

Introducción al aprendizaje de máquina

Aprendizaje de Máquina Aplicado

César Leandro Higueta Pérez
clhiguitap@eafit.edu.co

2023

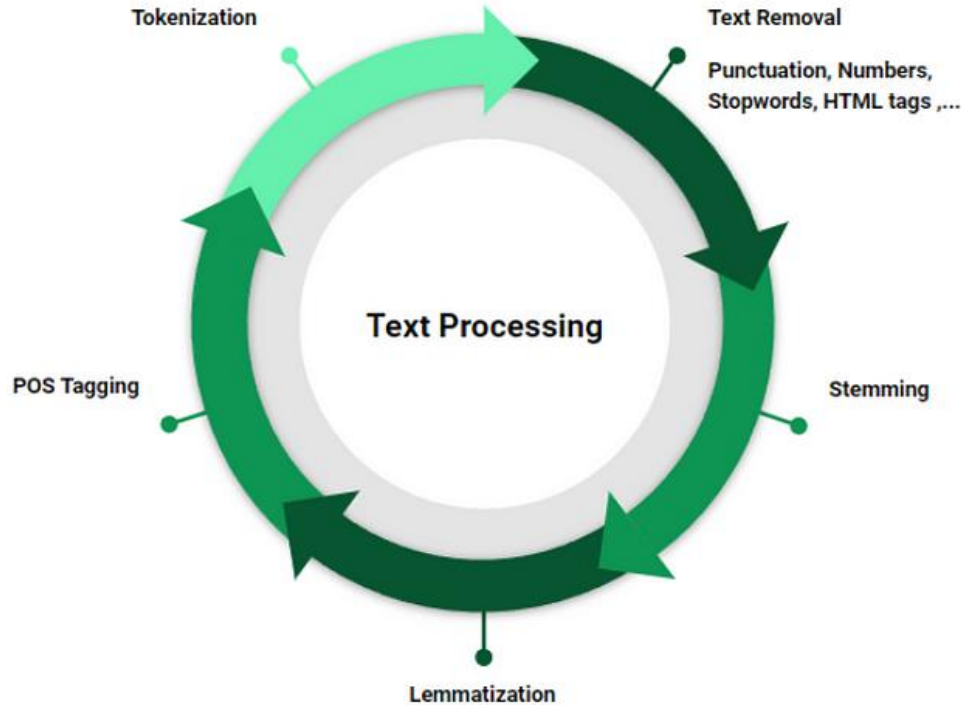
David Felipe Luna
dflunan@eafit.edu.co

Agenda

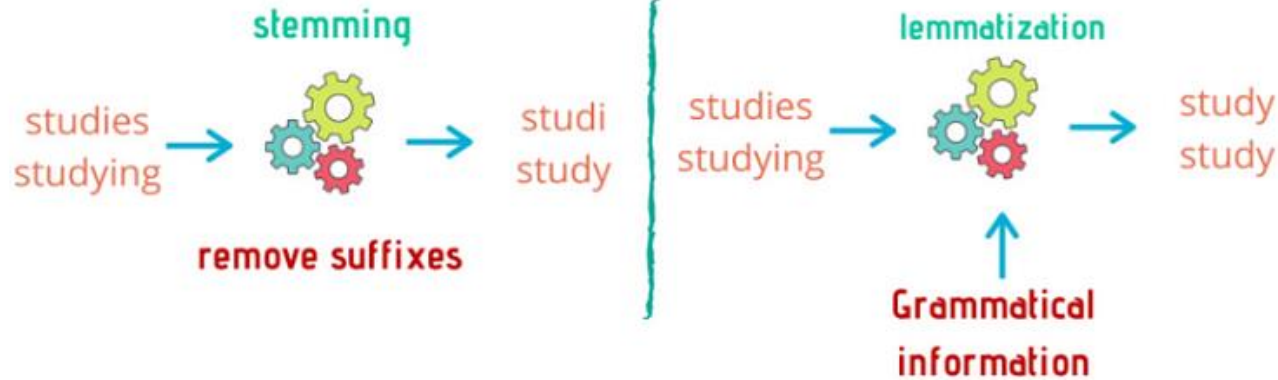


- Introducción a NLP
- Aplicaciones de NLP
- NLP con transformers
- Ejemplo práctico de Q&A usando Hugging Face.

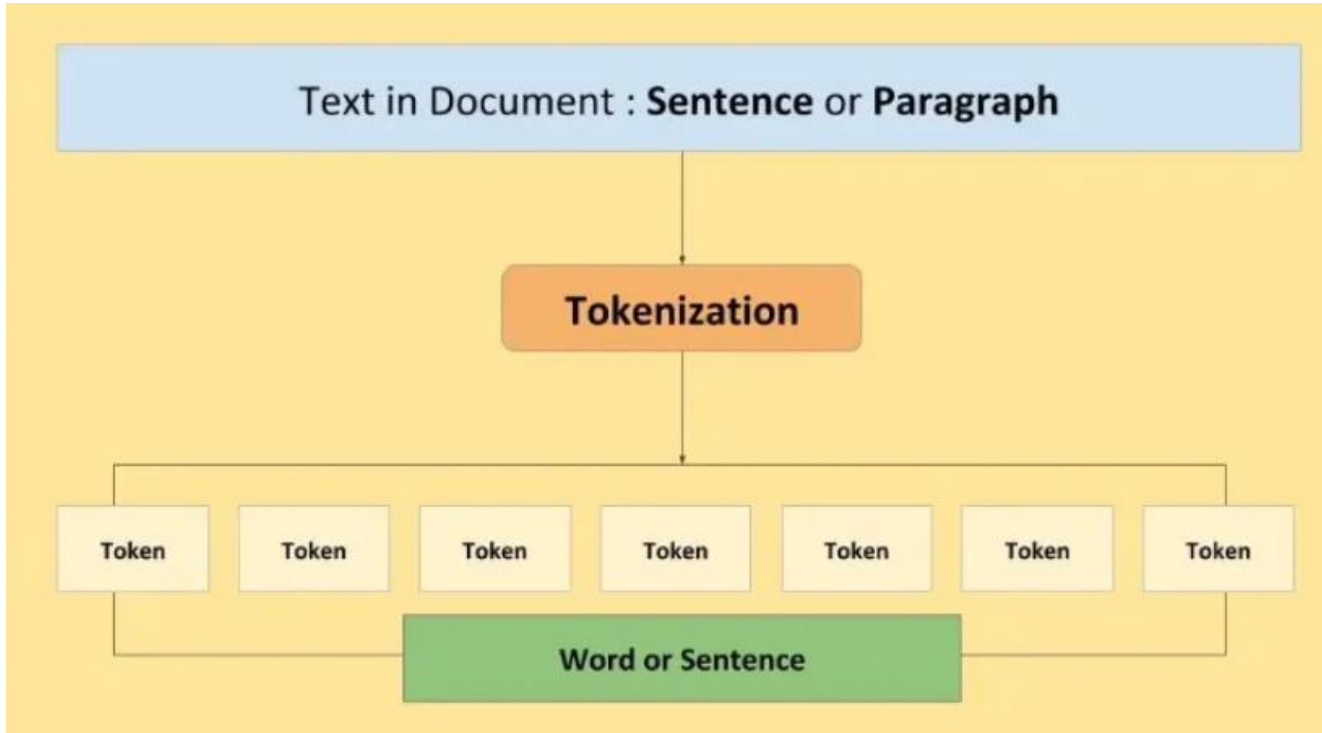
Introducción a NLP



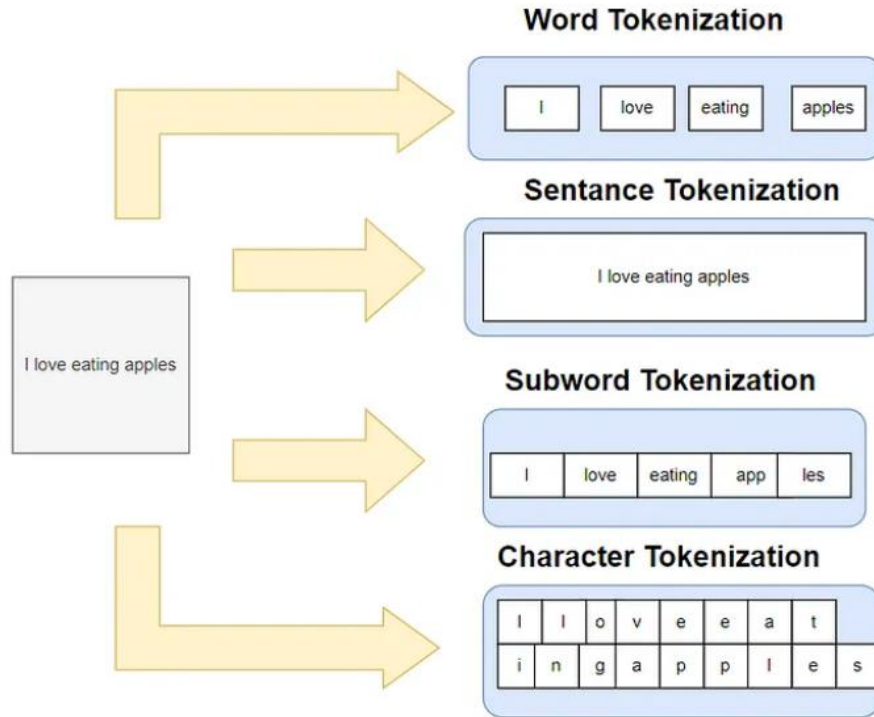
Stemming and Lemmatization



Tokenization

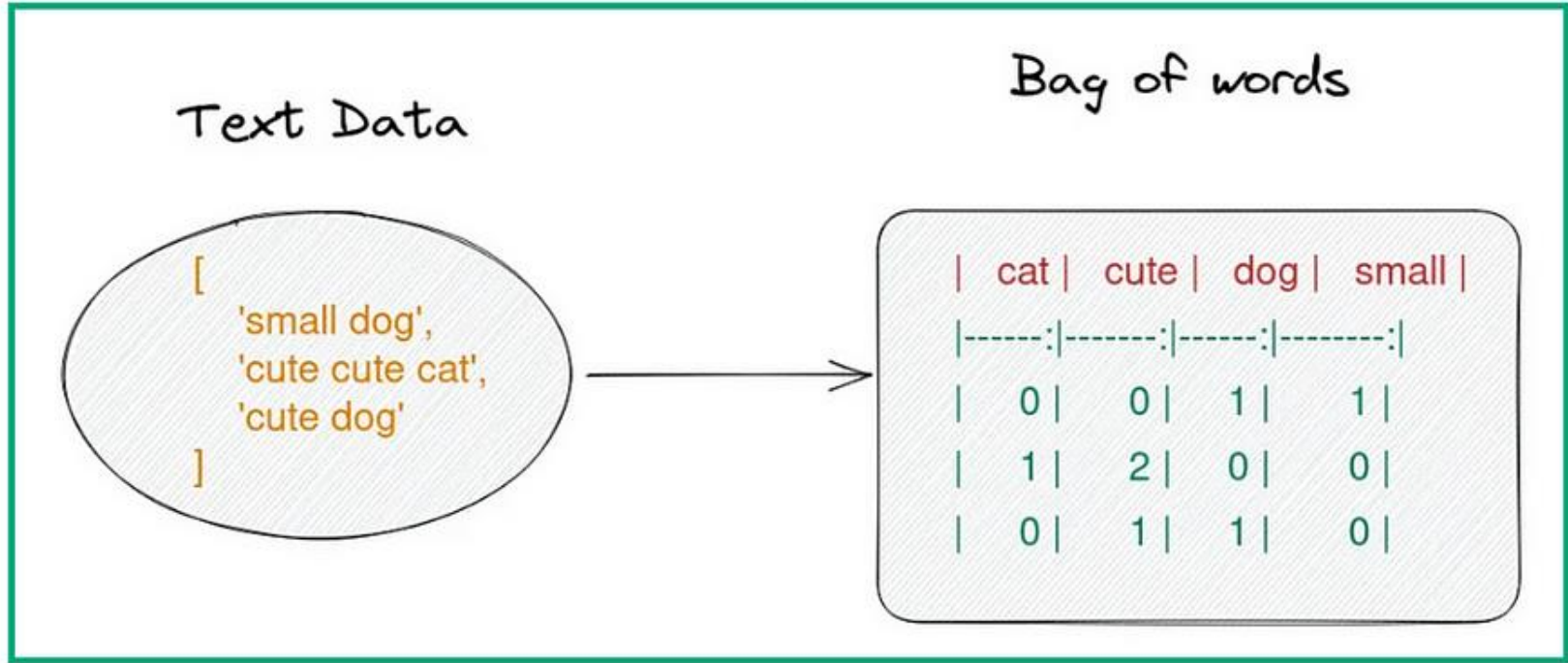


Tokenization

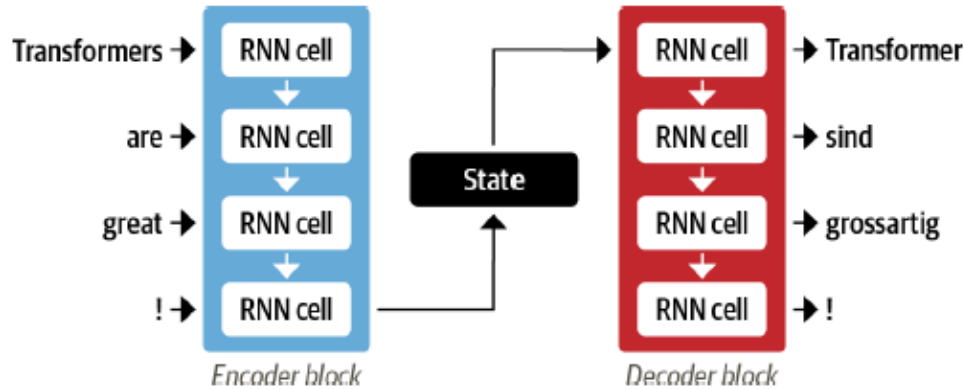
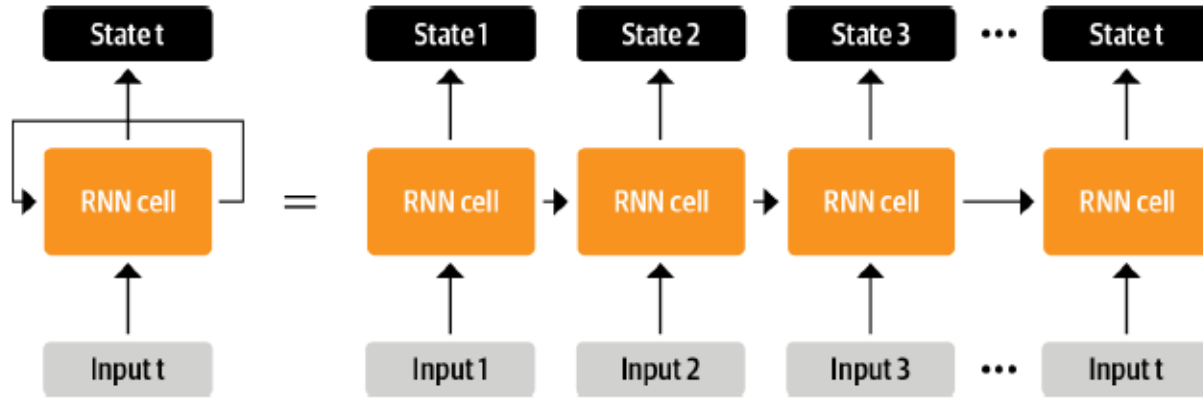


Tokenization techniques

Bag of words



RNN



Hello Transformers

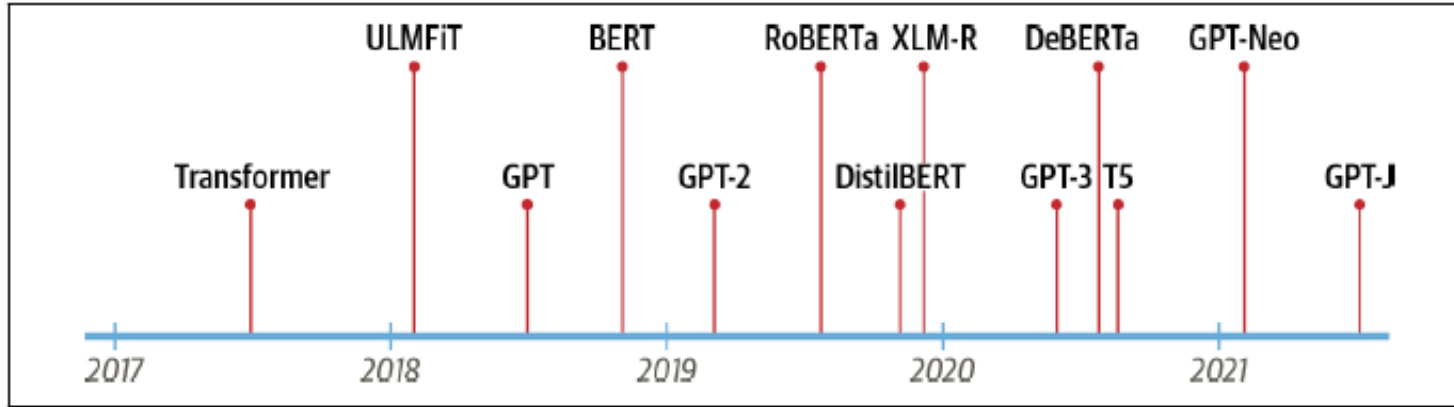
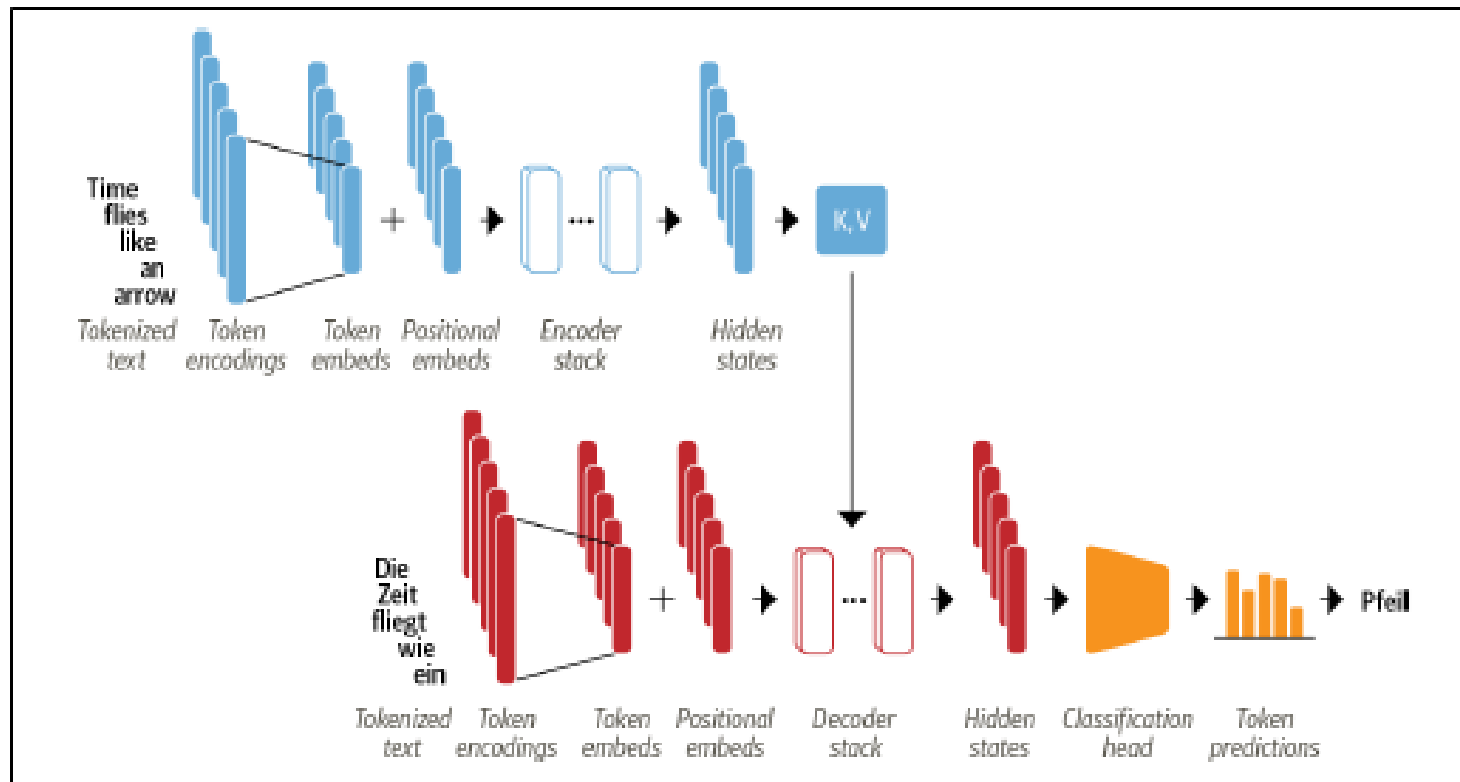
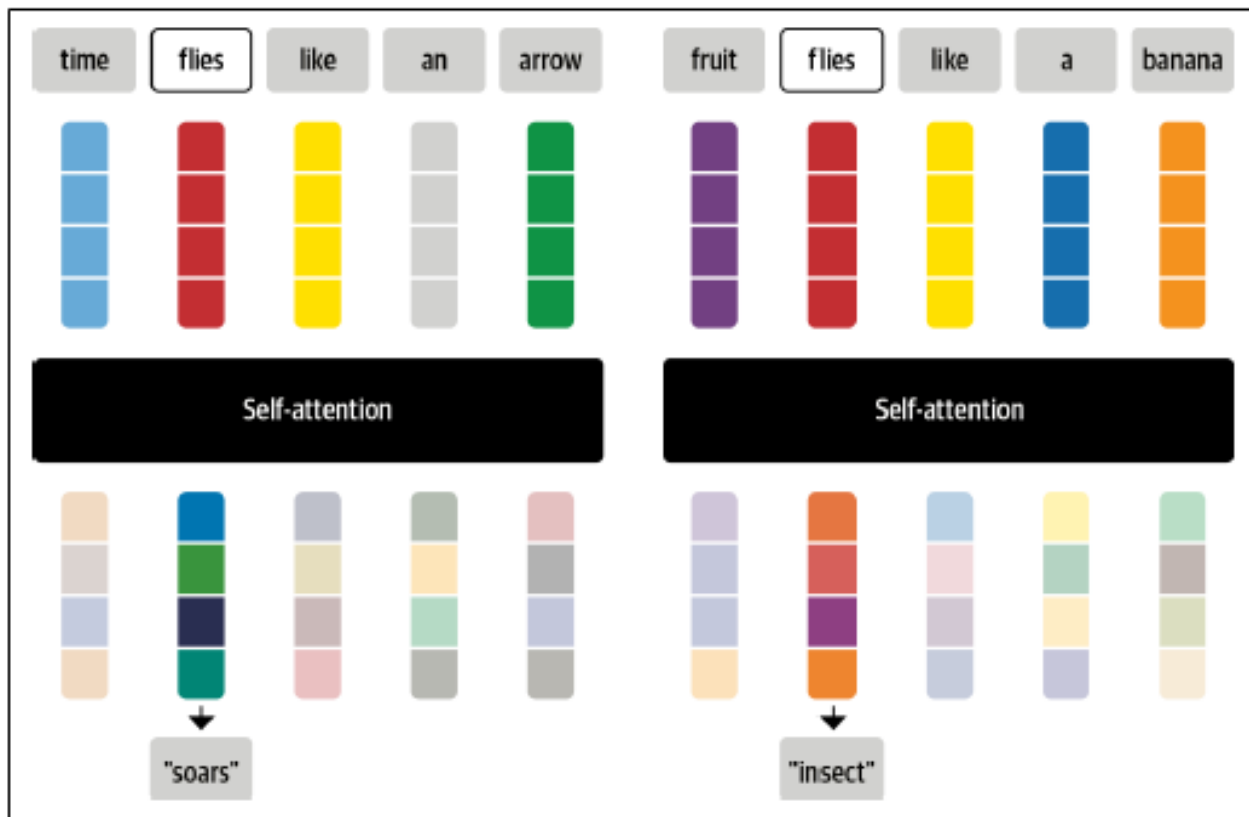


Figure 1-1. The transformers timeline

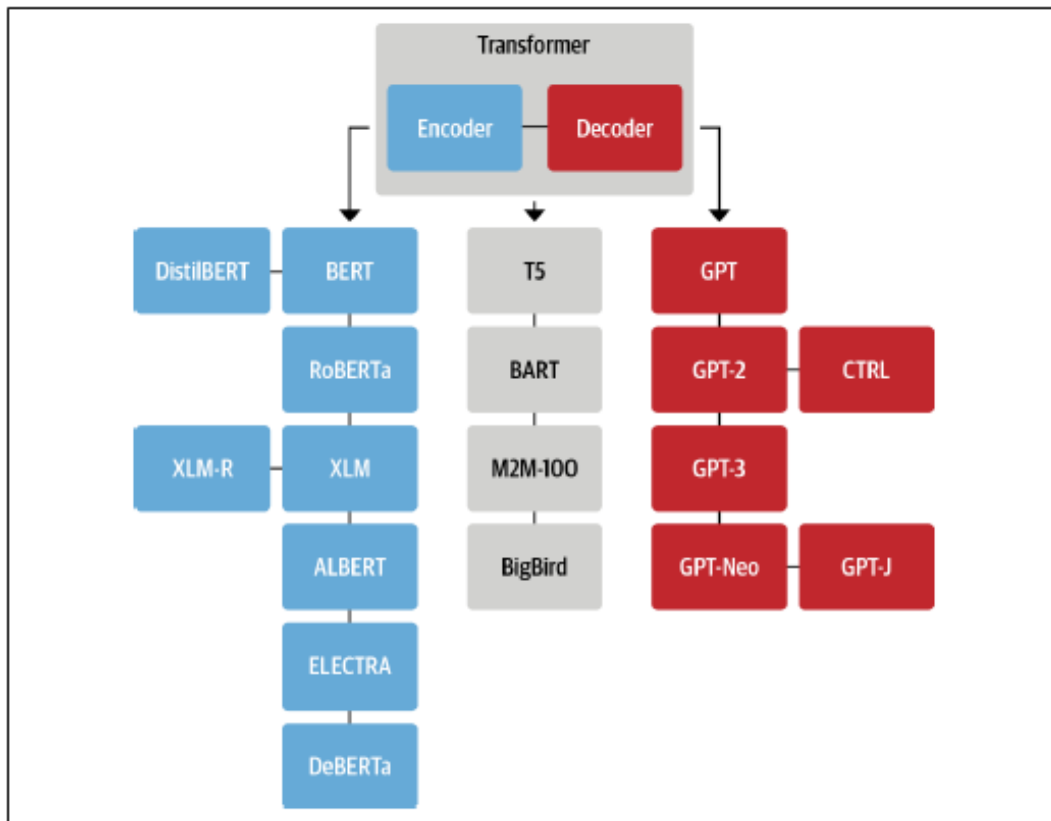
Transformer Architecture



Self-Attention



Algunas Arquitecturas de Transformers



Hugging Face Hub



Hugging Face

Models

Datasets

Spaces

Docs

Solutions

Pricing



Log In

Sign Up

Tasks Libraries Datasets Languages Licenses Other

Multimodal

- Feature Extraction
- Text-to-Image
- Image-to-Text
- Text-to-Video
- Visual Question Answering
- Document Question Answering
- Graph Machine Learning

Computer Vision

- Depth Estimation
- Image Classification
- Object Detection
- Image Segmentation
- Image-to-Image
- Unconditional Image Generation
- Video Classification
- Zero-Shot Image Classification

Natural Language Processing

- Text Classification
- Token Classification
- Table Question Answering
- Question Answering
- Zero-Shot Classification
- Translation
- Summarization
- Conversational
- Text Generation
- Text2Text Generation
- Fill-Mask
- Sentence Similarity

Models 372,969

new Full-text search

Sort: Trending

adept/fuyu-8b

Text Generation • Updated 4 days ago • \pm 14.2k • \heartsuit 508

segmind/SSD-1B

Text-to-Image • Updated about 8 hours ago • \pm 8.62k • \heartsuit 175

mistralai/Mistral-7B-v0.1

Text Generation • Updated 14 days ago • \pm 277k • \heartsuit 1.48k

stabilityai/stable-diffusion-xl-base-1.0

Text-to-Image • Updated 24 days ago • \pm 7.37M • \heartsuit 3.28k

meta-llama/Llama-2-7b-chat-hf

Text Generation • Updated 4 days ago • \pm 975k • \heartsuit 1.55k

CausalLM/7B

Text Generation • Updated 1 day ago • \pm 184 • \heartsuit 74

meta-llama/Llama-2-7b

Text Generation • Updated Jul 19 • \heartsuit 2.86k

SimianLuo/LCM_Dreamshaper_v7

Text-to-Image • Updated 1 day ago • \pm 24.1k • \heartsuit 78

jinaai/jina-embeddings-v2-base-en

Feature Extraction • Updated about 3 hours ago • \pm 4.17k • \heartsuit 183

HuggingFaceH4/zephyr-7b-alpha

Text Generation • Updated about 11 hours ago • \pm 58.6k • \heartsuit 799

CausalLM/14B

Text Generation • Updated about 5 hours ago • \pm 323 • \heartsuit 136

amazon/MistralLite

Text Generation • Updated 3 days ago • \pm 2.94k • \heartsuit 122

mistralai/Mistral-7B-Instruct-v0.1

Text Generation • Updated 15 days ago • \pm 220k • \heartsuit 864

teknium/OpenHeLMs-2-Mistral-7B

Text Generation • Updated about 19 hours ago • \pm 10.1k • \heartsuit 143

SkunkworksAI/BakLLaVA-1

Text Generation • Updated 3 days ago • \pm 523 • \heartsuit 175

runwayml/stable-diffusion-v1-5

Text-to-Image • Updated Aug 23 • \pm 7.65M • \heartsuit 9.47k

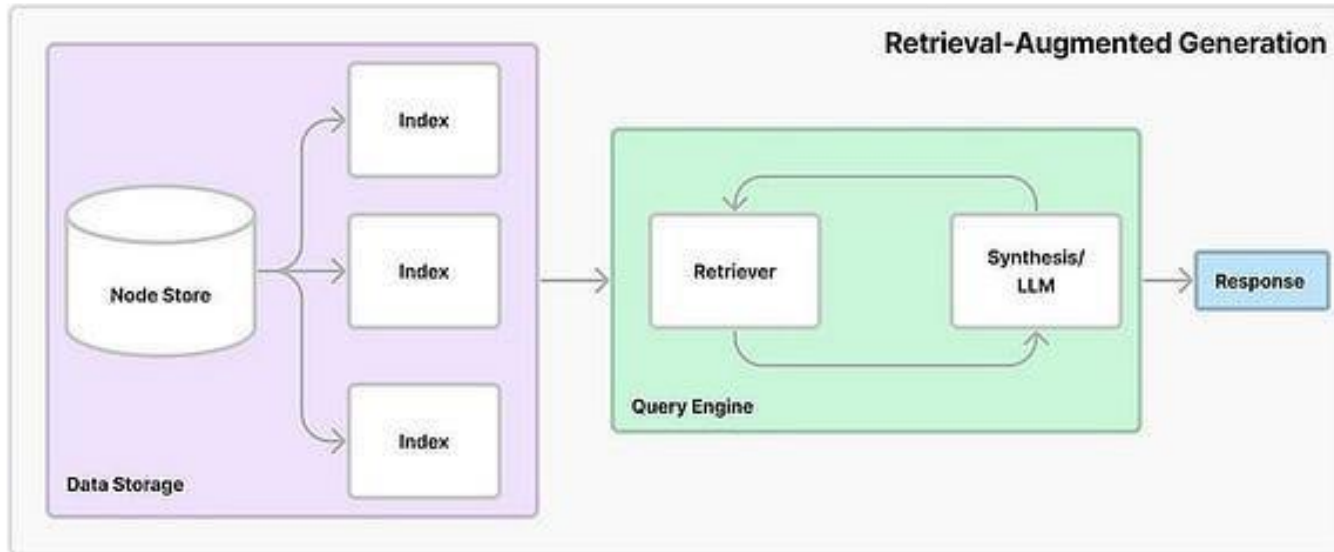
RAG

Retrieval-Augmented Generation (RAG)

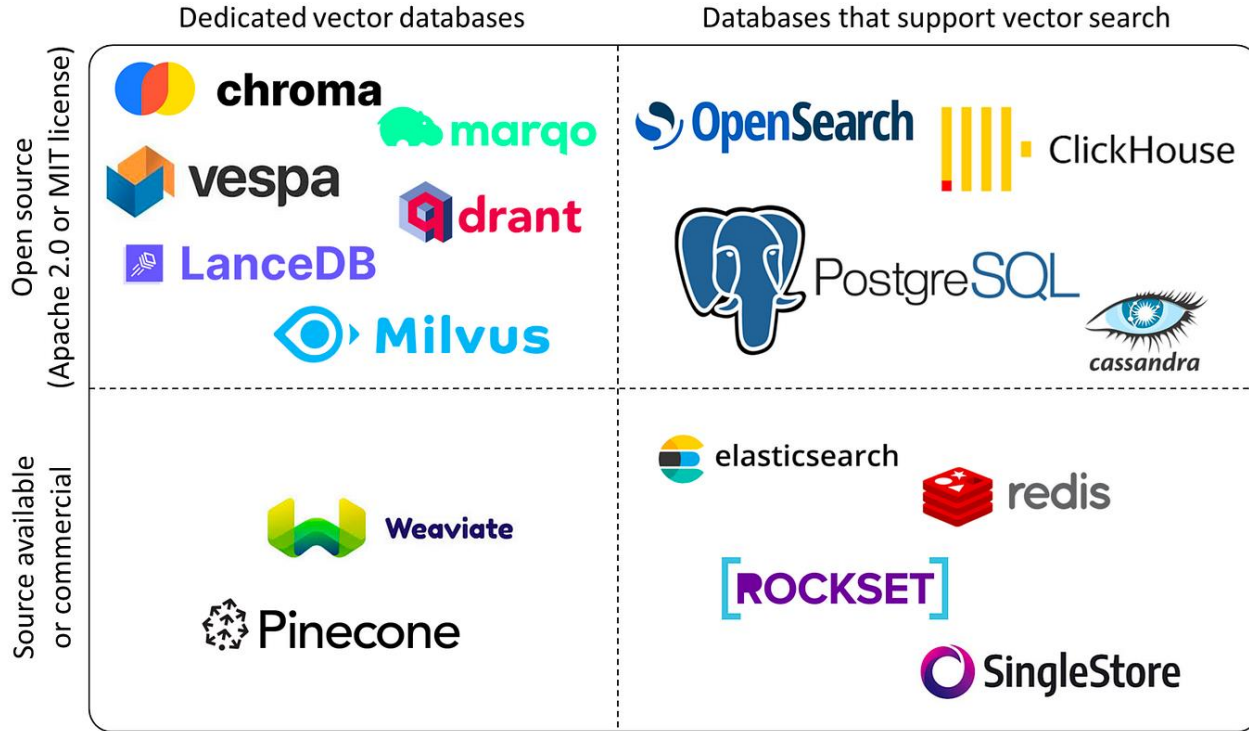
Retrieval-augmented generation (RAG) is a natural language processing (NLP) approach that combines the benefits of both retrieval-based and generation-based methods for content generation tasks. It aims to improve the quality and controllability of the generation tasks by leveraging a pre-trained language model in conjunction with a retrieval mechanism.



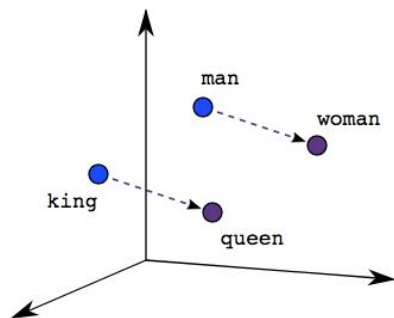
RAG: Componentes



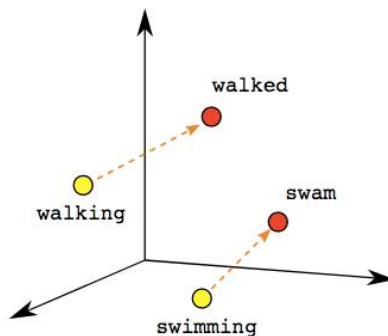
RAG: Database



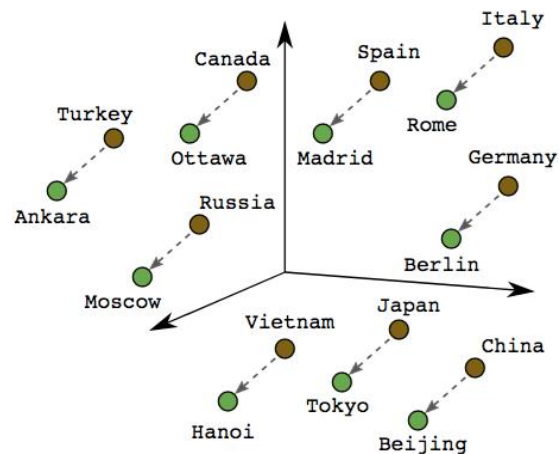
RAG: Embeddings



Male-Female



Verb Tense



Country-Capital



Hugging Face

The task illustrated in this tutorial is supported by the following model architectures:

[ALBERT](#), [BART](#), [BERT](#), [BigBird](#), [BigBird-Pegasus](#), [BLOOM](#), [CamemBERT](#), [CANINE](#), [ConvBERT](#), [Data2VecText](#), [DeBERTa](#), [DeBERTa-v2](#), [DistilBERT](#), [ELECTRA](#), [ERNIE](#), [ErnieM](#), [Falcon](#), [FlauBERT](#), [FNet](#), [Funnel Transformer](#), [OpenAI GPT-2](#), [GPT Neo](#), [GPT NeoX](#), [GPT-J](#), [I-BERT](#), [LayoutLMv2](#), [LayoutLMv3](#), [LED](#), [LiLT](#), [Longformer](#), [LUKE](#), [LXMERT](#), [MarkupLM](#), [mBART](#), [MEGA](#), [Megatron-BERT](#), [MobileBERT](#), [MPNet](#), [MPT](#), [MRA](#), [MT5](#), [MVP](#), [Nezha](#), [Nyströmformer](#), [OPT](#), [QDQBert](#), [Reformer](#), [RemBERT](#), [RoBERTa](#), [RoBERTa-PreLayerNorm](#), [RoCBert](#), [RoFormer](#), [Splinter](#), [SqueezeBERT](#), [T5](#), [UMT5](#), [XLM](#), [XLM-RoBERTa](#), [XLM-RoBERTa-XL](#), [XLNet](#), [X-MOD](#), [YOSO](#)