

AUTOMAÇÃO WEB PARA COLETA E RASPAGEM DE DADOS COM SELENIUM


...

Alexandre Cajazeira | Jonnison Lima



Alexandre Cajazeira

Me. em Ciência da
Computação (UFC) e Cientista
de Dados no Jornal O POVO

 @cajazeiramos



Jonnison Lima

Doutorando em Ciência da
Computação (UFMA)
Analista de TI (UFPI)

 @jonnisonLima

Para acessar este slide:



Códigos e material complementar:

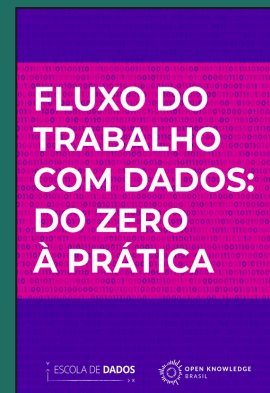


Cronograma

- Introdução
 - A importância da obtenção de dados para o fluxo de trabalho DDJ
 - O que é Web Scraping?
 - Por que raspar dados com automação de navegadores?
- Conceitos fundamentais
 - Arquitetura Cliente-Servidor e Web 2.0
 - Processamento de dados e geração de páginas do lado Cliente
 - Automatizando uso do navegador para raspagem dos dados
- A biblioteca Selenium
- Práticas e exemplos de coletas de dados
- Considerações finais e discussão

Coleta/ Obtenção de dados

- Busca-se ter em mãos conjuntos de dados para dar sustentação aos projetos de reportagem.
- Existem, basicamente, três maneiras de obtenção de dados:
 - Coleta primária (você constrói o seu *dataset*)
 - Encontrando bases disponíveis na web
 - Via Lei de Acesso à Informação (LAI)



O que é Web-Scraping

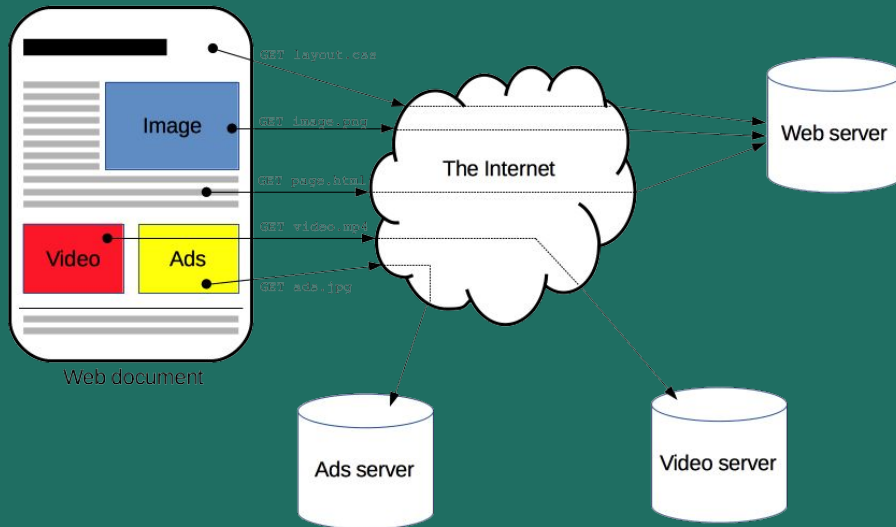
- Web Scraping consiste na elaboração de scripts que realizam requisições HTTP que simulam um usuário acessando determinados sites, extraindo dados e salvando de maneira organizada automaticamente.

Por que raspar dados com automação de navegadores?

- Cada vez mais os dados públicos são disponibilizados na Web
- Entretanto, existem...
 - Páginas que não disponibilizam os dados em formato aberto
 - Páginas com bases muito grandes que a coleta manual se torna inviável
 - Páginas que bloqueiam a coleta automatizada
 - Páginas com dados “apresentados” durante o carregamento

Revisão Teórica

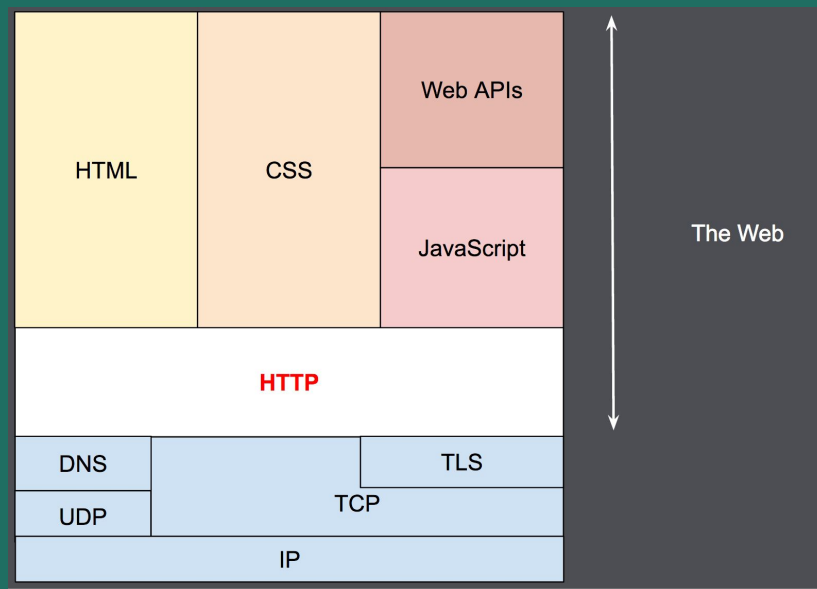
- Introdução a arquitetura cliente servidor



Fonte: <https://developer.mozilla.org/pt-BR/docs/Web/HTTP/Overview>

Revisão Teórica

- Funcionamento de páginas web



Fonte: <https://developer.mozilla.org/pt-BR/docs/Web/HTTP/Overview>

Revisão Teórica

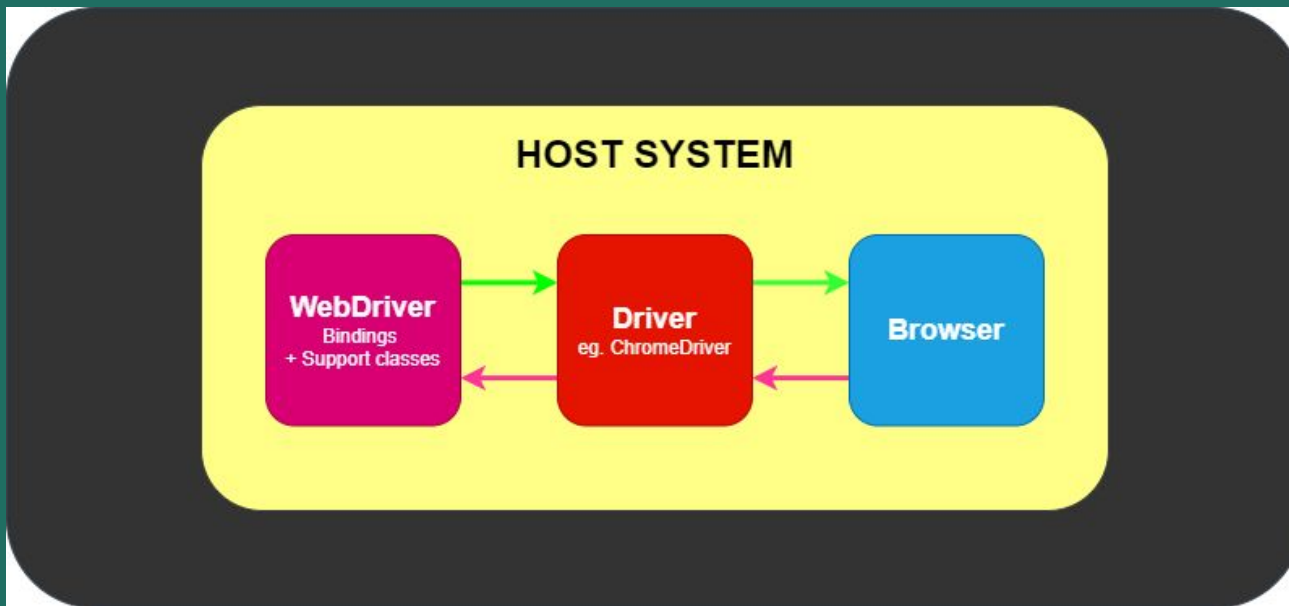
- Porque não consigo baixar só com requisição direta do conteúdo
- As páginas geram seu conteúdo dinamicamente com os dados vindo do servidor em diferentes momentos



A biblioteca Selenium - Visão Geral

- Ferramenta de testes automatizados de aplicações web para múltiplas plataformas
 - Windows, Linux, MAC OS
- Firefox, Chrome, EDGE, Safari, Opera
- Python, Java, Ruby, C, entre outras...
- Mas vamos fazer testes ou Web Scraping?

A biblioteca Selenium - Visão Geral



Fonte: <https://www.selenium.dev/documentation/overview/components/>

A biblioteca Selenium - Visão Geral

- Vamos ver isso na prática!

Práticas e exemplos de coletas com Selenium

- Prática 1 - [Painel Coronavírus / Ministério da Saúde](#) (20 min)
- Prática 2 - [Despesas por Função / Prefeitura de Fortaleza](#) (20 min)
- [Bônus] - Outros exemplos de recursos
- Exemplos de outras páginas para coletar com Selenium

Práticas e exemplos de coletas com Selenium

- Painel Coronavírus / Ministério da Saúde



CORONAVÍRUS // BRASIL

Painel Geral | Painel Interativo | OpenDATASUS | Sobre

COVID19
Painel Coronavírus
Atualizado em: 25/08/2023 15:58

CASOS CONFIRMADOS ↓

37.758.545
Acumulado | Casos novos

17967,7
Incidência*

ÓBITOS CONFIRMADOS ↓

705.170
Óbitos acumulados | Casos novos

1,9 % | **335,6**
Letalidade | Mortalidade*

Painel Interativo
Navegue pelos dados do Sistema Único de Saúde - SUS, com informações estratégicas e conheça tudo sobre a COVID-19 de forma transparente e analítica.
[Acesse](#)

Síntese de casos, óbitos, incidência e mortalidade


Pesquise uma localidade | **Brasil**

Arquivo CSV

Acesso em 25/08

Práticas e exemplos de coletas com Selenium

- Painel Coronavírus / Ministério da Saúde



CORONAVÍRUS // BRASIL

Painel Geral | Painel Interativo | OpenDATASUS | Sobre

Por que automatizar essa coleta?

COVID-19
Painel Coronavírus
Atualizado em: 25/08/2023 15:58

CASOS CONFIRMADOS

37.758.545
Acumulado | Casos novos

17967,7
Incidência*

ÓBITOS CONFIRMADOS

705.170
Óbitos acumulados | Casos novos

1,9 % | **335,6**
Letalidade | Mortalidade*

Painel Interativo

Navegue pelos dados do Sistema Único de Saúde - SUS, com informações estratégicas e conheça tudo sobre a COVID-19 de forma transparente e analítica.

Acesse

Síntese de casos, óbitos, incidência e mortalidade

Pesquise uma localidade | **Brasil**

Arquivo CSV

Acesso em 25/08

Práticas e exemplos de coletas com Selenium

- Painel Coronavírus / Ministério da Saúde

CORONAVÍRUS // BRASIL

Painel Geral | Painel Interativo | OpenDATASUS | Sobre

Como e por que descrever o processo de coleta manual?

COVID-19
Painel Coronavírus
Atualizado em: 25/08/2023 15:58

CASOS CONFIRMADOS ↓

37.758.545
Acumulado Casos novos
17967,7
Incidência*

ÓBITOS CONFIRMADOS ↓

705.170
Óbitos acumulados Casos novos
1,9 % **335,6**
Letalidade Mortalidade*

Painel Interativo
Navegue pelos dados do Sistema Único de Saúde - SUS, com informações estratégicas e conheça tudo sobre a COVID-19 de forma transparente e analítica.
[Acesse](#)

Síntese de casos, óbitos, incidência e mortalidade

Pesquise uma localidade: **Brasil** ▼

[Arquivo CSV](#)

Acesso em 25/08

Práticas e exemplos de coletas com Selenium

- Painel Coronavírus / Ministério da Saúde

CORONAVÍRUS // BRASIL

Painel Geral Painel Interativo OpenDATASUS Sobre

Quais as etapas e desafios da automatização?

COVID-19
Painel Coronavírus
Atualizado em: 25/08/2023 15:58

CASOS CONFIRMADOS ↓

37.758.545
Acumulado Casos novos
17967,7
Incidência*

ÓBITOS CONFIRMADOS ↓

705.170
Óbitos acumulados Casos novos
1,9 % **335,6**
Letalidade Mortalidade*

Painel Interativo
Navegue pelos dados do Sistema Único de Saúde - SUS, com informações estratégicas e conheça tudo sobre a COVID-19 de forma transparente e analítica.
Acesse

Síntese de casos, óbitos, incidência e mortalidade

Pesquise uma localidade **Brasil**

Arquivo CSV

Acesso em 25/08

Práticas e exemplos de coletas com Selenium

- Painel Covid - Quais as etapas da automatização?
 - Abrir e esperar o carregamento completo da página;
 - Localizar o botão para download, na estrutura HTML;
 - “Clicar” e realizar download;
 - Processar e armazenar resultado.

Práticas e exemplos de coletas com Selenium

- Painel Covid - Quais as etapas da automatização?

1ª etapa - Abrir e esperar o carregamento completo da página:

```
driver=webdriver.Firefox(firefox_profile=profile,options=options)
driver.get("https://covid.saude.gov.br/")
time.sleep(10)
```

Práticas e exemplos de coletas com Selenium

- Painel Covid - Quais as etapas da automatização?

2ª etapa - Localizar o botão de download, na estrutura HTML:

```
#SELECIONANDO O ELEMENTO E CLICANDO  
btn = [i for i in driver.find_elements(By.CSS_SELECTOR,"ion-button")  
if len(i.find_elements(By.CSS_SELECTOR,"ion-icon.icon-end"))>0][0]
```

Práticas e exemplos de coletas com Selenium

- Painel Covid - Quais as etapas da automatização?

3ª etapa - “Clicar” e realizar download:

```
btn.click()  
  
#FECHANDO O NAVEGADOR  
driver.close()
```

Práticas e exemplos de coletas com Selenium



- Painel Covid - Quais as etapas da automatização?

4ª etapa - Processar e armazenar resultado:

```
while(len(glob.glob(f"{download_path}*.zip"))==0):  
    time.sleep(1)  
  
with zipfile.ZipFile(glob.glob(f"{download_path}*.zip")[0], 'r') as zip_ref:  
    zip_ref.extractall(download_path)  
  
# ABRINDO O CSV  
csv_name = glob.glob(download_path+"*.csv")[0]  
pd_df = pd.read_csv(csv_name, sep=";")  
pd_df.head()
```

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Prefeitura de Fortaleza

[Portal Novo](#)

PORTAL DA TRANSPARÊNCIA DE FORTALEZA

INÍCIO RECEITAS DESPESAS CONTAS PÚBLICAS ORÇAMENTO PLANEJ. E MONIT. LICITAÇÕES CONTRATOS CONVÊNIOS CONTATO

ACESSIBILIDADE CONTRASTE A- A+ PDF Download Imprimir ?

DESPESAS POR FUNÇÃO DE GOVERNO EM 2023

2023

JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
DESCRIÇÃO DA DESPESA	VALOR ANUAL PREVISTO NA LOA (R\$)					VALOR MENSAL EMPENHADO(R\$)					VALOR ACUMULADO
LEGISLATIVA	269.609.626,00					49.517.830,52					49.517.830,52
JUDICIARIA	64.558.313,00					4.459.492,57					4.459.492,57
ESSENCIAL A JUSTICA	1.015.349,00					16.423,59					16.423,59
ADMINISTRACAO	924.062.836,00					103.941.444,76					103.941.444,76
SEGURANCA PUBLICA	348.887.131,00					16.643.669,48					16.643.669,48
ASSISTENCIA SOCIAL	290.173.373,00					11.593.513,09					11.593.513,09
PREVIDENCIA SOCIAL	1.120.618.222,00					103.129.620,46					103.129.620,46
SAUDE	2.935.721.638,00					258.916.794,89					258.916.794,89

Acesso em 25/08

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Prefeitura de Fortaleza


Portal Novo

PORTAL DA TRANSPARÊNCIA DE FORTALEZA

[INÍCIO](#)
[RECEITAS](#)
[DESPESAS](#)
[CONTAS PÚBLICAS](#)
[ORÇAMENTO](#)
[PLANEJ. E MONIT.](#)
[LICITAÇÕES](#)
[CONTRATOS](#)
[CONVÊNIOS](#)
[CONTATO](#)

ACESSIBILIDADE
CONTRASTE
A-
A
A+

PDF
Download
Imprimir
?

Por que automatizar essa coleta?

DESPESAS POR FUNÇÃO DE GOVERNO EM 2023

2023

JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
DESCRIÇÃO DA DESPESA	VALOR ANUAL PREVISTO NA LOA (R\$)					VALOR MENSAL EMPENHADO(R\$)					VALOR ACUMULADO
LEGISLATIVA	269.609.626,00					49.517.830,52					49.517.830,52
JUDICIARIA	64.558.313,00					4.459.492,57					4.459.492,57
ESSENCIAL A JUSTICA	1.015.349,00					16.423,59					16.423,59
ADMINISTRACAO	924.062.836,00					103.941.444,76					103.941.444,76
SEGURANCA PUBLICA	348.887.131,00					16.643.669,48					16.643.669,48
ASSISTENCIA SOCIAL	290.173.373,00					11.593.513,09					11.593.513,09
PREVIDENCIA SOCIAL	1.120.618.222,00					103.129.620,46					103.129.620,46
SAUDE	2.935.721.638,00					258.916.794,89					258.916.794,89

Acesso em 25/08

Práticas e exemplos de coletas com Selenium

- Despesas por Função -> BeautifulSoup:

```
page_info=urlopen('https://transparencia.fortaleza.ce.gov.br/index.php/despesa/exibirDespesasPorFuncao/2023')
soup = BeautifulSoup(page_info, 'html.parser')
tables = soup.findAll("table")

-----
HTTPError                                Traceback (most recent call last)
<ipython-input-48-f7b1f8212eb0> in <cell line: 1>()
----> 1 page_info=urlopen('https://transparencia.fortaleza.ce.gov.br/index.php/despesa/exibirDespesasPorFuncao/2023')
      2
      3 soup = BeautifulSoup(page_info, 'html.parser')
      4
      5 tables = soup.findAll("table")

-----
5 frames
/usr/lib/python3.10/urllib/request.py in http_error_default(self, req, fp, code, msg, hdrs)
    641 class HTTPDefaultErrorHandler(BaseHandler):
    642     def http_error_default(self, req, fp, code, msg, hdrs):
--> 643         raise HTTPError(req.full_url, code, msg, hdrs, fp)
    644
    645 class HTTPRedirectHandler(BaseHandler):

HTTPError: HTTP Error 403: Forbidden
```

Práticas e exemplos de coletas com Selenium

- Despesas por Função -> Requests:

```
✓ [6] a = 2023  
0s url = f'https://transparencia.fortaleza.ce.gov.br/index.php/despesa/exibirDespesasPorFuncao/{a}'  
x = requests.get(url)  
print(x.status_code)  
  
403
```

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Prefeitura de Fortaleza

Portal Novo

PORTAL DA TRANSPARÊNCIA DE FORTALEZA

INÍCIO RECEITAS DESPESAS CONTAS PÚBLICAS ORÇAMENTO PLANEJ. E MONIT. LICITAÇÕES CONTRATOS CONVÊNIOS CONTATO

ACESSIBILIDADE CONTRASTE A- A+?

PDF Download Imprimir

DESPESAS POR FUNÇÃO DE GOVERNO EM 2023

2023

JAN FEV MAR ABR MAI JUN JUL AGO SET OUT NOV DEZ

DESCRIÇÃO DA DESPESA	VALOR ANUAL PREVISTO NA LOA (R\$)	VALOR MENSAL EMPENHADO(R\$)	VALOR ACUMULADO
LEGISLATIVA	269.609.626,00	49.517.830,52	49.517.830,52
JUDICIARIA	64.558.313,00	4.459.492,57	4.459.492,57
ESSENCIAL A JUSTICA	1.015.349,00	16.423,59	16.423,59
ADMINISTRACAO	924.062.836,00	103.941.444,76	103.941.444,76
SEGURANCA PUBLICA	348.887.131,00	16.643.669,48	16.643.669,48
ASSISTENCIA SOCIAL	290.173.373,00	11.593.513,09	11.593.513,09
PREVIDENCIA SOCIAL	1.120.618.222,00	103.129.620,46	103.129.620,46
SAUDE	2.935.721.638,00	258.916.794,89	258.916.794,89

Acesso em 25/08

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Compreendendo a estrutura da página/dados:

DESPESAS POR FUNÇÃO DE GOVERNO EM 2023 ?

2023 ▼

JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
DESCRIÇÃO DA DESPESA	VALOR ANUAL PREVISTO NA LOA (R\$)	VALOR MENSAL EMPENHADO(R\$)	VALOR ACUMULADO								
LEGISLATIVA	269.609.626,00	49.517.830,52	49.517.830,52								

```
▼<li class="ui-state-default ui-corner-top">  
  <a href="#tabs-fev" onclick="clickAbaMesPDF('fev')">FEV</a> == $0  
</li>
```

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Identificando a url de acesso para cada ano/mês:

transparencia.fortaleza.ce.gov.br/index.php/despesa/exibirDespesasPorFuncao/2010#tabs-jul

DESPESAS POR FUNÇÃO DE GOVERNO EM 2010

2010

JAN

FEV

MAR

ABR

MAI

JUN

JUL

AGO

SET

OUT

NOV

DEZ

DESCRIÇÃO DA DESPESA	VALOR ANUAL PREVISTO NA LOA (R\$)
LEGISLATIVA	83.500.976,00
ESSENCIAL A JUSTICA	21.946.588,00
ADMINISTRACAO	417.682.590,00
SEGURANCA PUBLICA	82.077.294,00

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Identificando a url de acesso para cada ano/mês:

← → ↻ 🔒 transparencia.fortaleza.ce.gov.br/index.php/despesa/exibirDespesasPorFuncao/2010#tabs-jul

🔒 transparencia.fortaleza.ce.gov.br/index.php/despesa/exibirDespesasPorFuncao/2010#tabs-jul

JAN FEV MAR ABR MAI JUN JUL AGO SET OUT NOV DEZ

DESCRIÇÃO DA DESPESA

VALOR ANUAL PREVISTO NA LOA (R\$)

LEGISLATIVA

83.500.976,00

```
▼<li class="ui-state-default ui-corner-top">  
  <a href="#tabs-fev" onclick="clickAbaMesPDF('fev')">FEV</a> == $0  
</li>
```

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Coletar dados da tabela em exibição:

DESPESAS POR FUNÇÃO DE GOVERNO EM 2023

2023

JAN FEV MAR ABR MAI JUN JUL AGO SET OUT NOV DEZ

DESCRIÇÃO DA DESPESA	VALOR ANUAL PREVISTO NA LOA (R\$)	VALOR MENSAL EMPENHADO(R\$)	VALOR ACUMULADO
LEGISLATIVA	269.609.626,00	49.517.830,52	49.517.830,52
JUDICIARIA	64.558.313,00	4.459.492,57	4.459.492,57
ESSENCIAL A JUSTICA	1.015.349,00	16.423,59	16.423,59
ADMINISTRACAO	924.082.836,00	103.941.444,76	103.941.444,76
SEGURANCA PUBLICA	348.887.131,00	16.643.669,48	16.643.669,48
ASSISTENCIA SOCIAL	290.173.373,00	11.593.513,09	11.593.513,09
PREVIDENCIA SOCIAL	1.120.618.222,00	103.129.620,46	103.129.620,46
SAUDE	2.935.721.638,00	258.916.794,89	258.916.794,89

Práticas e exemplos de coletas com Selenium

- Relembrando a estrutura de uma tabela HTML:

```
<table>
  <tr>
    <th>Coluna1</th>
    <th>Coluna2</th>
  </tr>
  <tr>
    <td>Valor 1</td>
    <td>Valor 2</td>
  </tr>
  <tr>
    <td>Valor 1</td>
    <td>Valor 2</td>
  </tr>
</table>
```

Coluna1	Coluna2
Valor 1	Valor 2
Valor 1	Valor 2

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Script para coletar os dados da tabela em exibição:

```
url = 'https://transparencia.fortaleza.ce.gov.br/index.php/despesa/exibirDespesasPorFuncao/2022#tabs-{jan}'
wd = webdriver.Chrome(options=options)
wd.get(url)
time.sleep(3)

tabela = wd.find_elements(By.CLASS_NAME, 'table')[0]
rows = tabela.find_elements(By.TAG_NAME, "tr")

for row in rows[1:]:
    columns = row.find_elements(By.TAG_NAME, "td")
    for c in columns:
        print(c.text)
```

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Script para coletar os dados da tabela em exibição:

```
tabela = wd.find_elements(By.CLASS_NAME, 'table')[0]
rows = tabela.find_elements(By.TAG_NAME, "tr")

for row in rows[1:]:
    columns = row.find_elements(By.TAG_NAME, "td")
    for c in columns:
        print(c.text)
```

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Script para coletar os dados da tabela em exibição:

2022 ▼

JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
DESCRÇÃO DA DESPESA		VALOR ANUAL PREVISTO NA LOA (R\$)		VALOR MENSAL EMPENHADO(R\$)		VALOR ACUMULADO					
LEGISLATIVA		210.073.510,00		43.974.086,37		43.974.086,37					
JUDICIARIA		51.241.733,00		4.109.273,58		4.109.273,58					
ESSENCIAL A JUSTICA		965.204,00		0,00		0,00					
ADMINISTRACAO		830.313.303,00		66.589.408,95		66.589.408,95					

LEGISLATIVA
210.073.510,00
43.974.086,37
43.974.086,37
JUDICIARIA
51.241.733,00
4.109.273,58
4.109.273,58
ESSENCIAL A JUSTICA
965.204,00
0,00

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Script para coletar os dados da tabela em exibição:

```
tabela = wd.find_elements(By.CLASS_NAME, 'table')[0]
rows = tabela.find_elements(By.TAG_NAME, "tr")
resultados = []
for row in rows[1:]:
    columns = row.find_elements(By.TAG_NAME, "td")
    resultados.append({
        'Descrição da despesa': columns[0].text,
        'Valor Anual Previsto na LOA': columns[1].text,
        'Valor mensal empenhado': columns[2].text,
        'Valor Acumulado': columns[3].text,
    })
pd.DataFrame(resultados)
```

Descrição da despesa	Valor Anual Previsto na LOA	Valor mensal empenhado	Valor Acumulado
LEGISLATIVA	210.073.510,00	43.974.086,37	43.974.086,37
JUDICIARIA	51.241.733,00	4.109.273,58	4.109.273,58
ESSENCIAL A JUSTICA	965.204,00	0,00	0,00
ADMINISTRACAO	830.313.303,00	66.589.408,95	66.589.408,95
SEGURANCA PUBLICA	318.860.468,00	14.771.698,11	14.771.698,11
ASSISTENCIA SOCIAL	272.873.334,00	5.091.571,33	5.091.571,33
PREVIDENCIA SOCIAL	1.042.999.657,00	83.718.890,58	83.718.890,58
SAUDE	2.702.685.442,00	148.813.321,64	148.813.321,64
TRABALHO	17.985.423,00	102.553,04	102.553,04
EDUCACAO	2.413.922.501,00	116.424.185,33	116.424.185,33
CULTURA	102.462.129,00	786.397,84	786.397,84
DIREITOS DA CIDADANIA	80.470.288,00	3.748.559,88	3.748.559,88

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Coleta anual (mês a mês):

```
meses = ['jan', 'fev', 'mar', 'abr', 'mai', 'jun', 'jul', 'ago', 'set', 'out', 'nov', 'dez']

def coleta_anual(ano):
    resultado = []
    for i, m in enumerate(meses):
        url = f'https://transparencia.fortaleza.ce.gov.br/index.php/despesa/exibirDespesasPorFuncao/{ano}#tabs-{m}'
        wd = webdriver.Chrome(options=options)
        wd.get(url)
        time.sleep(2)

        tabela = wd.find_elements(By.CLASS_NAME, 'table')[i]
        rows = tabela.find_elements(By.TAG_NAME, "tr")
```



Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Coleta anual (mês a mês):

✓ [18] coleta_anual(2023)

	Descrição da despesa	Valor Anual Previsto na LOA	Valor mensal empenhado	Valor Acumulado	Mês	Ano
0	LEGISLATIVA	269.609.626,00	49.517.830,52	49.517.830,52	jan	2023
1	JUDICIARIA	64.558.313,00	4.459.492,57	4.459.492,57	jan	2023
2	ESSENCIAL A JUSTICA	1.015.349,00	16.423,59	16.423,59	jan	2023
3	ADMINISTRACAO	924.082.836,00	103.941.444,76	103.941.444,76	jan	2023
4	SEGURANCA PUBLICA	348.887.131,00	16.643.669,48	16.643.669,48	jan	2023
...

Práticas e exemplos de coletas com Selenium

- Despesas por Função / Fortaleza - Etapas da automatização:
 - Coleta anual (mês a mês):

```
resultado_final = []  
  
for a in range(2010,2023):  
  
    resultado_final.append(coleta_anual(a))  
  
df = pd.concat(resultado_final)
```


Outros exemplos de páginas para coletar com Selenium

- Decretos presidenciais
- Medidas Provisórias
- Painel de Assentamentos Federais / Incra
- Valores do NOVO PAC / Casa Civil
- Dentre outros

Considerações Finais

- Resolve problemas de páginas carregadas dinamicamente
- Agrega liberdade à coleta de dados
- Entretanto,
 - Os códigos são muito específicos para cada problema
 - Mudanças na fonte dos dados comprometem as soluções