



escola
britânica de
artes criativas
& tecnologia

Profissão: Analista de dados



GLOSSÁRIO



Coleta de dados II



Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

- **Captando dados da web (crawling)**
- **Extraindo dados da web (scrapping)**
- **Consumindo dados de API públicas**



Captando dados da web (crawling)



Captando dados da web (crawling)

Glossário

Browser

É um *software* utilizado para visualizar e interagir com páginas da web. Também conhecido como navegador, ele permite que os usuários acessem e naveguem em *sites* na internet.

Códigos de retornos

São códigos numéricos que indicam o resultado de uma solicitação HTTP entre o cliente e o servidor. Eles indicam se a solicitação foi bem sucedida ou se houve algum problema. Alguns exemplos de códigos de retorno incluem 200 (OK), 404 (Não encontrado) e 500 (Erro interno do servidor).



Captando dados da web (crawling)

Glossário

● Cliente servidor

É um modelo de comunicação em que um computador (o cliente) solicita e recebe dados de outro computador (o servidor), geralmente através da internet. O cliente envia uma solicitação ao servidor e aguarda a resposta, que pode ser um arquivo, uma página da web, entre outros dados.

● For loop

É uma estrutura de controle de fluxo em programação que permite executar um bloco de código repetidamente enquanto uma condição é verdadeira, iterando sobre uma sequência de elementos.



Captando dados da web (crawling)

Glossário

● HTTP

É um protocolo de comunicação utilizado para transferir dados pela internet. Ele define as regras para como os dados devem ser formatados e transmitidos entre o cliente e o servidor.

● Método get

É um dos métodos HTTP utilizados para solicitar dados de um servidor. Ele envia uma solicitação para recuperar dados especificados em uma URL.

● Método post

É um dos métodos HTTP utilizados para enviar dados para um servidor. Ele é geralmente usado para enviar informações de formulário ou fazer *upload* de arquivos.

● Servidor

É um computador ou sistema que fornece serviços ou recursos para outros computadores ou sistemas em uma rede. Na internet, um servidor é frequentemente usado para hospedar sites, fornecer serviços de e-mail, armazenar arquivos, entre outras funções.



Captando dados da web (crawling)

Glossário

Web crawling

É uma técnica de mineração de dados em que um software automatizado, conhecido como "rastreador" ou "crawler", percorre a web de forma sistemática e coleta informações sobre sites e páginas da web.



Extraíndo dados da web (scrapping)



Extraindo dados da web (scrapping)

Glossário

● BeautifulSoup

É uma biblioteca em Python utilizada para extrair dados de arquivos HTML e XML. Ela fornece uma interface simples e intuitiva para analisar e manipular o conteúdo de páginas web, permitindo a extração de informações de forma automatizada.

● Objeto funcional

É um conceito em programação funcional que se refere a uma função que é passada como argumento para outra função, ou que retorna uma outra função como resultado. Essas funções são usadas para construir programas a partir de pequenas peças reutilizáveis, facilitando a modularidade e a manutenção do código. Em resumo, um objeto funcional é uma função que é tratada como um valor, podendo ser usada de forma flexível e dinâmica em um programa.



Consumindo dados de API públicas



Consumindo dados de API públicas

Glossário

• **Anaconda**

É uma distribuição de Python que inclui diversas bibliotecas e ferramentas para análise de dados e ciência de dados.

• **API RESTful**

É uma API baseada em HTTP que utiliza os verbos HTTP (GET, POST, PUT, DELETE) para permitir a comunicação entre sistemas.

• **Autenticação**

Processo de verificação da identidade do usuário, utilizado para proteger sistemas e serviços.

• **Bot (abreviação de robot)**

É um programa de computador que realiza tarefas automatizadas na web, como enviar mensagens, coletar dados ou realizar ações em sites. Podem ser programados para realizar tarefas simples ou complexas e podem ser usados para uma variedade de finalidades, como marketing, monitoramento de preços, análise de dados e atendimento ao cliente.



Consumindo dados de API públicas

Glossário

• Dicionário Python

É uma estrutura de dados que mapeia chaves a valores, permitindo a recuperação rápida de um valor a partir de sua chave.

• Django

É um *framework web* de alto nível para Python que é conhecido por sua escalabilidade e robustez. É usado para criar aplicativos de web complexos e de grande escala, com recursos como autenticação de usuário, manipulação de formulários e administração de banco de dados.

• Flask

É um framework web leve e flexível para Python. Ele permite que os desenvolvedores criem aplicativos web de forma rápida e fácil usando uma sintaxe simples e elegante. Flask é popular entre os desenvolvedores web devido à sua simplicidade e flexibilidade.



Consumindo dados de API públicas

Glossário

• **HTTPError**

É uma exceção que pode ocorrer quando se tenta acessar um recurso na web usando o protocolo HTTP. Ela indica que houve um erro na comunicação entre o cliente (geralmente um navegador da web) e o servidor (onde o recurso está hospedado). O erro pode ser causado por diversos motivos, como a página solicitada não existir, o servidor estar fora do ar, ou o cliente não ter permissão para acessar o recurso.

• **IDE (Integrated Development Environment)**

Ambiente de desenvolvimento integrado utilizado para escrever, depurar e testar programas Python.




• **Importação**

Processo de incluir um módulo ou pacote em um programa Python para utilização de suas funções e classes.



Consumindo dados de API públicas

Glossário

- 
Interpretador
 É um programa que executa o código fonte de um programa diretamente, sem a necessidade de compilar o código em um arquivo executável separado. Em Python, o interpretador é responsável por executar o código Python linha por linha.
- 
Jupyter Notebook
 É um ambiente interativo de desenvolvimento web para criar e compartilhar documentos que contenham código, texto, imagens e gráficos.
- 
JSON
 Um formato de dados leve e fácil de ler e escrever, comumente usado para representar objetos em APIs e intercâmbio de dados.



Consumindo dados de API públicas

Glossário

● Keras

É uma biblioteca de aprendizado de máquina de código aberto para Python que é conhecida por sua simplicidade e facilidade de uso. Ele é usado para criar modelos de aprendizado profundo e é popular entre os desenvolvedores por sua sintaxe simples e elegante.

● OpenCV

É uma biblioteca de visão computacional para Python que é usada para processar e analisar imagens e vídeos. Ela oferece recursos avançados como reconhecimento de rosto, rastreamento de objetos e detecção de movimento.



Consumindo dados de API públicas

Glossário

● PIP (Python Package Index)

É o gerenciador de pacotes padrão para Python. Ele permite que os usuários instalem e gerenciem pacotes de terceiros que adicionam funcionalidade ao Python. O Python Package Index é um repositório de pacotes Python que podem ser instalados usando o PIP.

● Pygame

É uma biblioteca Python que fornece um conjunto de módulos projetados para o desenvolvimento de jogos 2D e 3D. Ela permite que desenvolvedores criem jogos com gráficos, som, manipulação de entrada e recursos de rede. O Pygame é construído sobre a biblioteca SDL (Simple DirectMedia Layer), que fornece acesso de baixo nível a gráficos, áudio e dispositivos de entrada.



Consumindo dados de API públicas

Glossário

PyTorch

É uma biblioteca de aprendizado de máquina de código aberto para Python que é conhecida por sua flexibilidade e facilidade de uso. Ele é usado para criar modelos de aprendizado profundo e é popular entre os desenvolvedores devido à sua sintaxe simples e elegante.

SciPy

É uma biblioteca para Python que oferece suporte para cálculo científico e matemático avançado. Ela inclui funções para otimização, álgebra linear, integração numérica, estatística e outras áreas de matemática e ciência.



Consumindo dados de API públicas

Glossário

● **Scrapy**

É um framework de web crawling e scraping escrito em Python. Ele fornece uma estrutura para extrair dados de sites da web de forma programática e automatizada. Com o Scrapy, os desenvolvedores podem criar spiders (aranhas) que seguem links em uma página da web para coletar dados de outras páginas. O Scrapy é usado para extrair informações de sites da web em larga escala, como para monitorar preços de produtos em lojas online ou coletar notícias de vários sites de notícias.

● **Selenium**

É uma ferramenta de automação de testes para a web. Ele permite que os desenvolvedores escrevam testes automatizados para aplicativos web em vários navegadores, simulando a interação do usuário com o aplicativo.



Consumindo dados de API públicas

Glossário

TensorFlow

É uma biblioteca de aprendizado de máquina de código aberto para Python. Ele é usado para criar modelos de aprendizado profundo e é conhecido por sua eficiência em processar grandes quantidades de dados. O TensorFlow é usado em áreas como reconhecimento de voz, detecção de imagens e análise de dados.

URL (Uniform Resource Locator)

É um endereço que identifica uma página da web ou qualquer outro recurso na internet. A URL geralmente começa com o protocolo de transferência (como HTTP ou HTTPS), seguido pelo nome do servidor, caminho para o recurso e, opcionalmente, parâmetros de consulta.



Consumindo dados de API públicas

Glossário

Web development

É o processo de criação de sites, aplicativos e outros recursos da web usando tecnologias como HTML, CSS, JavaScript e Python. Os desenvolvedores web usam essas ferramentas para criar e implementar designs de sites, interatividade e funcionalidades avançadas na web.

Web spider (aranha da web)

É um programa de computador que automatiza a coleta de dados na web. Ele segue os links em uma página da web para coletar dados de outras páginas, e assim sucessivamente, até que um conjunto de dados seja coletado. Os web spiders são frequentemente usados para coletar informações de vários sites da web em larga escala, como para fins de análise de dados ou monitoramento de preços. O Scrapy é um exemplo de um framework para construir web spiders em Python.



Bons estudos!

