

Quebrando captchas com Python



Quem sou eu?



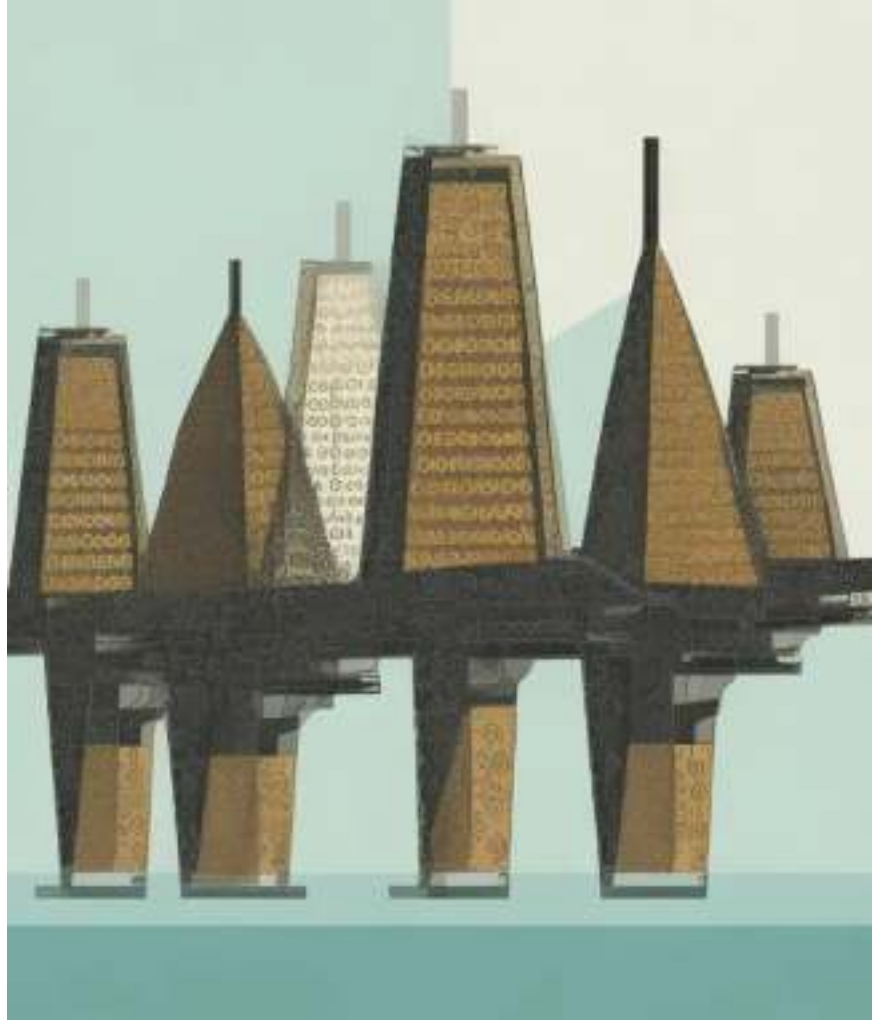
- Construo crawlers desde 2019
- Trabalho na Data Lawyer
- Cursei informática pelo IFRN
- Graduando em TI pela UFRN

Redes sociais: @joseluan ou @_joseluan

O que vamos ver?

- O que é um captcha, e onde ele aparece
- Evolução dos captchas
- Formas quebrar um captcha de imagem
- Quebrando em recaptcha v2
- 3 dicas de bypass em captcha

**Dados é o
novo
petróleo**



0 que é um captcha, e onde ele aparece

C - Completely

A - Automated

P - Public

T - Turing test

C - Computers

H - Humans

A - Apart

Na tradução livre, Teste de
Turing Público Completamente
Automatizado para
Diferenciação entre
Computadores e Humanos.



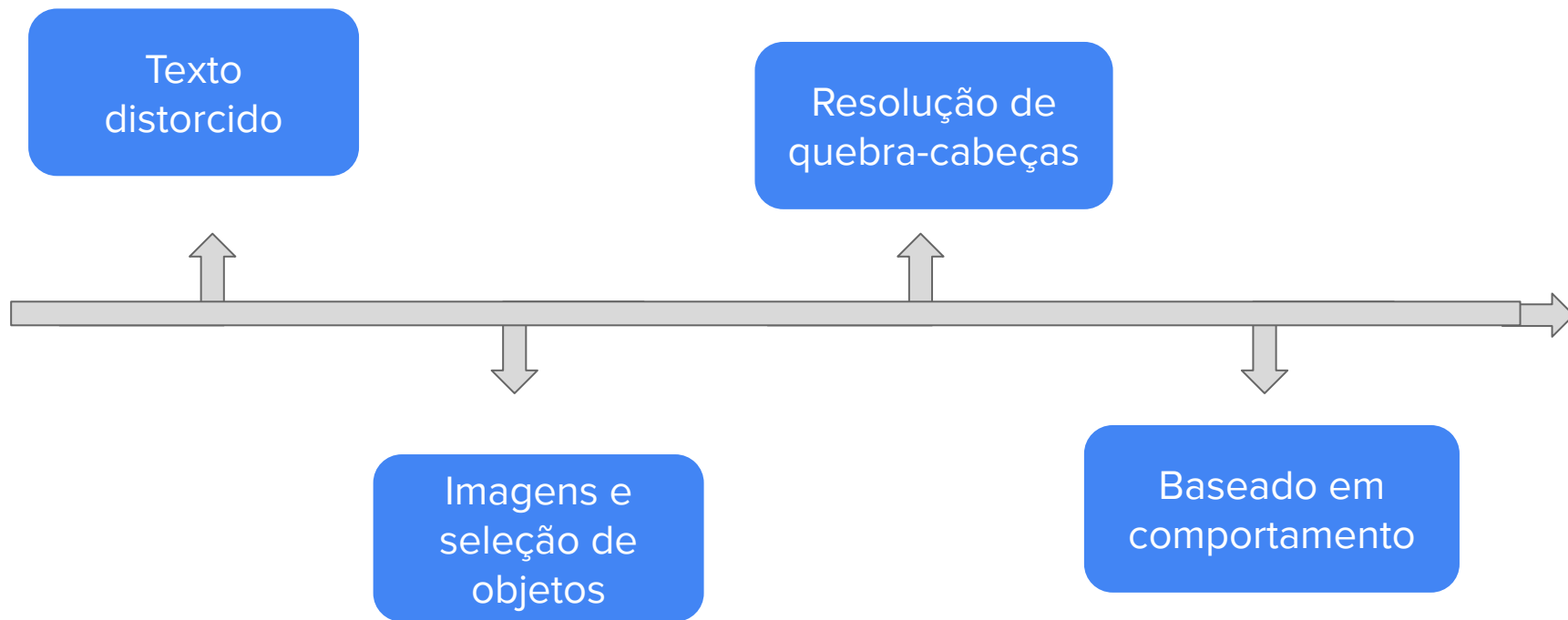
Onde podem aparecer?



gov.br

amazon

Evolução dos captchas





Portal Correios > Rastreamento >

Rastreamento

Deseja acompanhar sua encomenda?
Digite seu CPF/CNPJ ou código* de rastreamento.

AA123456785BR



* limite de 20 objetos



Digite o texto contido na imagem



I'm not a robot



I'm not a robot



Digite sua senha

CPF

Por favor, clique em cada
imagem contendo um gato
com cabelo curto.

Se não houver nenhum, clique em Pular.



PT

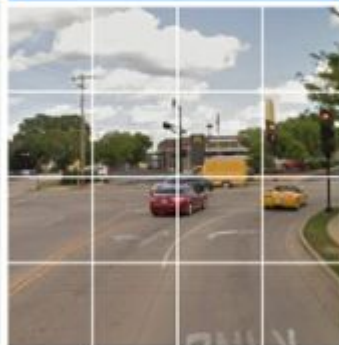


Pular

Select all squares with

traffic lights

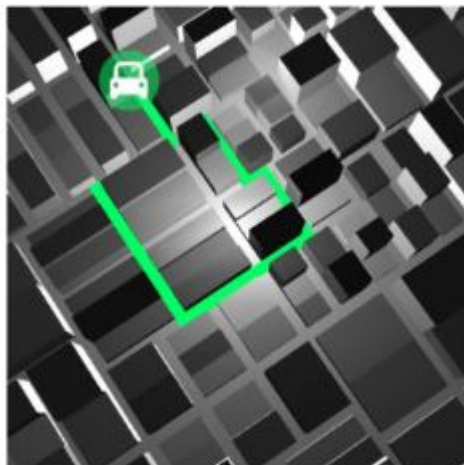
If there are none, click skip



SKIP

Solve the puzzle

Place a dot at the end of the car's path



English



Submit



Toque nas setas
para rodar a bola



Formas quebrar um captcha de imagem

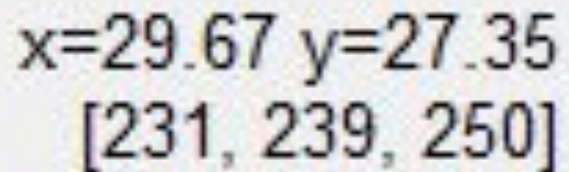


Formas quebrar um captcha de imagem

- Uma imagem, pode ser representada por uma matriz x, y
- O valor são os canais de cores RGB.

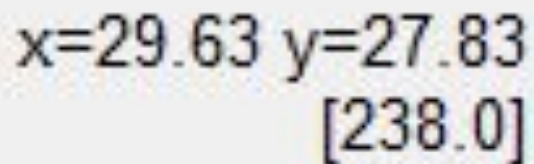


Formas quebrar um captcha de imagem



x=29.67 y=27.35
[231, 239, 250]

Imagem com 3 canais de cores



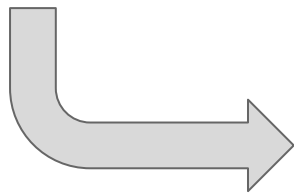
x=29.63 y=27.83
[238.0]

Imagem com 1 canais de cor

escala_cinza.py > ...

```
1  from PIL import Image
2  import matplotlib.pyplot as plt
3
4  image = Image.open('captcha_imagem.jpg')
5  gray_image = image.convert('L') # Transforma na escala de cinza
6
7  #Mostra a imagem
8  imgplot = plt.imshow(gray_image, cmap='gray')
9  plt.title('Imagem em Escala de Cinza')
10 plt.axis('off')
11 plt.show()
```

Formas quebrar um captcha de imagem



Formas quebrar um captcha de imagem

Profundidade
numa imagem é
a quantidade
de tons que
pode ser
representada
por cada pixel



Formas quebrar um captcha de imagem

- O valor do fundo da imagem gira em torno de 240
- Vamos alterar a profundidade da imagem
- Os pixels que forem acima de 230 vai ser branco e abaixo desse valor vai ser 0



profundidade.py > ...

```
1  from PIL import Image
2  import matplotlib.pyplot as plt
3  import numpy as np
4
5  image = Image.open('captcha_imagem.jpg')
6  gray_image = image.convert('L') # Transforma na escala de cinza
7
8  gray_array = np.array(gray_image) # Transforma a imagem num array numpy
9  mask = (gray_array >= 230) # Mascara para profundidade
10 gray_array_masked = np.where(mask, gray_array, 0)
11
12 # Plotar a imagem profundidade
13 plt.imshow(gray_array_masked, cmap='gray')
14 plt.title('Imagem com profundidade igual a 2')
15 plt.axis('off')
16 plt.show()
```

Formas quebrar um captcha de imagem

Imagem binarizada



V G R H

Formas quebrar um captcha de imagem

O - Optical

C - Character

R - Recognition

Vamos passar a imagem por um OCR (Reconhecimento Óptico de Caracteres)

No python podemos usar a biblioteca **pytesseract**

Formas quebrar um captcha de imagem

```
2 gray_image = np.where(mask, gray_array, 0) # Matriz profundidade
3
4 # Substitua o caminho correto para o executável do Tesseract
5 pytesseract.pytesseract.tesseract_cmd = r'C:\Program Files\Tesseract-OCR\tesseract.exe'
6
7 imagem_pil = Image.fromarray(gray_image) # Converter a matriz numa imagem do pillow
8 texto = pytesseract.image_to_string(imagem_pil)#, config=custom_config) # Passando a imagem no OCR
9
10 # Mostrar o texto obtido pelo OCR
11 print(f"Texto extraído pelo OCR:", texto)
```

Formas quebrar um captcha de imagem

Texto extraído pelo OCR: NGRH

Formas quebrar um captcha de imagem

Usando um serviço externo

Instalação: **pip3 install 2captcha-python**



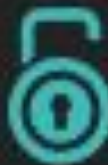
2captcha_imagem.py > ...

```
1  from twocaptcha import TwoCaptcha
2
3  api_key = 'SUA CHAVE'
4  solver = TwoCaptcha(api_key)
5
6  try:
7      result = solver.normal('captcha_imagem.jpg')
8  except Exception as e:
9      print(e)
10
11  else:
12      print('Resultado: ' + str(result))
```

Quebrando em recaptcha v2



CapMonster Cloud



2Captcha

Quebrando em recaptcha v2

<https://www.google.com/recaptcha/api2/demo>

```
▼ <div id="recaptcha-demo" class="g-recaptcha" data-  
sitekey="6Le-wvkSAAAAAPBMRTvw0Q4Muxq9bi0DJwx_mJ-"  
data-callback="onSuccess" data-action="action">  
▼ <div style="width: 304px; height: 78px;">
```

Exemplo de formulário com ReCAPTCHA

Nome

Sobrenome

E-mail

Escolha sua cor favorita:
☒ Vermelho
☐ Verde

A verificação expirou. Marque a caixa de seleção novamente.

☐ Não sou um robô

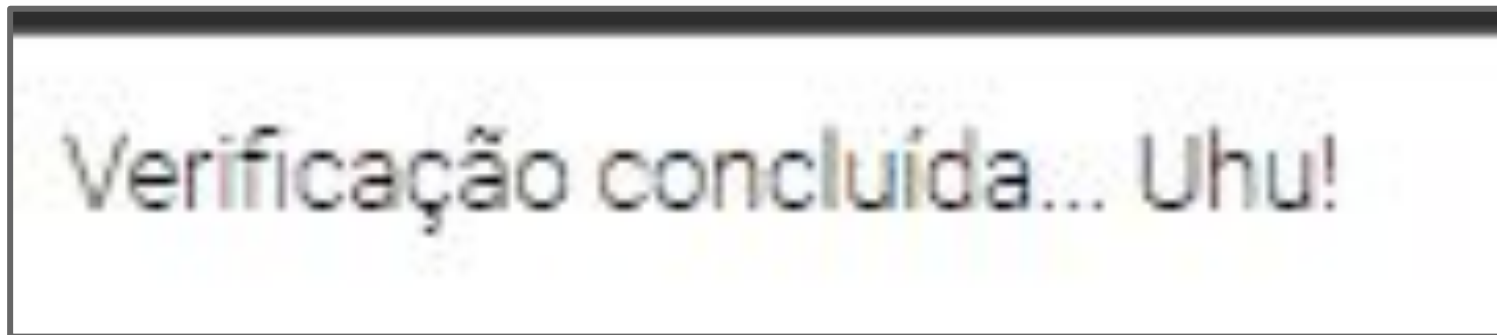
 reCAPTCHA
Privacidade • Termos

```
twocaptcha_recaptcha_v2.py > ...
```

```
1  import requests
2  from twocaptcha import TwoCaptcha
3
4
5  session = requests.session()
6  > headers = { ...
22
23  url = 'https://www.google.com/recaptcha/api2/demo'
24  response_inicial = session.get(url, headers=headers)
25
26  # Resolvendo o captcha
27  # chave que estar na pagina
28  site_key = '6Le-wvkSAAAAAPBMRTvw0Q4Muexq9bi0DJwx_mJ-'
29  # chave do 2 captcha
```

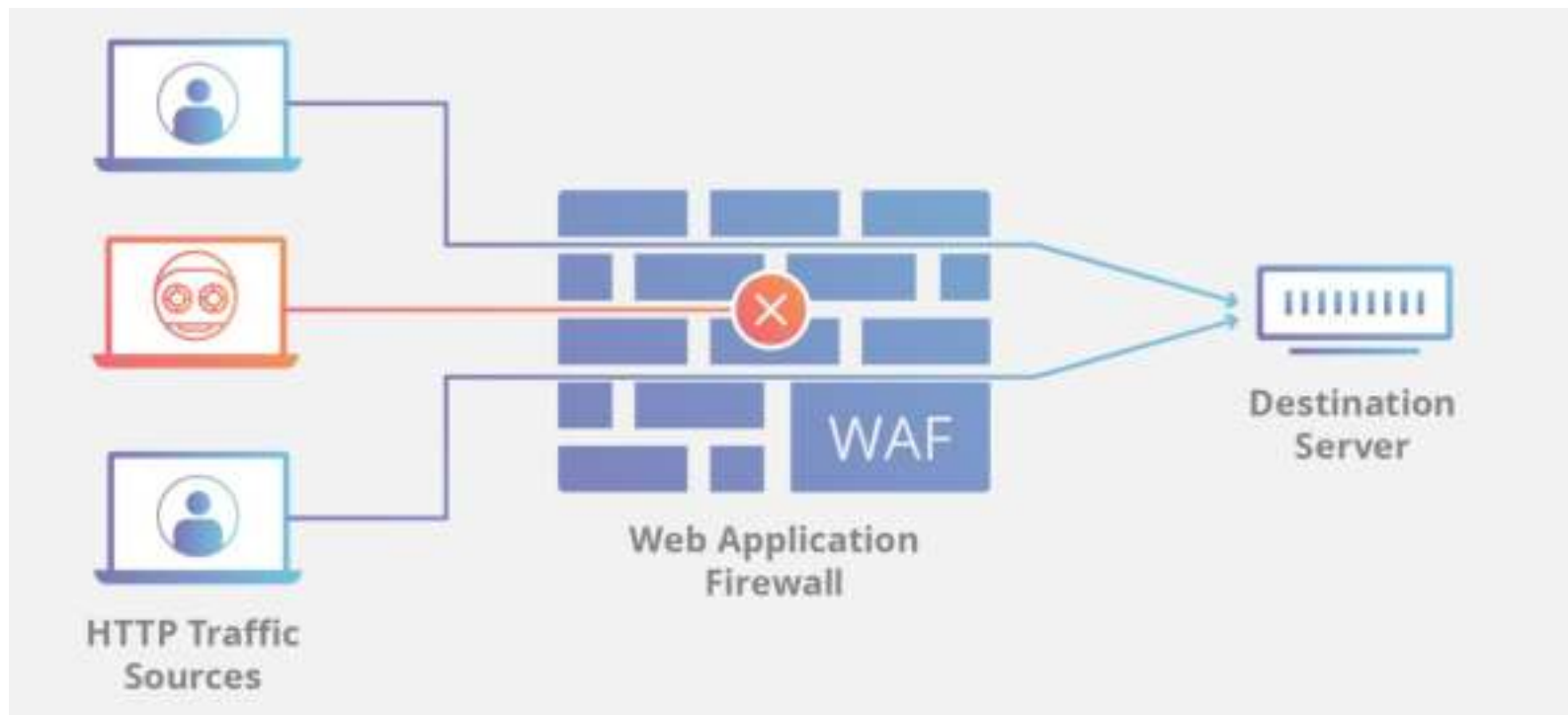
```
31 api_key = 'SUA CHAVE'
32 solver = TwoCaptcha(api_key)
33
34 try:
35     # Resolvendo o captcha
36     result = solver.recaptcha(sitekey=site_key, url=url)
37 except Exception as e:
38     print(e)
39
40 data = {'g-recaptcha-response': result['code']}
41 response_teste = session.post(url, headers=headers, data=data)
42 print(response_teste.text)
43
```

Quebrando em recaptcha v2



```
class="recaptcha-success">Verificação concluída... Uhu!</div></body></html>
```

O que é um WAF?



Bypass WAF

The ZenRows logo features a circular, multi-colored gradient (red, orange, yellow, green, blue) with a soft, glowing effect. The word "ZENROWS" is written in a bold, black, sans-serif font, positioned to the right of the gradient circle.

ZENROWS



SHODAN



2Captcha



LOLGIFS.NET

3 dicas de bypass em captcha

Fazer o bot por aplicativo

Geralmente a aplicativo não tem captcha

A api mobile é mais rápida que o site e a api web



Fazer o bot por aplicativo

O Frida é um framework que injeta uma engine javascript dentro da aplicação em tempo de execução

Ele pode ser utilizado para capturar requisições



Validação do token

- Muitos sistemas não validam se um token passado é o correto
- Verificam apenas se estar passando uma informação
- Caso não forneça no post o campo do token, ele libera a requisição

Reaproveitamento do token

- Os tokens tem uma vida útil.
- Pode ser apenas para uma request ou por um determinado tempo .
- Vale a pena salvar o token, e ir utilizando ele até que o site retorne algum erro.



Obrigado !

Fontes das imagens

