



18%

ATIVIDADES
5 de 10

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



16.1k xp



05

Contando palavras do Corpus

Transcrição

Para contarmos apenas as palavras presentes em nosso corpus, criaremos a variável

```
lista_palavras .
```

Em seguida, chamaremos a função `separa_palavras()` recebendo `lista_tokens` , porém também precisaremos criar esta última da mesma maneira que fizemos para a frase de exemplo, utilizando `nltk.tokenize.word_tokenize()` recebendo `artigos` .

Com isso, poderemos aplicar `len()` com a `lista_palavras` para sabermos a quantidade de palavras.

Por fim, imprimiremos a frase "O número de palavras é {}" após o `f` que faz a interpolação de strings , seguido do método `len()` entre as chaves.

```
lista_tokens = nltk.tokenize.word_tokenize(artigos)
lista_palavras = separa_palavras(lista_tokens)
print(f"O número de palavras é {len(lista_palavras)}")
```

COPIAR CÓDIGO



18%

ATIVIDADES

5 de 10

DISCORD

ALURA

FÓRUM DO

CURSO

VOLTAR

PARA

DASHBOARD

**MODO
NOTURNO**



16.1k xp

a

Feita a execução, teremos o número total de 393916 palavras no corpus, o que é um valor bastante considerável.