

Web Scraping

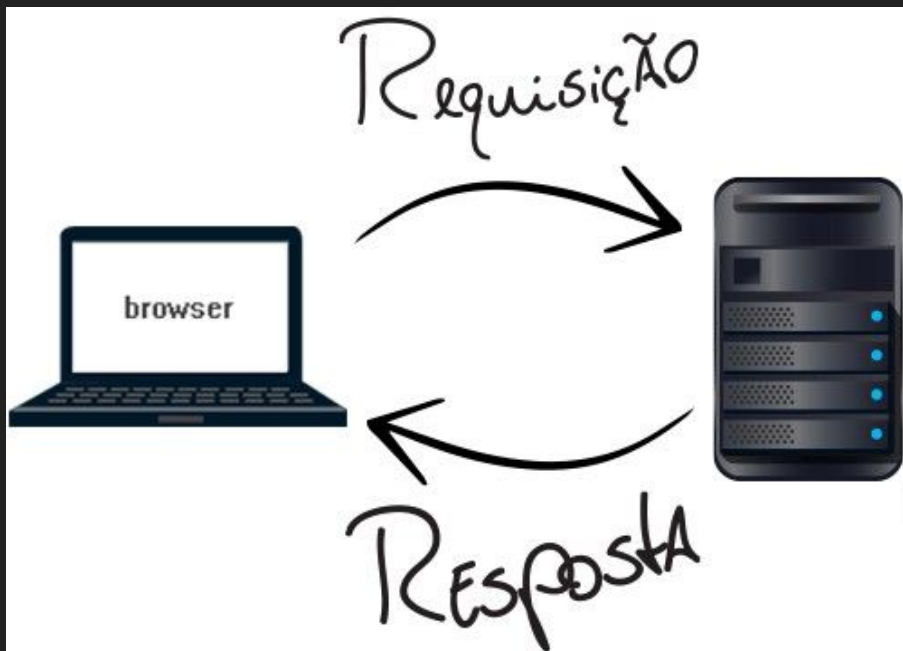
com Python



Protocolo HTTP

- Protocolos de Interconexão de Redes de Computadores.
- **Protocolo**: padronização do envio e recebimento de informações.
- **HTTP**: *HyperText Transfer Protocol*.
 - Define o modo como sites são acessados na Internet.
 - Requisição → Resposta

Protocolo HTTP



Protocolo HTTP

RESPOSTAS DE REQUISIÇÕES

- *Strings* padronizadas:
 - Código de status;
 - Cabeçalho;
 - Conteúdo.

Métodos HTTP

- **GET:** Solicita um recurso para o servidor (uma página da web, por exemplo).
- **POST:** Enviamos uma informação para o servidor.

»» Veja mais métodos ««

Códigos de Status

Família	Descrição
Informativo (1XX)	Respostas sem conteúdo, contendo apenas informação sobre a comunicação.
Sucesso (2XX)	A mensagem chegou ao servidor e era válida.
Redirecionamento (3XX)	O recurso buscado está em outro servidor.
Erro do cliente (4XX)	A requisição possui algum erro.
Erro do servidor (5XX)	O servidor não pode atender à requisição.

Códigos de Status

Família	Exemplo
Informativo (1XX)	Processing (102) – o servidor ainda está processando a requisição e, por hora, não tem conteúdo a enviar.
Sucesso (2XX)	OK (200) – tudo ocorreu exatamente como planejado: a requisição estava OK e o servidor respondeu adequadamente.
Redirecionamento (3XX)	Moved Permanently (301) – o recurso buscado mudou permanentemente de endereço.
Erro do cliente (4XX)	Not Found (404) – o recurso buscado não existe.
Erro do servidor (5XX)	Service Unavailable (503) – o servidor não é capaz de atender à solicitação no momento (por exemplo, por estar sobrecarregado ou em manutenção).

VAMOS AO CÓDIGO!

