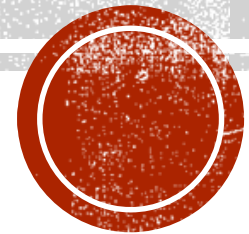


CAPTURANDO DADOS DE PÁGINAS WEB UTILIZANDO TÉCNICAS DE WEB SCRAPING COM A LINGUAGEM JAVA



Paulo Henrique da Cruz

APRESENTAÇÃO PESSOAL

- **34 anos**
- **Formado em Análise e Desenvolvimento de Sistemas - IFSP Araraquara (2016)**
- **Especializado em Desenvolvimento de Software para Dispositivos Móveis - IFSP São Carlos (2019)**
- **Pós-graduado em Arquitetura de Software Distribuído – PUC Minas (2022)**
- **MBA em Mercado Financeiro – BTG Pactual (2024)**
- **Diretor Associado de Tecnologia da Informação - Banco BTG Pactual**



JAVA WEB SCRAPING

DESENVOLVENDO UM BOT PARA CAPTURAR DADOS DA WEB, ARMAZENAR E EXPOR EM UMA API REST



▪ Objetivos:

1. Conceitos
2. Aprender a capturar dados de página Web
3. Automatizar a captura de dados
4. Armazenar dos dados
5. Expor os dados via API REST
6. Simular a API do Campeonato Brasileiro 2024



CONCEITO E BENEFÍCIOS



O conceito é bem simples, é um processo que visa extrair informações de determinado site para depois serem analisadas. A implementação de um web scraping, permite a coleta de dados em sites específicos, e pode gerar insights valiosos para o seu negócio.

▪ Benefícios

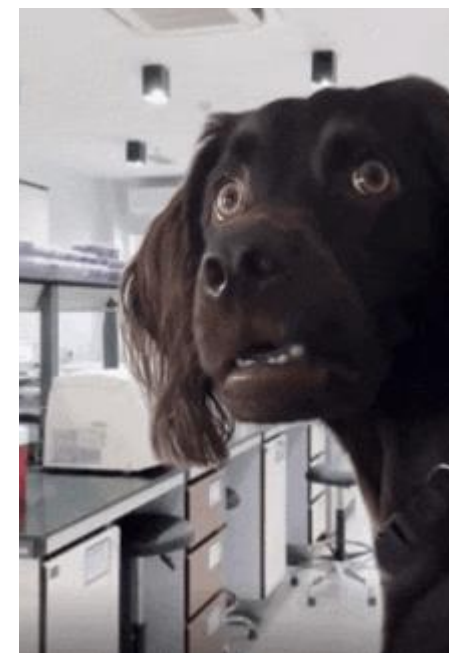
1. Realizar tarefas específicas
2. Coletar dados de fontes precisas
3. Economizar tempo e esforço
4. Automatizar processos de análise de dados



WEB SCRAPING É LEGAL OU ILEGAL?

O processo de scraping em uma página web, não é ilegal em si, mas vários sites tem políticas e ações para atrapalhar o “garimpo” de dados.

Então é muito importante, ficarmos atentos as políticas de privacidade de onde formos coletar dados.



WEB SCRAPING: POSSIBILIDADES



QUAIS FERRAMENTAS E TECNOLOGIAS VAMOS UTILIZAR?



- Java 11 ou superior
- Spring Boot
- Spring Data JPA
- Swagger
- Lombok
- Modelmapper
- Jsoup
- Banco de dados H2
- Maven



JSOUP



- É uma biblioteca java para se trabalhar com HTML. Ele fornece uma API muito rica para buscar urls, extrair e manipular HTML, utilizando-se de métodos do HTML5 e os seletores CSS.
- Site : <https://jsoup.org/>
- Documentação : <https://jsoup.org/apidocs/index-all.html>



JSOUP



- `Jsoup.connect(url).get()`

Conecta em uma url e obtém o HTML da página

- *Document*

Estrutura do documento HTML

- *Elements*

Estrutura de elementos de uma determinada parte do HTML

- *Attributes*

São os atributos de um determinado elemento



ESTRUTURA DO GOOGLE

Argentina x França

Copa do Mundo da FIFA 2022™ · 18/12/22

Encerrado



3

×

3



Argentina

Pênaltis: 4-2

França

Final

Lionel Messi 23' (P), 108'
Ángel Di María 36'



Kylian Mbappé 80' (P), 81', 118' (P)



Resumo da partida · 2:00



Resumo da partida

Feedback



Mais sobre este jogo

Argentina x França

Copa do Mundo da FIFA 2022™ · 18/12/22

Encerrado



3

×

3



Argentina

Pênaltis: 4-2

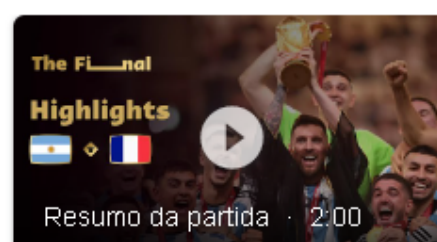
França

Final

Lionel Messi 23' (P), 108'
Ángel Di María 36'



Kylian Mbappé 80' (P), 81', 118' (P)



Resumo da partida · 2:00



Resumo da partida

Feedback



Mais sobre este jogo

OBRIGADO



“O trabalho duro vence o dom natural”

Rock Lee



OBRIGADO!



[HTTPS://WWW.LINKEDIN.COM/IN/PAULO-HENRIQUE-DA-CRUZ/](https://www.linkedin.com/in/paulo-henrique-da-cruz/)



HENRYQUE_PHC@YAHOO.COM.BR



[HTTPS://GITHUB.COM/PHCRUZ](https://github.com/PHCRUZ)

