



MÓDULO 03

HTTP E APIs

mentorama.



Apresentação

HTTP E APIs

mentorama.

Protocolos HTTP e HTTPS

Python professional

mentorama.

Requisições web utilizando Requests

Python professional

mentorama.

Web scraping & Web Crawling

Python professional

mentorama.

Utilizando o Scrapy

Python professional

mentorama.

O que é uma API?

Python profissional

mentorama.

Trabalhando com APIs

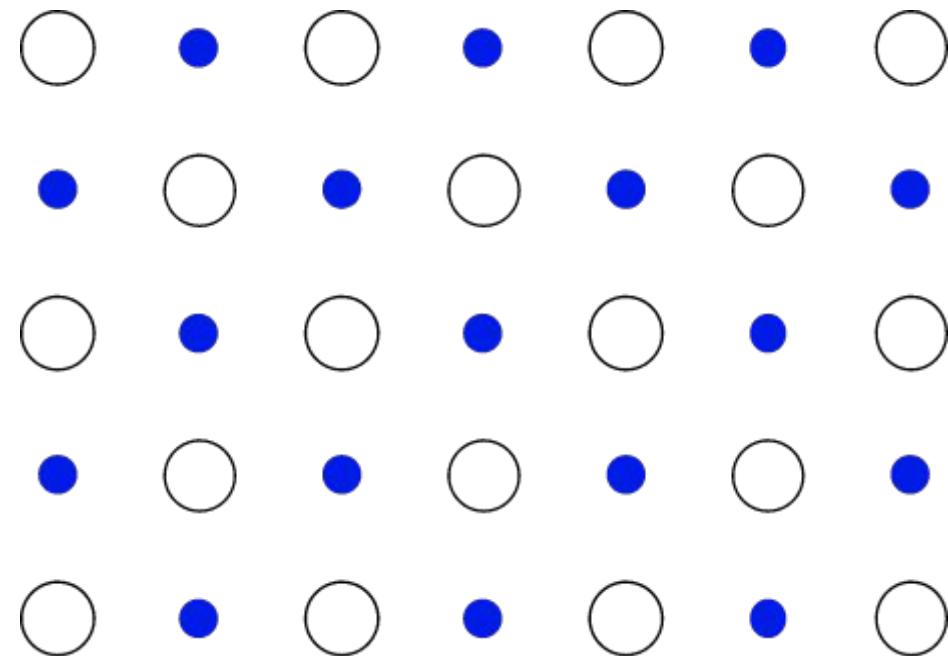
Python professional

mentorama.

**Mãos à obra: Conhecendo
um pouco mais do Scrapy.**

Mãos à obra: Vamos explorar a API do IBGE para exercitar.

mentorama.



O **Scrapy** é um framework muito potente utilizado para a criação de Web Scrapers, Web Crawlers e Spiders.

Ele traz, em apenas um framework, as ferramentas necessárias para extração dos dados e exploração do conteúdo HTML.

mentorama.

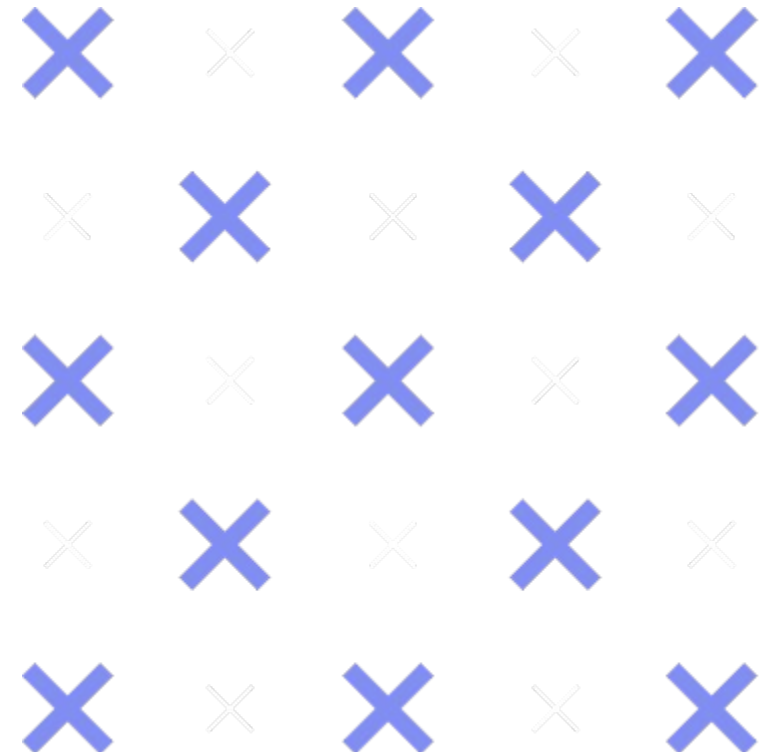
Application Programming Interface

API Web

“...é um conjunto definido de mensagens de requisição e resposta HTTP, geralmente expresso nos formatos XML ou JSON.”

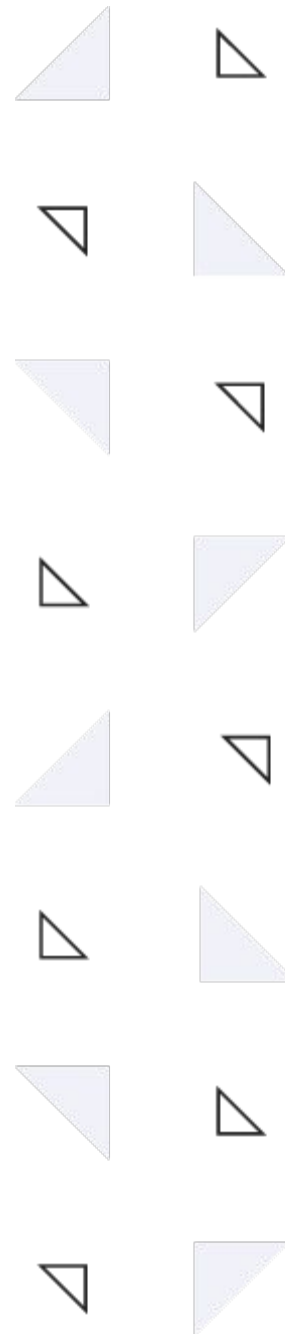
Wikipédia

mentorama.



Mãos à obra: Realizando a nossa primeira requisição com Python.

mentorama.



**Mãos à obra: Vamos obter links para
“navegação” utilizando o nosso Web
Crawler.**



mentorama.



Web Scraping:

Técnica de extração de dados da web automatizada por meio de scripts

Web Crawler:

Script com a finalidade de encontrar novos sites buscando links nos sites acessados

Spider:

Nome dados ao robô que reúne as funções de Scraping e Crawler

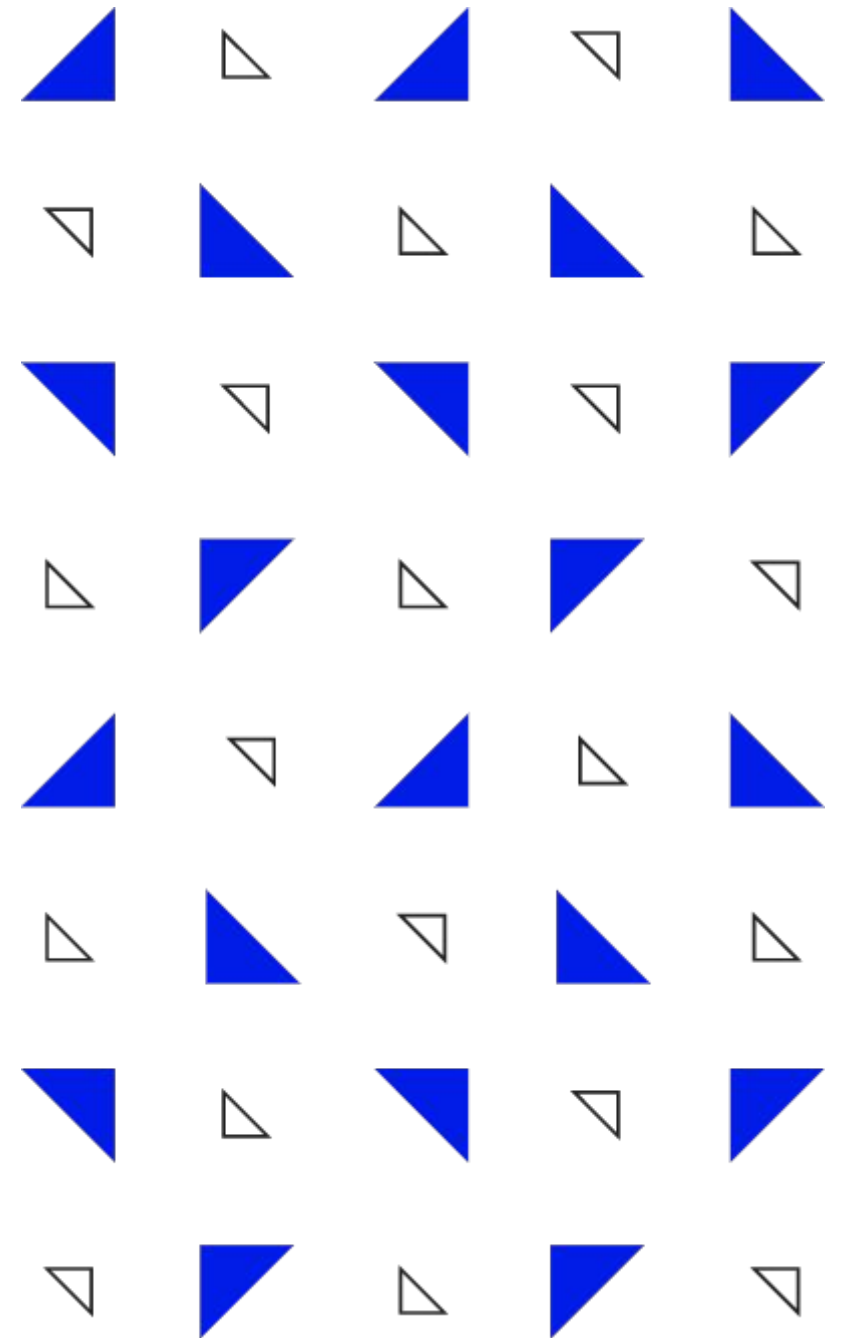
Para realizar a navegação pelas páginas Web vamos utilizar a biblioteca BeautifulSoup.

mentorama.

O que veremos?

- ◆ Protocolos HTTP e HTTPS
- ◆ Requisições web utilizando Requests
- ◆ Web scraping
- ◆ Utilizando o Scrapy
- ◆ O que é API?
- ◆ Trabalhando com APIs

mentorama.



Tarefa

Desenvolva um Web Crawler para “navegar” pelas páginas da Wikipédia:

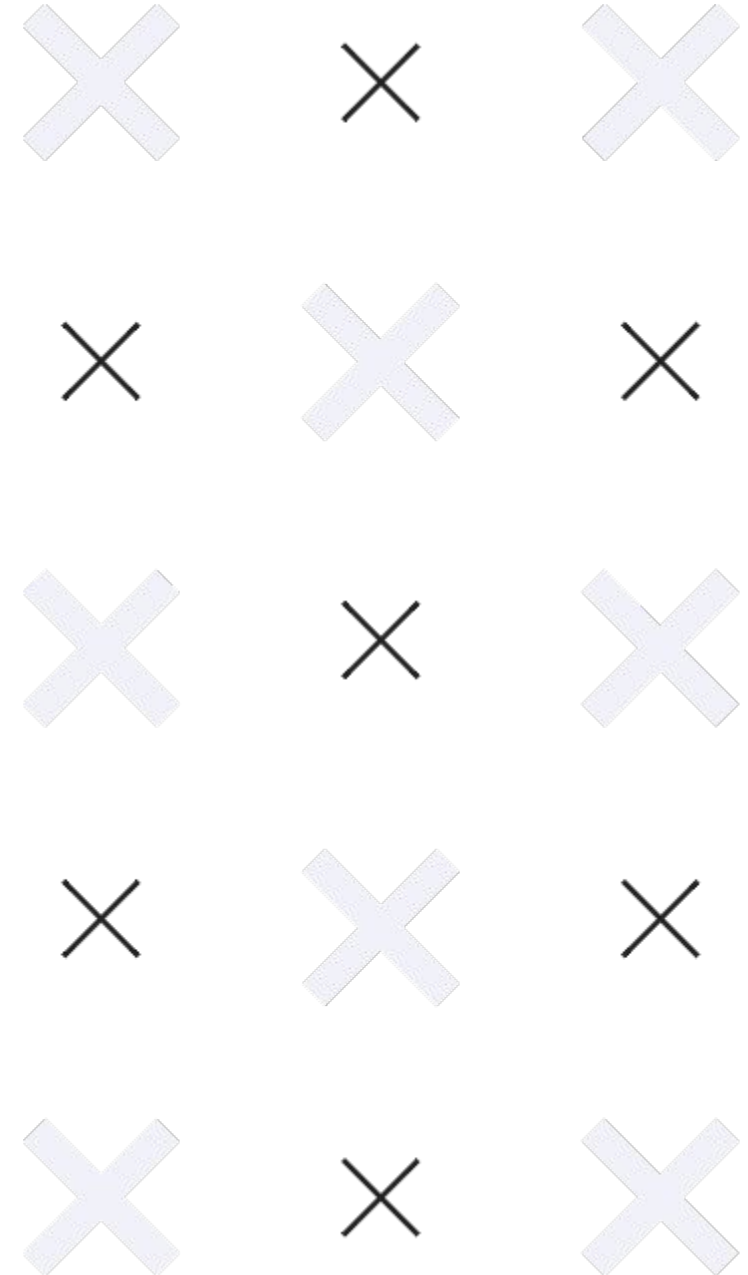
- ◆ Escolha uma página da Wikipédia e realize a raspagem dos dados da página.
- ◆ Imprima na tela a mensagem “Página principal: {Nome da página visitada}”
- ◆ Busque no texto da página em que foi feita a raspagem todos os links que apontem para outras páginas da Wikipédia.
- ◆ Faça uma nova raspagem para cada novo link capturado e imprima em cada um deles a mensagem “Página secundária: {Nome da página visitada}”.

mentorama.

- ◆ Hypertext Transfer Protocol
- ◆ Utilizado desde 1990
- ◆ Hipertextos são textos de estrutura não-linear
- ◆ O HTTP é o principal protocolo de comunicação na internet, mas não o único
- ◆ Atua na Camada de Comunicação segundo o modelo OSI



- ◆ Com o HTTP estabelecemos uma sequência de requisições e respostas entre cliente e servidor
- ◆ Os métodos mais comuns são GET e POST
- ◆ HTTPS (Hypertext Transfer Protocol Secure)
- ◆ HTTPS encripta as informações para tornar a comunicação mais segura



- ◆ Por que utilizar APIs em suas aplicações?
- ◆ Arquitetura de Microsserviços
- ◆ É importante seguir boas práticas para construção de APIs

◆ Python possui diversas bibliotecas para realizar requisições HTTP & HTTPS

- ◇ urllib
- ◇ requests
- ◇ http
- ◇ aiohttp

◆ A biblioteca requests é amplamente utilizada por sua simplicidade e eficiência

