



Processamento de Linguagem Natural

UNIDADE 02

A base do processamento de textos

Olá, estudante!

Nesta semana, vamos iniciar nossos trabalhos com os processos e ferramentas que formam a base de desenvolvimento de algoritmos de PLN.

Vamos também explorar o mundo das expressões regulares, que nos possibilitam manipular dados ao encontrar padrões textuais.

Vamos lá?

| Quais processos e ferramentas formam a espinha dorsal do PLN?

Quando desenvolvemos algoritmos de processamento de texto, há certos processos e ferramentas que se repetem na maioria das tarefas que desejamos realizar. Na Unidade de Aprendizagem **A base do processamento de textos**, disponível na semana, vamos conhecer esses processos, bem como visualizar algumas demonstrações práticas de execução em Python.

| Como podemos encontrar padrões no texto com as expressões regulares?

Além dos processos apresentados anteriormente, outra ferramenta muito comum no processamento de textos são as expressões regulares, que permitem que o algoritmo busque por padrões em meio ao texto, podendo apoiar diversas tarefas de PLN.

Na Unidade de Aprendizagem **Expressões regulares**, disponível na semana, você verá como lidar com as expressões regulares na linguagem de programação Python, apesar de a base teórica se repetir em diferentes linguagens de programação; portanto, o que você aprender aqui poderá ser replicado em outros ambientes de desenvolvimento.

Agora, vamos trabalhar de forma mais prática com as expressões regulares. No vídeo a seguir, você encontrará um tutorial prático que ilustra diversas funcionalidades das expressões regulares. Além do vídeo, você pode também acessar o *notebook*, que contém algumas atividades propostas ao final.

Neste vídeotutorial, vamos explorar diversas funcionalidades das expressões regulares em Python.

Expressões regulares em Python



Notebook – Google Colab

Explore o *notebook* a seguir, que demonstra a utilização de expressões regulares.

<https://colab.research.google.com/drive/1zrBgCG06FR2jugQiKwv2M2LWJWJN3uY->

Nesta unidade, você aprendeu a:

- Identificar métodos de PLN.

| Conclusão

- Descrever processos de pré-processamento de textos.
- Definir o papel das expressões regulares como recurso para o PLN.
- Reconhecer as diferentes funções disponíveis para utilizar expressões regulares.
- Programar algoritmos de PLN empregando expressões regulares.

| Referências

BIRD, S.; KLEIN, E.; LOPER, E. **Natural language processing with Python: analyzing text with the natural language toolkit**. 2009. Disponível em: <https://www.nltk.org/book/>. Acesso em: 4 maio 2020.

JARGAS, A. M. **Expressões regulares: uma abordagem divertida**. São Paulo: Novatec, 2016.

JURAFSKY, D.; MARTIN, J. H. **Speech and language processing**. Disponível em: <https://web.stanford.edu/~jurafsky/slp3/>. Acesso em: 15 dez. 2020.

PYTHON SOFTWARE FOUNDATION. **Re-regular expression operations**. 2020. Disponível em: <https://docs.python.org/3/library/re.html>. Acesso em: 4 maio 2020.

