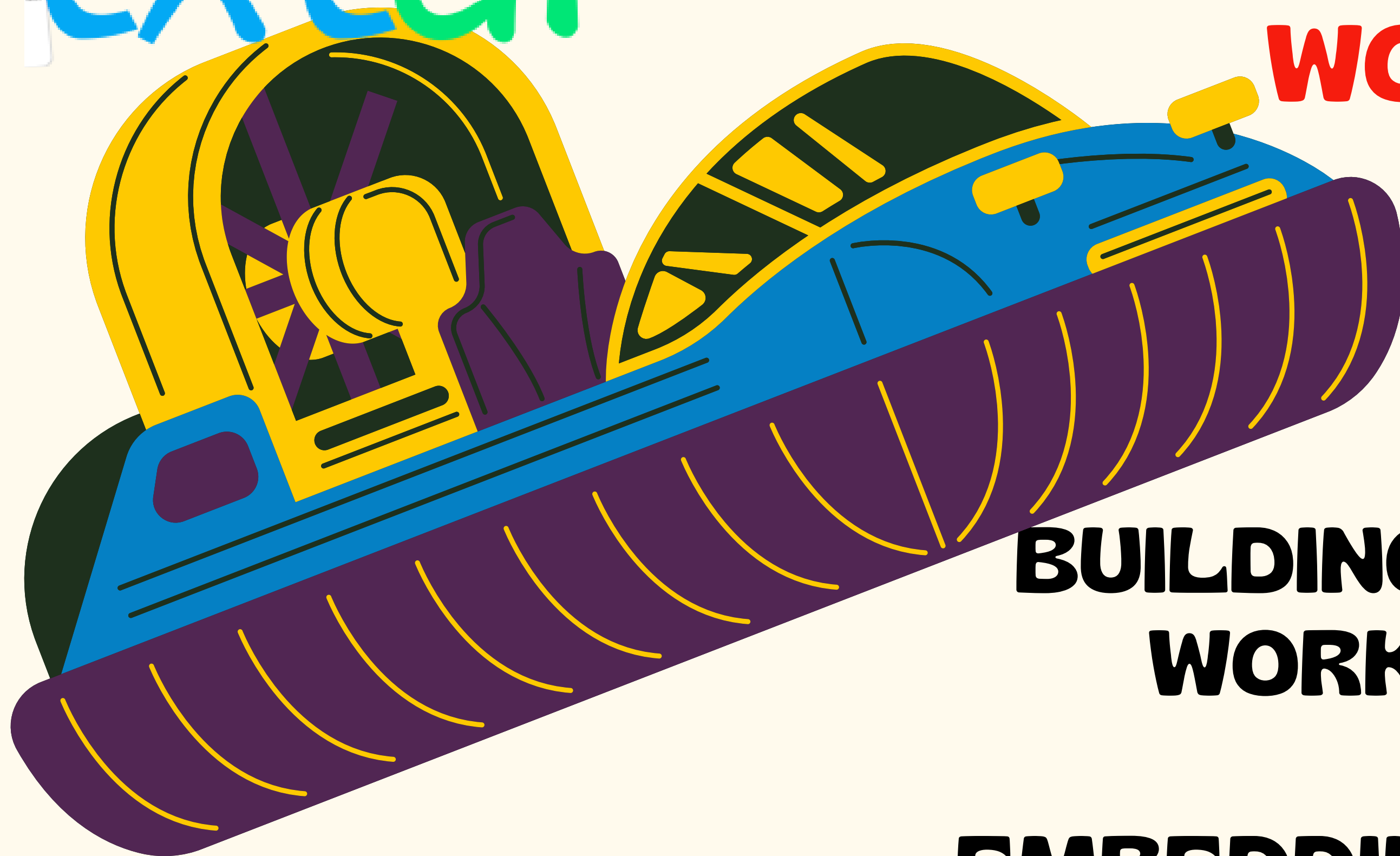


txtai

DIVING INTO TXTAI: THE ENGINE OF NLP WORKFLOWS



**BUILDING PIPELINES,
WORKFLOWS AND
RDBMS FOR
EMBEDDING VECTORS**

CHALLENGES SOLVED

1) EMBEDDING DATA AND QUERYING INDEX MADE INTUITIVE

2) TEXT, AUDIO, IMAGES & VIDEO ALIKE CAN BE EMBEDDED WITH EASE

2) EMBEDDING INDEX CAN BE USED WITH PIPELINES INSIDE TXTAI

3) MULTIPLE FLOWS ALREADY CODED, THAT CAN USE OPEN SOURCE MODELS

4) PIPELINES AND FLOWS CAN BE BUILT WITH PYTHON OR YAML

EXAMPLES TO REVIEW

OUT OF 40 EXAMPLES, FOLLOWING WILL BE MOST IMPORTANT FOR THE BEGINNER

- [HTTPS://GITHUB.COM/NEUML/TXTAI/BLOB/MASTER/EXAMPLES/01_INTRODUCING_TXTAI.IPYNB](https://github.com/NEUML/TXTAI/blob/master/EXAMPLES/01_INTRODUCING_TXTAI.IPYNB)
- [HTTPS://GITHUB.COM/NEUML/TXTAI/BLOB/MASTER/EXAMPLES/03_BUILD_AN_EMBEDDINGS_INDEX_FROM_A_DATA_SOURCE.IPYNB](https://github.com/NEUML/TXTAI/blob/master/EXAMPLES/03_BUILD_AN_EMBEDDINGS_INDEX_FROM_A_DATA_SOURCE.IPYNB)
- [HTTPS://GITHUB.COM/NEUML/TXTAI/BLOB/MASTER/EXAMPLES/30_EMBEDDINGS_SQL_CUSTOM_FUNCTIONS.IPYNB](https://github.com/NEUML/TXTAI/blob/master/EXAMPLES/30_EMBEDDINGS_SQL_CUSTOM_FUNCTIONS.IPYNB)
- [HTTPS://GITHUB.COM/NEUML/TXTAI/BLOB/MASTER/EXAMPLES/33_QUERY_TRANSLATION.IPYNB](https://github.com/NEUML/TXTAI/blob/master/EXAMPLES/33_QUERY_TRANSLATION.IPYNB)
- [HTTPS://GITHUB.COM/NEUML/TXTAI/BLOB/MASTER/EXAMPLES/42_PROMPT_DRIVEN_SEARCH_WITH_LLMS.IPYNB](https://github.com/NEUML/TXTAI/blob/master/EXAMPLES/42_PROMPT_DRIVEN_SEARCH_WITH_LLMS.IPYNB)
- [HTTPS://GITHUB.COM/NEUML/TXTAI/BLOB/MASTER/EXAMPLES/14_RUN_PIPELINE_WORKFLOWS.IPYNB](https://github.com/NEUML/TXTAI/blob/master/EXAMPLES/14_RUN_PIPELINE_WORKFLOWS.IPYNB)
- [HTTPS://GITHUB.COM/NEUML/TXTAI/BLOB/MASTER/EXAMPLES/26_ENTITY_EXTRACTION_WORKFLOWS.IPYNB](https://github.com/NEUML/TXTAI/blob/master/EXAMPLES/26_ENTITY_EXTRACTION_WORKFLOWS.IPYNB)
- [HTTPS://GITHUB.COM/NEUML/TXTAI/BLOB/MASTER/EXAMPLES/SIMILARITY.PY](https://github.com/NEUML/TXTAI/blob/master/EXAMPLES/SIMILARITY.PY)
- [HTTPS://GITHUB.COM/NEUML/TXTAI/BLOB/MASTER/EXAMPLES/ARTICLE.PY](https://github.com/NEUML/TXTAI/blob/master/EXAMPLES/ARTICLE.PY)
- [HTTPS://GITHUB.COM/NEUML/TXTAI/BLOB/MASTER/EXAMPLES/WORKFLOWS.PY](https://github.com/NEUML/TXTAI/blob/master/EXAMPLES/WORKFLOWS.PY)

[HTTPS://GITHUB.COM/INSIGHTBUILDER](https://github.com/INSIGHTBUILDER)

KEY TOPICS

- **EMBEDDING DATABASES : UNION OF VECTOR INDEXES + RELATIONAL DB**

- **STEPS FOR CREATING INDEXES:**

- **DATA IN LIST FORMAT**
- **EMBEDDING FUNCTION INITIATED**
- **EMBEDDING INDEX INITIATED**
- **SEARCH EXECUTED ON EMBEDDING**
- **SAVING & LOADING THE INDEX**
- **QUERYING WITH SQL**
- **DOING ETL WITH SQL**

- **WHAT ARE PIPELINES?**

A CLASS THAT APPLIES ML MODELS TO THE INPUT DATA AND PROVIDES OUT

1. QUESTIONS - ANSWER QUESTIONS USING A TEXT CONTEXT
2. LABELS - APPLY LABELS TO TEXT USING A ZERO-SHOT CLASSIFICATION MODEL. ALSO SUPPORTS SIMILARITY COMPARISONS.
3. SUMMARY - ABSTRACTIVE TEXT SUMMARIZATION
4. TEXTRACTOR - EXTRACT TEXT FROM DOCUMENTS
5. TRANSCRIPTION - TRANSCRIBE AUDIO TO TEXT
6. TRANSLATION - MACHINE TRANSLATION

- **WHAT ARE WORKFLOWS**

SERIES OF TASKS THAT ARE PIPELINES OR PYTHON FUNCTIONS. THESE TASKS CAN BE CHAINED TOGETHER TO DO MULTIPLE STEPS

LETS SEE SOME ACTION

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



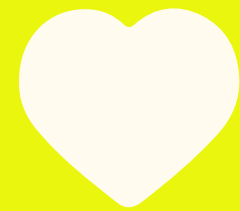
python_de_learners_data/building_index_txtAi.ipynb at main · insightbuilder/python_de_learners_data

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

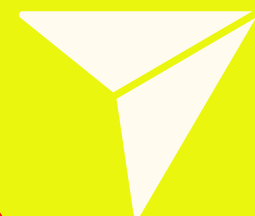


GitHub

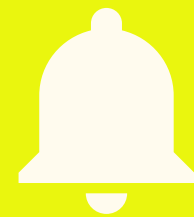
THANKS FOR WATCHING



LIKE



SHARE



SUBSCRIBE