

 04

Tokenização: crucial ao PLN

Júlia teve contato com uma pequena parcela da biblioteca **NLTK** e achou incrível. Gostou tanto que resolveu pesquisar mais sobre o processo de **tokenização**.

Qual das alternativas **descreve** corretamente o **processo de tokenização**?

Seleciona uma alternativa

A A tokenização é o processo de receber uma frase separada juntando suas palavras ou pontuação de acordo com um tipo de unificador específico, gerando os chamados tokens, que são as frases separadas por ponto final. Nesta aula usamos como unificador o espaço em branco, mas temos outras formas de unificar nosso texto.

B A tokenização é o processo de receber um texto e separar suas palavras ou pontuação de acordo com um tipo de separação específica, gerando os chamados tokens, que são as partes segregadas daquele texto. Nesta aula usamos como separador o espaço em branco, que é a única forma que temos para separar nosso texto.

C A tokenização é o processo de receber um texto e separar suas palavras ou pontuação de acordo com um tipo de separação específica, gerando os chamados tokens, que são as partes segregadas daquele texto. Nesta aula usamos como separador o espaço em branco, mas temos outras formas de separar nosso texto.