

# 7 LIBRARIES TO SUPERCHARGE YOUR NLP JOURNEY



**OPEN!**  
SOURCE

**EMBED, MODEL AND  
EXECUTE SEMANTIC  
SEARCH LIKE A PRO**

# COMMON CHALLENGES

**FOREST OF NLP (NATURAL LANGUAGE PROCESSING) IS VERY DENSE WITH MANY CHALLENGES**

**1) SPLITTING OR TOKENISING TEXT**

**2) EMBEDDING OR VECTORISING TEXT**

**3) MODELING TOPICS**

**4) MULTIPLE SEARCH ALGORITHMS (SEMANTIC / NEAREST NEIGHBOURS)**

**5) QA MODELING**

**6) KEY WORD / PHRASE EXTRACTION**

**7) MAKING MODELS ACCESSIBLE**

**8) UNDERSTANDING HOW MODELS TAKE INPUT AND RETURN OUTPUT**

**9) PERFORMANCE BOTTLENECKS DUE TO THE CHOSEN ALGORITHMS**

# 7 LIBRARIES TO RESCUE

MODELING, SEARCHING & EMBEDDING MULTI-LINGUAL SENTENCES & MORE

- 1) **HAYSTACK** : AN ALL ROUND LIBRARY THAT CAN WORK WITH MULTIPLE MODELS (SIMILAR TO LANGCHAIN)
- 2) TOP2VEC: DETECTS TOPICS PRESENT IN TEXT AND GENERATES JOINTLY EMBEDDED TOPIC, DOCUMENT AND WORD VECTORS
- 3) GENSIM: GENSIM IS DESIGNED TO PROCESS RAW, UNSTRUCTURED DIGITAL TEXTS (“PLAIN TEXT”) USING UNSUPERVISED MACHINE LEARNING ALGORITHMS.
- 4) **TXTAI**: TXTAI BUILDS EMBEDDINGS DATABASES, WHICH ARE A UNION OF VECTOR INDEXES AND RELATIONAL DATABASES.
- 5) BERTTOPIC: TOPIC MODELING TECHNIQUE THAT LEVERAGES TRANSFORMERS AND C-TF-IDF TO CREATE DENSE CLUSTERS ALLOWING FOR EASILY INTERPRETABLE TOPICS WHILST KEEPING IMPORTANT WORDS
- 6) KEYBERT: IS A MINIMAL AND EASY-TO-USE KEYWORD EXTRACTION TECHNIQUE THAT LEVERAGES BERT EMBEDDINGS TO CREATE KEYWORDS AND KEYPHRASES THAT ARE MOST SIMILAR TO A DOCUMENT.
- 7) **BACKPROP** - NATURAL LANGUAGE ENGINE THAT MAKES USING STATE-OF-THE-ART LANGUAGE MODELS EASY, ACCESSIBLE AND SCALABLE.

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

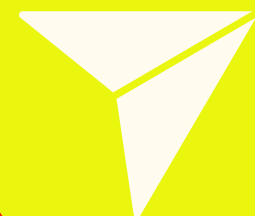
# LETS READ ABOUT THE LIBRARIES

- [HTTPS://GITHUB.COM/DEEPSET-AI/HAYSTACK](https://github.com/deepset-ai/haystack)
- [HTTPS://GITHUB.COM/DDANGELOV/TOP2VEC](https://github.com/ddangelov/top2vec)
- [HTTPS://GITHUB.COM/RARE-TECHNOLOGIES/GENSIM](https://github.com/rare-technologies/gensim)
- [HTTPS://GITHUB.COM/NEUML/TXTAI](https://github.com/neuml/txtai)
- [HTTPS://GITHUB.COM/MAARTENGR/BERTOPIC](https://github.com/maartengr/bertopic)
- [HTTPS://GITHUB.COM/MAARTENGR/KEYBERT](https://github.com/maartengr/keybert)
- [HTTPS://GITHUB.COM/MILANLPROC/CONTEXTUALIZED-TOPIC-MODELS](https://github.com/milanlproc/contextualized-topic-models)
- [HTTPS://GITHUB.COM/BACKPROP-AI/BACKPROP](https://github.com/backprop-ai/backprop)

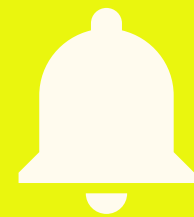
# THANKS FOR WATCHING



**LIKE**



**SHARE**



**SUBSCRIBE**