COLOMBIA POTENCIA DE LA VIDA

Socialize key words about

Natural Language Processing (NLP)

TIC

Tokenization:

is the process of breaking down a text into smaller units called tokens. These tokens can be words, phrases, or sentences, depending on the level of granularity needed.

Stemming:

is a technique used to reduce words to their base or root form. It involves removing suffixes or prefixes from words, so variations of a word are represented by the same stem.

Part-of-Speech Tagging:

involves assigning a specific part-of-speech (such as noun, verb, adjective, etc.) to each word in a sentence.

Word Embeddings:

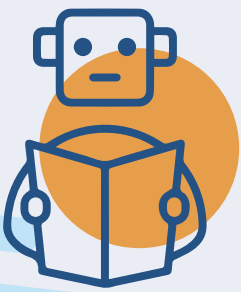
are dense vector representations of words in a continuous vector space. Each word is assigned a unique vector, and the distance and direction between these vectors capture semantic relationships between words.

Recurrent Neural Networks (RNNs):

are a type of artificial neural network designed for sequential data processing. Unlike traditional feedforward networks, RNNs have connections that create loops, allowing them to maintain a memory of previous inputs.

Deep Learning:

a subset of machine learning that involves neural networks with multiple layers (deep neural networks).



Ambiguity in NLP:

the presence of multiple possible interpretations or meanings in a given language context.

Bias in NLP:

systematic and unfair preferences or prejudices in language processing models.