



29%

ATIVIDADES
2 de 10

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



16.3k xp



02

Detalhando o corretor

Transcrição

Relembrando como nosso corretor ortográfico funcionará, queremos que este receba o termo "lgica" escrito erroneamente e que devolva-o corrigido para "lógica".

Descobrimos a palavra correta porque está entre os mais de 17.000 vocábulos conhecidos pela contagem feita anteriormente.

Porém, ainda não implementamos a transição de “lgica” para “lógica” de fato.

Pegaremos a palavra digitada errada e geraremos novas palavras, onde uma delas poderá ser a correta. Neste nosso problema específico, temos a letra "ó" a menos.

Nosso modelo de NLP precisará de um algoritmo capaz de **inserir uma letra** a mais neste termo digitado equivocadamente.

No Processamento de Linguagem Natural, temos a entrada da palavra equivocada e a saída da correta. Entre esses pontos, elaboramos uma tabela com três colunas representando os passos da correção deste algoritmo que desenvolveremos.



29%

ATIVIDADES
2 de 10

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



16.3k xp

a

A primeira coluna possui as letras que poderão ser inseridas, ou seja, todas as que existem no alfabeto e as vogais com os acentos possíveis.

A segunda coluna representa a **operação de inserção** que realizaremos, e a terceira possui o resultado com as candidatas de possíveis palavras para correção.

Como primeira operação, dividiremos as `strings` em **duas partes** e inseriremos uma letra entre elas. No caso de "lgica", um lado será a letra "l", e o outro lado será o resto da minha palavra, "gica".

Para inserirmos a letra, a operação será a primeira parte "l" somada com a letra que queremos inserir. Na primeira linha da primeira coluna desta tabela, temos a letra "a", resultando na palavra candidata a correção "lagica" na terceira coluna.

A mesma coisa acontecerá com a letra "b" da segunda linha da primeira coluna, resultando em "lbgica".

Quando chegarmos em "ó" com o acento correto, teremos a correção com o termo "lógica" que queremos, afinal os outros resultados anteriores da terceira coluna formam palavras que não existem e não são conhecidas a partir do corpus.

LETRAS	OPERAÇÃO	RESULTADO
a	"l" + a + "gica"	lagica
b	"l" + b + "gica"	lbgica
...
ó	"l" + ó + "gica"	lógica



29%

ATIVIDADES
2 de 10

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



16.3k xp

a

Porém, o algoritmo só está resolvendo o caso da palavra “lógica”, onde apenas adicionamos uma letra entre “l” e “gica”.

Então faremos a generalização desta operação em nosso algoritmo, pegando a primeira parte da `string` chamada de lado **esquerdo**, adicionando uma nova **letra**, e concatenando com o restante da palavra no lado **direito** como já sabemos:

esquerdo (primeira parte) +** **letra** **+ direito (segunda parte)

Com a operação generalizada, deveremos pensar em como geraremos a separação da palavra em duas partes para recebermos a nova letra no meio, afinal teremos situações diferentes no dia-a-dia.

Elaboraremos uma nova tabela com quatro colunas: o lado **esquerdo**, **direito**, **operação** e **resultado**. Como título, temos a palavra errada "lgica" para realizarmos a divisão e correção neste caso.

Começaremos dividindo entre a primeira parte e segunda parte, e pegaremos a primeira posição `0` da `string` em questão, ou seja, antes da primeira letra "l" da palavra. Portanto, a primeira linha do lado esquerdo será `null`, pois não possui nada.

Para este mesmo ponto do `index`, o lado direito será a palavra toda. Em seguida, faremos a mesma a operação generalizada, resultado em “ológica”.

Este tipo de operação serve para quando a primeira letra não for digitada, como no caso de termos escrito apenas "ógica" por exemplo, onde a inserção de "l" no início seria necessária.

Prosseguindo para a divisão do caractere `1` da `string`, o lado esquerdo será a letra “l”, o lado direito será “gica”, e a operação resultará na palavra correta "lógica".



29%

ATIVIDADES
2 de 10

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



16.3k xp

a

Em um terceiro caso, a primeira parte será o “lg” e a segunda será “ica”, então o resultado da operação ficará "lgóica", que também não existe no corpus. E assim sucessivamente ao longo de toda a palavra até a última posição.

LGICA

ESQUERDO	DIREITO	OPERAÇÃO	RESULTADO
null	lgica	null + ó + lgica	óligca
l	gica	l + ó + gica	lógica
lg	ica	lg + ó + ica	lgóica
lgi	ca	lgi + ó + ca	lgióca
lgic	a	lgic + ó + a	lgicóa
lgica	null	lgica + ó + null	lgicaó

No caso da última posição, a operação resolve o problema da ausência da última letra da palavra, como "lógic" por exemplo.

Ao final, as palavras da quarta coluna serão as possíveis candidatas à correção. Consultando nosso corpus textual, saberemos que a palavra correta é “lógica”.

A seguir, transformaremos o algoritmo pensado através da tabela em código.