

PROCESSAMENTO DE LINGUAGEM NATURAL

Prof^a. MSc. Elene Ohata



Novembro/2023

ELENE OHATA

- Engenharia de Mecatrônica - IFCE
- Mestrado em Engenharia de Teleinformática - UFC
- Doutorado em Engenharia de Teleinformática - UFC - em andamento
- Membro do LAPISCO

TÓPICOS

- O que é PLN?
- Aplicações
- Python para textos
- Expressões regulares
- Tokenização
- Stemming e Lemmatization
- Stop words



O QUE É PLN/NLP?

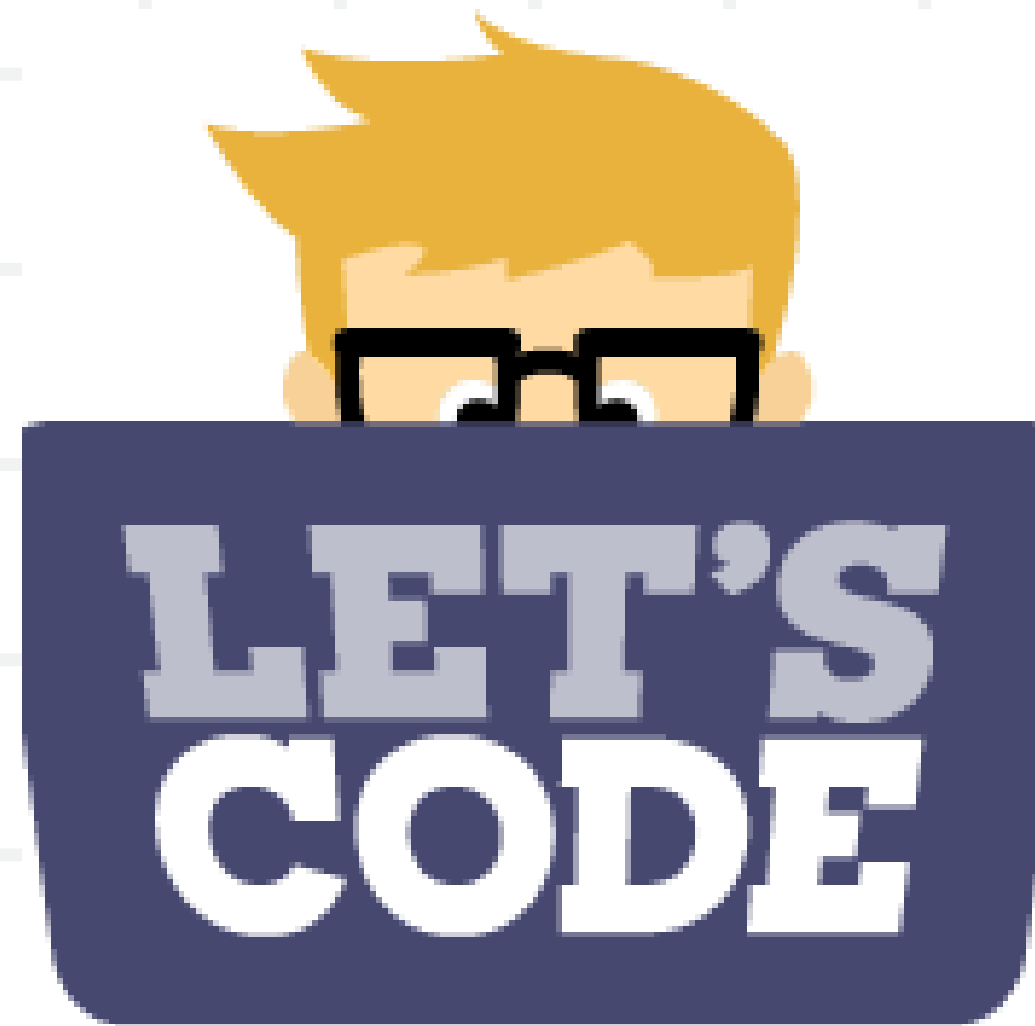
- Natural Language Processing (NLP) ou Processamento de Linguagem Natural (PLN) é a tecnologia usada para ajudar dispositivos tecnológicos a entenderem a linguagem do ser humano de maneira a responder suas demandas.
- É um ramo da inteligência artificial que ajuda os computadores a entender, interpretar e manipular a linguagem humana.

APLICAÇÕES

- Chatbots e Assistentes Virtuais
- Tradução automática
- Análise de sentimentos
- Resumo automático
- Extração de informações
- Classificação de textos



PYTHON BÁSICO PARA TEXTOS



EXPRESSÕES REGULARES

Uma expressão regular ou regex é uma sequência de caracteres que define um padrão de pesquisa específico e com o qual você pode combinar ou substituir padrões dentro de um texto com a menor quantidade de código.



EXPRESSÕES REGULARES

O Google foi fundado em 1998.

Como apagar o número?

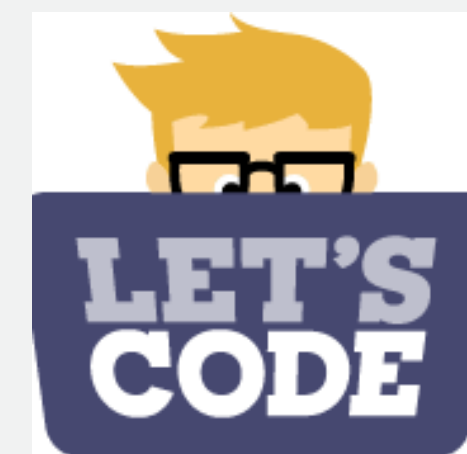
- Loop pela frase
- Procurar por um número
- Apagar o número



EXPRESSÕES REGULARES

O Google foi fundado em 1998.

```
re.sub(r"\d", "", sentence)
```



TOKENIZAÇÃO

Tokenização é o processo de dividir o texto em componentes (tokens).



TOKENIZAÇÃO



Let's tokenize! Isn't this easy?

Por pontuação:

Let ' s tokenize ! Isn ' t this easy ?

Por espaços:

Let's tokenize! Isn't this easy?

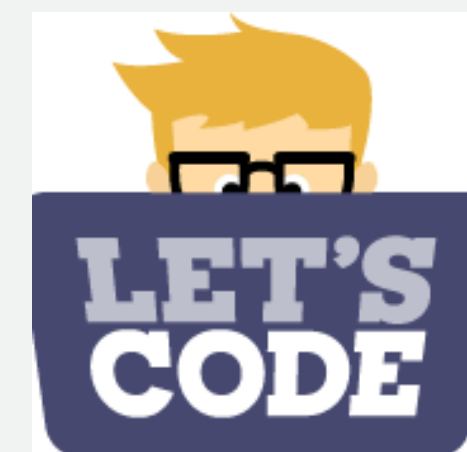
Por regras:

Let 's tokenize ! Is n't this easy ?



STOP WORDS

- São palavras que podem ser consideradas irrelevantes.
- Normalmente, possuem uma frequência muito grande.
- Exemplos: as, os, uns, de, para, com, por.



STEMMING

- Cada palavra é individualmente reduzida à sua raiz (radical) ou, como é chamado na técnica, ao seu stem.
- Algoritmos têm um conjunto de regras para decidirem como fazer os cortes até chegar à raiz da palavra.
- A redução pode gerar uma palavra gramaticalmente incorreta.
- Exemplo: gato, gata, gatos, gatas -> “gat”
- Vantagem: Reduz a quantidade de palavras únicas

LEMMATIZATION

- Redução de uma palavra à sua forma básica ou de dicionário, conhecida como lemma.
- Considera o significado da palavra, sua classe gramatical e a análise morfológica.
- Exemplo: tiver, tenho, tinha, tem -> “ter”





OBRIGADA!

elene.ohata@lapisco.ifce.edu.br

