



21%

ATIVIDADES
7 de 10

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



16.1k xp

a

07

Tipos de palavras

Transcrição

Agora que já temos nossa lista normalizada, poderemos **remover as repetições** de palavras, e existem algumas formas de fazer isso.

Poderemos tanto construir um algoritmo, quanto trabalhar com ideias mais interessantes no Python, pois este tem métodos que implementam a ideia de **conjuntos** matemáticos, como o `set()`.

O conjunto dos números naturais é "0, 1, 2, 3, 4" e assim por diante, e obviamente não possui repetições.

O `set()` pegará uma lista com o exemplo `[1, 2, 3, 3, 3, 3, 4, 5, 6]`, o qual possui valores repetidos. O retorno será o conjunto dessa lista sem nenhuma repetição.

```
set([1, 2, 3, 3, 3, 3, 4, 5, 6])
```

COPIAR CÓDIGO

Rodando o comando, receberemos o conjunto `{1, 2, 3, 4, 5, 6}`, e veremos que as repetições foram eliminadas com o uso deste método.



21%

ATIVIDADES
7 de 10

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODOS
NOTURNOS



16.1k xp

a

Com isso, passaremos a nossa `lista_normalizada` para o `set()` na célula seguinte, removendo as repetições.

Para descobrirmos o tamanho deste conjunto, aplicaremos o `len()`.

```
len(set(lista_normalizada))
```

COPIAR CÓDIGO

Executando o código, obteremos a quantidade de 17654 de tipos diferentes, o que é um valor bastante interessante.

Este será o número de palavras que nosso corretor ortográfico aprenderá a corrigir, lembrando que temos um total de 393.916 vocábulos no corpus.

Segundo [este link com um artigo \(https://forum.duolingo.com/comment/7642297/Quantas-palavras-voc%C3%AA-precisa-saber-para-ser-fluente\)](https://forum.duolingo.com/comment/7642297/Quantas-palavras-voc%C3%AA-precisa-saber-para-ser-fluente) do **Duolingo**, é necessário saber cerca de 800 a 1.000 palavras de um idioma para estar em nível básico, 1.500 a 2.000 para o intermediário, e de 3.000 a 4.000 vocábulos para ter fluência.

Portanto, como nosso algoritmo aprenderá a corrigir mais de **17.000** questões, é bem mais do que o suficiente para construirmos um bom corretor ortográfico, sem levar em consideração as variações entre as palavras e outros detalhes que veremos adiante.