



Raspagem de Dados

**Outubro 2022**



# Quem sou eu?

## Aramuni

- 10 anos na área de desenvolvimento de sistemas
  - 4 anos VB6, VBA, Asp.Net, C#
  - 5 anos Java
  - 1 ano Python
- 5 anos como professor em cursos de tecnologia
- 2 anos e meio na Trybe
- Doutor e mestre em Sistemas de Informação
- 15 artigos publicados em revistas científicas de nível A2, B1 e B2
- 31 TCC's orientados no curso de ciência da computação



Atualmente: Tech Lead do time de currículo de Ciência da Computação

Agenda: <https://calendly.com/aramuni/30min> (Vamos conversar sobre carreira?)

LinkedIn: <https://www.linkedin.com/in/joaopauloaramuni/>

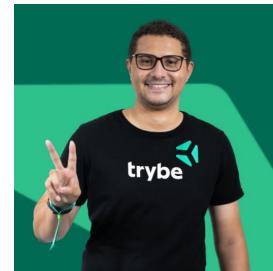
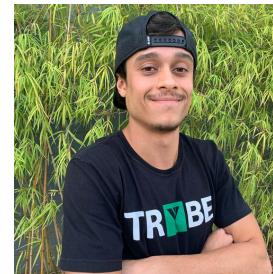
GitHub: <https://github.com/joaopauloaramuni>



# Estudantes me trouxeram até aqui



Horácio Slughorn



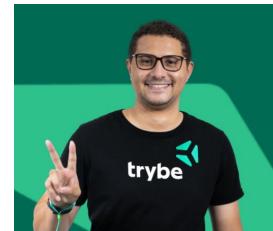
Mantinha fotos dos estudantes como troféus



# Estudantes me trouxeram até aqui



Horácio Slughorn



Os 6 fazem um ótimo trabalho na Trybe!

Não poderia ser diferente, aprenderam comigo 😂

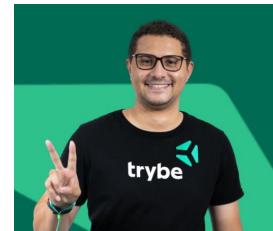
Mantinha fotos dos estudantes como troféus



# Estudantes me trouxeram até aqui



Horácio Slughorn



Um dormia na aula mas tudo bem,  
começava 07:40 da manhã

Mantinha fotos dos estudantes como troféus

# Meus hobbies





**MU**

**Hadaward**

**Normal Element Skill Enhar Master**

**Class Soul Wizard**  
**Level 930**  
**Guild DOMINIO**  
**Server [Altheim-7(Non-PvP) Serve]**

**Pts Remaining**  
**Fruit Create 107 / 127 (40%)**  
**Fruit Decrease 0 / 127 (100%)**

**Strength 318**  
**\* Attack Power 131 ~ 175**  
**\* Atk Succ rate 9836**  
**\* PvP Atk rate 16226**  
**Agility 1920**  
**\* Spi Def rate 2039**  
**\* Spi PvP Atk rate 2805**  
**\* Defense 2763**  
**\* Attack Speed 235 / 284**  
**\* Defense rate 1045 (+104)**  
**\* PvP Def rate 1280**  
**Stamina 15**  
**Energy 500**  
**\* Spi Wiz Dmg 1 ~ 4**  
**\* Wizardry Dmg 92 ~ 165 (+153)**

**< DOMINIO - DOMINIO >**  
**Hadaward**

**MM**

**Expanded Inventory**

**1 40**

**2**

**168,663,454 6,000 Personal s Expanded**

**39 / 1089**

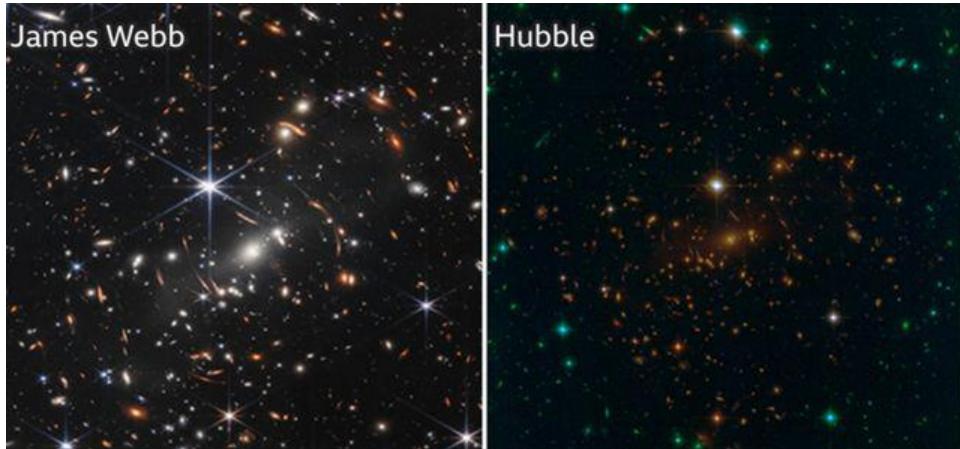
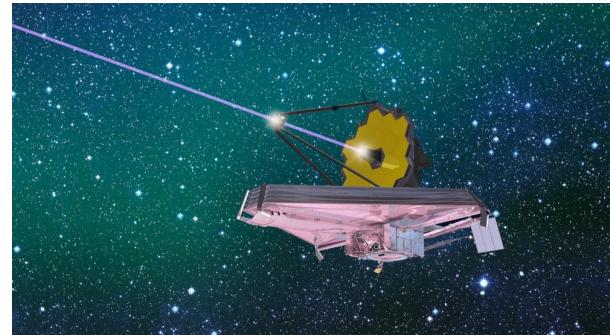
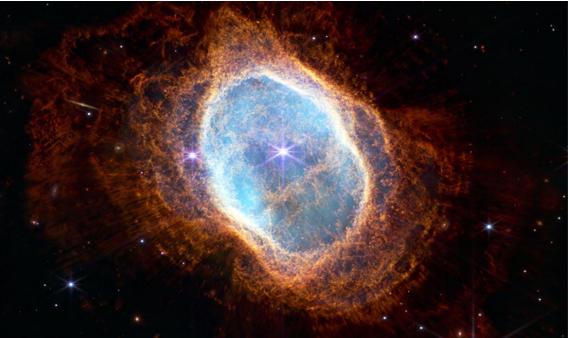
**38411113235 / 58024369465**

A composite image showing three screenshots from the game MU Online. The top left shows a character's stats window for 'Hadaward' with various attributes and levels. The middle right shows a combat scene where the player character, 'Hadaward', is fighting against multiple opponents in a dark, atmospheric setting. The bottom right shows the player's inventory screen, displaying various items and equipment.

# Curiosidade



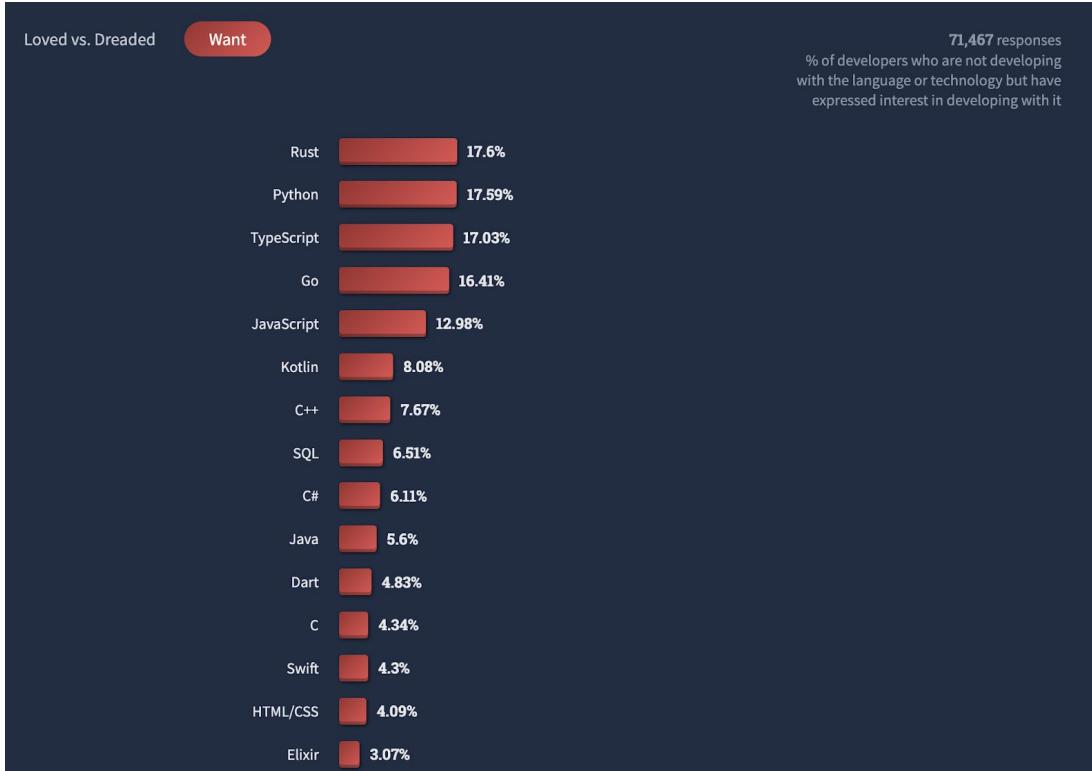
As fotos tiradas pelo telescópio James Webb foram processadas utilizando-se **PYTHON**



# Curiosidade



Pesquisa do Stack Overflow 2022 aponta o python como a segunda linguagem mais **amada** no mundo. (Diferença de 0,01 para o primeiro lugar)



# Curiosidade

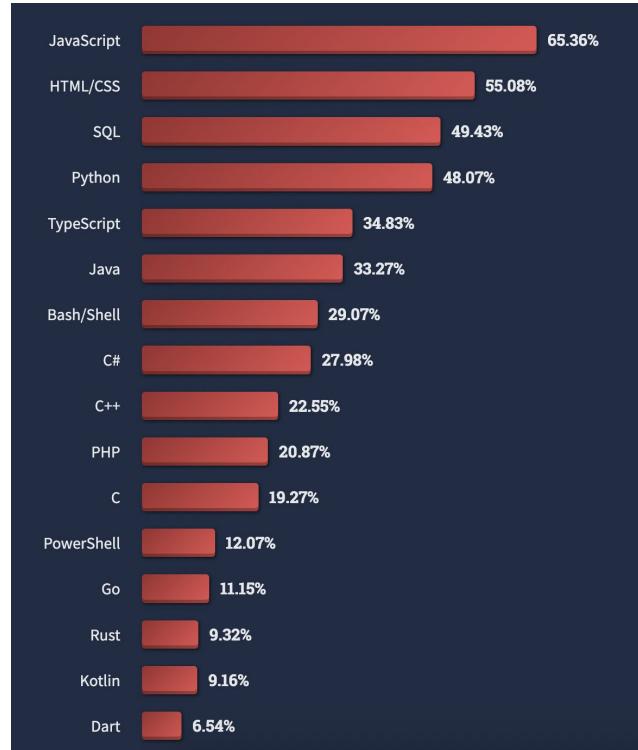


**Python** aparece em quarto lugar na pesquisa do stack overflow 2022 entre as mais populares.



Fonte:

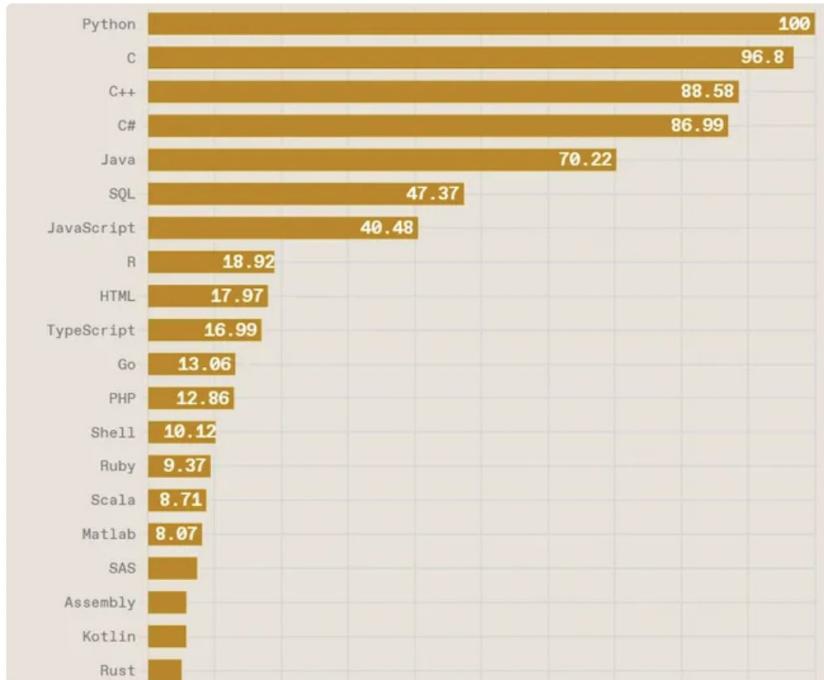
<https://survey.stackoverflow.co/2022/>



# Curiosidade



20 linguagens de programação mais **populares** de acordo com o Instituto de Engenheiros Elétricos e Eletrônicos (IEEE)



Confira as 20 linguagens de programação mais populares de acordo com o Instituto de Engenheiros Elétricos e Eletrônicos (IEEE). (Imagem: IEEE Spectrum/Divulgação)

# Curiosidade



**Python** aparece em primeiro lugar na pesquisa da analytics insight como a mais popular do mundo (14 de agosto de 2022)

Fonte:

<https://www.analyticsinsight.net/top-10-most-used-programming-languages-among-software-developers/>



# Curiosidade



Em 2016 o **Instagram** anunciou o python como a sua principal linguagem de programação.



# Curiosidade



Outros exemplos:





# Raspagem de Dados

Não teremos um slide hoje, podem ficar à vontade para levantar a mão ao final da apresentação.





# Raspagem de Dados

## O que seria raspagem de dados?

Técnica computacional na qual um programa extrai dados de páginas da web.

(Teremos alguns spoilers sobre o bloco que se inicia hoje)



# Raspagem de Dados

Termos comuns de

- Crawler
- Web Crawler (Rastejar pela web)
- Scraping
- Web Scraping
- Scrap
- Data Scraping
- Spyder (Web = Teia)
- Raspagem/Extração de dados
- 

Meu preferido: Crawleada

```
print('');  
print('');
```



# Raspagem de Dados

**Se o site não é meu, e estou extraíndo informações dele, a raspagem de dados é uma atividade ILEGAL?**





# Raspagem de Dados

**Se o site não é meu, e estou extraíndo informações dele, a raspagem de dados é uma atividade ILEGAL?**

Não.





# Raspagem de Dados

**Se o site não é meu, e estou extraíndo informações dele, a raspagem de dados é uma atividade ILEGAL?**

Lei n. 13.709/2018, é evidente que o emprego da raspagem de dados pessoais automatizada para construção de banco de dados sintetizados e disponibilização desses dados à venda trata-se de prática violadora dos direitos dos titulares e dos princípios da LGPD, que foge do legítimo interesse.





# Raspagem de Dados

**Posso usar a raspagem de dados para o mal?**

Sim.

Informação = Poder/Dinheiro

Python é a linguagem preferida dos hackers (sintaxe simples porém uma linguagem poderosa)

Veremos adiante algumas estratégias de **segurança** para evitar isso (captchas, tokens, camadas de segurança no front e no back, etc)





# Raspagem de Dados

## **Em 2018, o LinkedIn processou quase 100 pessoas**

O LinkedIn processou quase 100 pessoas que anonimamente rasparam seu site (não existe anonimato na web). E por que razões eles processaram essas pessoas?

- Violação da Lei de Fraude e Abuso de Computadores
- Violação do Código Penal da Califórnia
- Violação do Digital Millennium Copyright Act (DMCA)
- Quebra de contrato
- Apropriação indevida
- Trespass



# Raspagem de Dados

**Em 2018, o LinkedIn processou quase 100 pessoas**

Lembrando que esse assunto ainda não está pacificado e bem definido no Direito brasileiro.

Então podem haver interpretações diferentes acerca do tema e sobre a relação da LGPD com os robôs.





# Raspagem de Dados



NOTÍCIAS CORONAVÍRUS EDITORIAS SUPORTE OD SEGURANÇA OFERTAS

Olhar Digital > Segurança e Privacidade > Volvo sofre ataque por grupo de ransomware, com roubo de dados

SEGURANÇA E PRIVACIDADE

## Volvo sofre ataque por grupo de ransomware, com roubo de dados

Por Ronnie Mancuso, editado por Jeniffer Cardoso

16/12/2021 09h23, atualizada em 17/12/2021 22h11

JORNAL NACIONAL

## Brasil completa uma semana sem dados oficiais da vacinação por causa do ataque hacker ao ConecteSUS

Esse apagão de dados preocupa especialistas. Os sistemas da Fiocruz que monitoram os casos de influenza e a pandemia do coronavírus no Brasil ficaram prejudicados.

Por Jornal Nacional

17/12/2021 21h53 · Atualizado há 2 semanas



FOLHA DE S.PAULO



FOLHAJUS

## PF ainda investiga ataques hacker e 'megavazamento' de dados um ano depois

Casos de ofensivas a STJ, TSE e Ministério da Saúde, além da venda de dados de 223 milhões, estão sem desfecho



NOTÍCIAS CORONAVÍRUS EDITORIAS SUPORTE OD SEGURANÇA OFERTAS



Olhar Digital > Segurança e Privacidade > Grupo hacker envolvido no ataque ao Ministério da Saúde e à Claro no Brasil alega ter atacado sites de imprensa de Portugal

SEGURANÇA E PRIVACIDADE

## Grupo hacker envolvido no ataque ao Ministério da Saúde e à Claro no Brasil alega ter atacado sites de imprensa de Portugal

Ronnie Mancuso | 03/01/2022 16h29



# Raspagem de Dados

E o que eu posso fazer com um Crawler afinal?

The screenshot shows a news article from g1 (globo.com) titled "SP cancela blocos de rua e mantém desfiles das escolas de samba". The article discusses how São Paulo has canceled street carnival blocks but kept school samba parades. Below the main headline, there's a link to a list of more than 10 capitals that have also canceled street carnivals. To the right of the main article, there are two smaller news snippets: one about a doctor who lost his daughter to Covid and another about Fernando Fernandes becoming the new host of 'No Limite'.



JSON

```
{  
    "noticia1": {  
        "titulo": "SP cancela blocos de rua...",  
        "data": "06/01/22"  
    },  
    "noticia2": {  
        "titulo": "Vacina faz a diferença...",  
        "data": "06/01/22"  
    },  
    "noticia3": {  
        "titulo": "Fernando Fernandes é o novo apresentador...",  
        "data": "06/01/22"  
    }  
}
```



# Raspagem de Dados

**E como a raspagem é feita?**

Pela frente da aplicação, ou seja, o front-end:





# Raspagem de Dados

ASSINE JÁ E-MAIL JOÃO ARAMUNI

g1

Interior de SP

'Vacina faz a diferença', diz médico que perdeu filha para a Covid

Estreia em abril

Fernando Fernandes é o novo apresentador de 'No Limite'

Paulo Vieira vai comandar quadro inédito no BBB

DevTools is now available in Portuguese!

Always match Chrome's language Switch DevTools to Portuguese Don't show again

28 ▲ 2 2

[https://g1.globo.com/sp/ribeirao-preto-franca/noticia/2022/01/06/\\_preto-sp-celebra-inclusao-de-criancas-no-plano-nacional-de-vacinacao.ghtml](https://g1.globo.com/sp/ribeirao-preto-franca/noticia/2022/01/06/_preto-sp-celebra-inclusao-de-criancas-no-plano-nacional-de-vacinacao.ghtml)

Elements Console Sources Network ↗

<a class="bstn-hl-link" href="https://g1.globo.com/sp/ribeirao-preto-franca/noticia/2022/01/06/\_preto-sp-celebra-inclusao-de-criancas-no-plano-nacional-de-vacinacao.ghtml">

  <ul class="bstn-hl-list"> flex

    <li class="bstn-hl-itemlist bstn-hl-chapeuitem">..</li>

    <li class="bstn-hl-itemlist bstn-hl-spaceritem"></li>

    <li class="bstn-hl-itemlist bstn-hl-playbutton "></li>

    <li class="bstn-hl-itemlist bstn-hl-playbutton bstn-hl-video"></li>

  flex

  <li class="bstn-hl-itemlist bstn-hl-mainitem">

    <div>

      <div class="\_evt">

        <span class="bstn-hl-title gui-color-primary gui-color-hover gui-color-primary-bg-after">'Vacina faz a diferença', diz médico que perdeu filha para a Covid</span> == \$0

      ... 1-hl-mainitem div div\_evt span.bstn-hl-title.gui-color-primary.gui-color-hover.gui-color-primary-bg-after ...

    Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessibility

Filter :hov .cls +

element.style { }

@media (min-width: 1000px)

.is-half-size .bstn-hl-mainitem .bstn-hl-title { bastian-202...rest.css:1

  font-family: var(--font-family-display),var(--font-family-display-fallback)!important;

  font-size: var(--font-size-80-responsive-display)!important;

  letter-spacing: var(--font-size-80-responsive-display-letter-spacing)!important;

  line-height: var(--font-line-height-compact)!important;

  font-variation-settings: var(--font-variation-settings-display-normal-roman)!important;

  font-feature-settings: var(--font-feature-settings-display-roman)!important;

}

.bstn-hls:not(.variation-uber) .with-photo .bstn-hl-relateditem .bstn-hl-title, .with-photo .bstn-hl-mainitem .bstn-hl-title { bastian-202...rest.css:1

  text-shadow: □ 0 1px 3px □ rgb(0 0 0 / 80%);

  color: ■#fff!important;



# Raspagem de Dados

**E como a raspagem é feita?**

Os nomes de classes e IDs na página são a principal referência de onde está o dado que queremos:





# Raspagem de Dados

ASSINE JÁ E-MAIL JOÃO ARAMUNI

g1

Interior de SP

'Vacina faz a diferença',  
diz médico que perdeu  
filha para a Covid

Estreia em abril

Fernando Fernandes é o  
novo apresentador de 'No  
Limite'

Paulo Vieira vai comandar quadro inédito no BBB

DevTools is now available in Portuguese!

Always match Chrome's language Switch DevTools to Portuguese Don't show again

28 ▲ 2 2

`<a class="bstn-nl-link" href="https://g1.globo.com/sp/ribeirao-preto-franca/noticia/2022/01/06/_preto-sp-celebra-inclusao-de-criancas-no-plano-nacional-de-vacinacao.ghtml">`

`<ul class="bstn-hl-list"> flex`

`<li class="bstn-hl-itemlist bstn-hl-chapeuitem"></li>`

`<li class="bstn-hl-itemlist bstn-hl-spaceritem"></li>`

`<li class="bstn-hl-itemlist bstn-hl-playbutton "></li>`

`<li class="bstn-hl-itemlist bstn-hl-playbutton bstn-hl-video"></li>`

`flex`

`<li class="bstn-hl-itemlist bstn-hl-mainitem">`

`<div>`

`<div class="_evt">`

`<span class="bstn-hl-title gui-color-primary gui-color-hover gui-color-primary-bg-after">'Vacina faz a diferença', diz médico que perdeu filha para a Covid</span>`

`== $0`

`... 1-hl-mainitem div div_evt span.bstn-hl-title.gui-color-primary.gui-color-hover.gui-color-primary-bg-after ...`

Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessibility

Filter :hov .cls +

element.style { }  
@media (min-width: 1000px)  
.is-half-size .bstn-hl-mainitem .bstn-hl-title { font-family: var(--font-family-display),var(--font-family-display-fallback)!important; font-size: var(--font-size-80-responsive-display)!important; letter-spacing: var(--font-size-80-responsive-display-letter-spacing)!important; line-height: var(--font-line-height-compact)!important; font-variation-settings: var(--font-variation-settings-display-normal-roman)!important; font-feature-settings: var(--font-feature-settings-display-roman)!important; }

.bstn-hls:not(.variation-uber) .with-photo .bstn-hl-relateditem .bstn-hl-title, .with-photo .bstn-hl-mainitem .bstn-hl-title { text-shadow: □ 0 1px 3px □ rgb(0 0 0 / 80%); color: ■#ffff!important; }



# Raspagem de Dados

**Posso votar mil vezes para alguém sair do BBB?**

Podia, no passado. (Ou será que ainda dá?)





# Raspagem de Dados

## BBB 20: hacker mostra como criar bot para votar no paredão do reality



O programador Gabriel Pato desenvolveu a tecnologia com propósito experimental, mas acabou tendo seu vídeo derrubado pela Globo

Fonte

TecMundo

• 21/04/2020 01:30:00



# Paredão BBB20: Quem você quer eliminar? Babu, Flayslane ou Marcela?

Babu



Flayslane



div.gc\_\_1YQFv | 269 x 113,167

Para confirmar seu voto, selecione:

balão



⟳ Recarregar Imagens

Marcela



```
<div class="2E7_C5qCG0v-8RqBa0A3YA"> (event) {flex}
  <div class="21TgL80MNap3Ua4HvZTb4t">Flayslane</div>
  <div class="34Ne0Rx2aJ101kcITajV0f">⊗</div>
</div>
<div class="3xDixt59TdumA-tXdgvxM_2DsRxsoPgkhrgSexq-TSV0_2c2tC...
<div class="2qoDA6DYehhW7fpTiYA0d6"> (flex)
  <div class="gc_pjHD9"> (flex)
    <div class="gc__1j2k0">⊗</div> (flex)
    <div class="gc__2pMn3">
      <div class="gc__1YQFv">
        <div class="gc__20tpa">⊗</div>
        <div>
           event
        </div>
      </div>
    </div>
    <div class="gc__2muJ3">⊗</div>
  </div>
</div>
<div class="3HY6tUrdeykwokPXP7PdnT">⊗</div> (flex)
</div>
<div class="10NB5ccTCqPCo-VSuA0PnC"></div> (flex)
</div>
<div class="1-XdgvxM_2DsRxsoPgkh...> div.2qoDA6DYehhW7fpTiYA0d6 > div.gc_pjHD9 > div.gc__2pMn3 >
  <div>
    <div>
      <div>
        <div>
          <div>
            <div>
              <div>
                <div>
                  <div>
                    <div>
                      <div>
                        <div>
                          <div>
                            <div>
                              <div>
                                <div>
                                  <div>
                                    <div>
                                      <div>
                                        <div>
                                          <div>
                                            <div>
                                              <div>
                                                <div>
                                                  <div>
                                                    <div>
                                                      <div>
                                                        <div>
                                                          <div>
                                                            <div>
                                                              <div>
                                                                <div>
                                                                  <div>
                                                                    <div>
                                                                      <div>
                                                                        <div>
                                                                          <div>
                                                                            <div>
                                                                              <div>
                                                                                <div>
                                                                                  <div>
                                                                                    <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
                                                                                      <div>
................................................................
```



# Raspagem de Dados

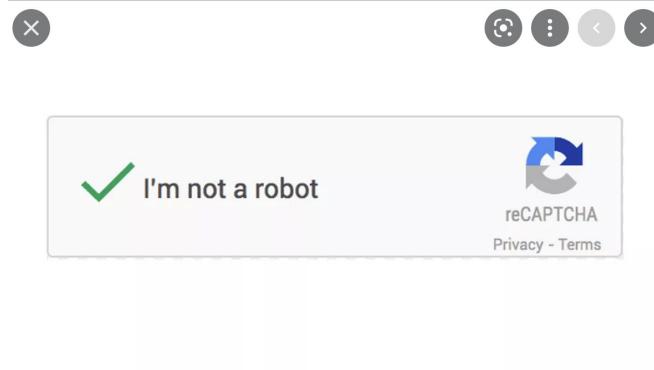
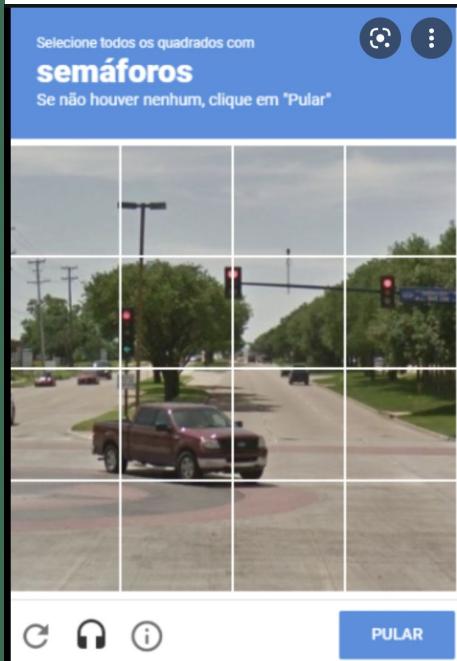
É muito provável que o captcha da edição deste ano seja diferente na hora de votar





# Raspagem de Dados

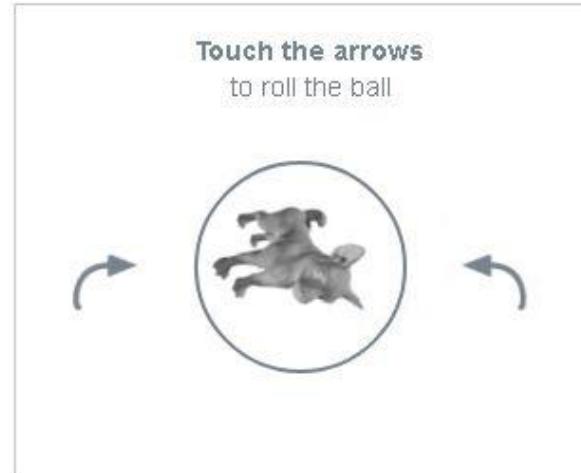
Como evitar (ou ao menos atrasar) que um crawler (ou robô) faça raspagem na minha página?





# Raspagem de Dados

Um captcha difícil de pular via código,  
o fun captcha:





# Raspagem de Dados

**Consigo pular um captcha do Google (hidrante, faixa de pedestre), pelo código python?**





# Raspagem de Dados

**Consegui pular um captcha do Google (hidrante, faixa de pedestre), pelo código python?**

Sim. Porém para resolver um captcha do Google se leva aproximadamente 1 minuto, o que diminui significativamente o número de requisições ao website.

**Como isso é feito?**





# Raspagem de Dados

**Consegui pular um captcha do Google (hidrante, faixa de pedestre), pelo código python?**

Seria impossível burlar toda a base de imagens existente no Google. Por conta disso, os captchas muitas vezes são resolvidos por **seres humanos**.

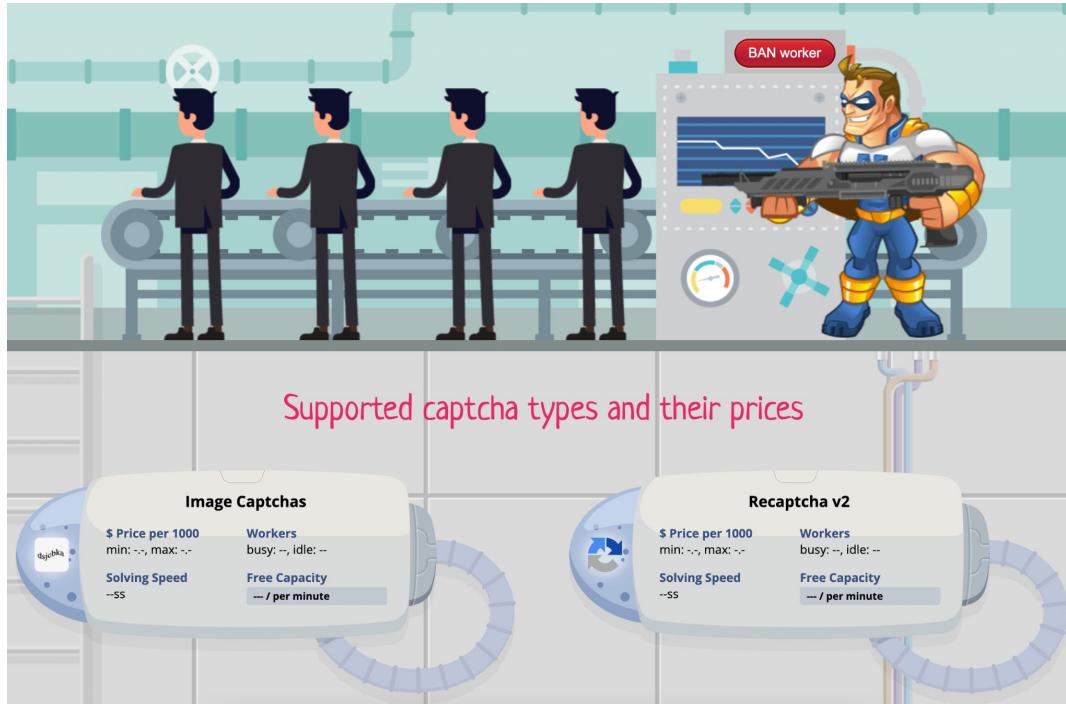
Seu crawler envia o ID do captcha para uma empresa especializada nisso e ela te devolve o ID da resposta para aquele captcha.

Também é possível contar com uma **inteligência artificial**, porém elevaria a complexidade do projeto significativamente.





# Raspagem de Dados



**2Captcha** Serviço de reconhecimento de captchas

Para trabalhadores

Dinheiro online introduzindo captcha

A introdução do captcha é o mais simples e eficaz meio de ganhar dinheiro na internet.



# Raspagem de Dados

## Consegui pular um captcha pelo código python?

```
from python_anticaptcha import AnticaptchaClient, NoCaptchaTaskProxylessTask\n\n\ndef solvecaptcha(url: str, site_key: str) -> str:\n    """\n        Solve Captcha\n        :param url:\n        :param site_key:\n        :return: solution response\n    """\n\n    api_key = "api_key_here"\n    client = AnticaptchaClient(api_key)\n    task = NoCaptchaTaskProxylessTask(url, site_key, is_invisible=True)\n    job = client.createTask(task)\n    job.join()\n\n    return job.get_solution_response()
```



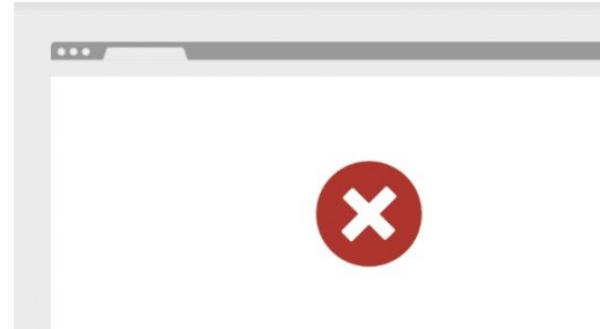


# Raspagem de Dados

## Outras estratégias de combate aos crawlers:

- Monitoramento de requisições (usuário comum: 1, robô: 10.000)
- Bloqueio temporário do IP que enviou muitas requisições
- Mudar o nome das classes e IDs do site
- Identificar a origem da requisição (é um browser verdadeiro?)
- Captchas aleatórios

Sorry, you have been blocked  
You are unable to access shahid4u.info





# Raspagem de Dados

**O ambiente onde um crawler irá raspar é extremamente INSTÁVEL.**

- A pessoa desenvolvedora que criou o site pode alterar a qualquer momento os nomes de **classes** e **IDs** que você utilizou para guiar o seu crawler.
- Camadas de segurança podem atrasar seu crawler e gerar **timeouts**.
- Seu **IP** pode ser identificado e bloqueado, etc.
- Seu **Proxy** (caso esteja utilizando) pode não funcionar corretamente.
- Pode haver um **robots.txt** no site, que é um *Protocolo de Exclusão de Robôs* (Utilizado para informar aos robôs visitantes quais diretórios de um site não devem ser raspados por ele).



# Raspagem de Dados

**O ambiente onde um crawler irá raspar é extremamente INSTÁVEL.**

Por conta disso, os contratos firmados envolvendo crawlers estabelecem uma porcentagem de sucesso, por exemplo 70%:

Se 1000 requisições forem enviadas ao site do G1, pelo menos 700 requisições voltarão com as notícias desejadas.

**100% é praticamente impossível, então não desanime caso o seu projeto não retorne de primeira a informação que você deseja!**





# Raspagem de Dados

## **Outros exemplos de uso de crawlers:**

- E-commerce (este produto hoje custa quanto em cada site? Americanas, Magalu, Amazon, etc)
- Bancos de dados públicos disponibilizados pelo governo
- Redes sociais, etc





# Raspagem de Dados

amazon.com.br Enviar para João Belo Hor... 30421 Todos notebook dell

Todos Venda na Amazon Mais Vendidos Ideias para Presente Eletrônicos Amazon Moda Computadores Novidades na Amazon Brinquedos e Jogos Comprar novamente

Computadores e Informática Mais Vendidos Ofertas Notebooks Desktops PC Gaming Monitores Acessórios Impressoras e Acessórios

« Voltar aos resultados



Notebook Dell Inspiron i15-3501-A10P 15.6"  
HD 11ª Geração Intel Pentium Gold 4GB  
128GB SSD Windows 10 Preto

Visite a loja Dell

5 estrelas 273 avaliações de clientes | 141 perguntas respondidas

Amazon Indica para "notebook dell"

Por: R\$ 2.565,00 ✓prime

à vista (4% off)

ou R\$ 2.699,99 em até 10x de R\$ 270,08 sem juros Ver parcelas disponíveis



Pagamentos e Segurança



Enviado pela Amazon



Política de devolução

Série Inspiron 15 3000 Series 3501

Marca Dell

Usos específicos Personal, Gaming, Business do produto

Tamanho da tela 15.6

Sistema operacional Windows 10 Home

Detalhes do produto

Detalhes do produto

americanas • mais

busque aqui seu produto



oi, João minha conta



Rua Junquinhos, Nova Sul...

empresas

baixe o app

receba hoje

produtos internacionais marcas próprias

venda na americanas americanas social oferta dia

compre por departamento

mercado

material escolar

celulares

ar condicionado eletrodomésticos móveis tv e home theater notebooks

oba, cupom! com o cupom: LIQUIDA10 ganhe 10% de desconto

confira as regras

página inicial > informática > notebooks > notebooks

favoritar comentar

compartilhar

R\$ 2.699,99 77%  
R\$ 2.510,99 à-mais  
em 1x no cartão de crédito com Ame e receba R\$ 15,00 de volta  
mais formas de pagamento



Notebook Dell Inspiron i15-3501-A10P Intel Pentium Gold-7505 4GB 128GB W10 HD 15,6" Preto  
5 estrelas (114) 143 perguntas  
mais informações do produto

cor: preto

voltagem: bivolt

política de troca e devolução

R\$ 2.699,99 77%  
R\$ 2.510,99 à-mais

em 1x no cartão de crédito com Ame e receba R\$ 15,00 de volta  
mais formas de pagamento

Rua Junquinhos, Nova Suiça - Belo...

receba até amanhã à-mais grátils

retire na loja a partir de amanhã à-mais grátils

mais formas de entrega

comprar

comprar com Ame



# Raspagem de Dados

- Alguns crawlers que fiz

The screenshot shows the homepage of the HotMilhas website. At the top, there's a navigation bar with links for 'Início', 'Como vender suas milhas', 'Quem somos', 'Dúvidas', 'Blog', and 'Minha conta'. The main header features the 'HotMilhas' logo. Below the header, there's a form for users to enter their name and email to receive a quote via email. To the right of the form, a button says 'Recebimento agendado'. On the left, there's a section titled 'Quero vender minhas milhas:' with options for LATAM, GOL, AZUL, and LATAM Platinum/Black milhas. A large 'COTAR AGORA' button is at the bottom of the form. On the right side of the page, two men wearing 'HotMilhas' t-shirts are standing against a blue background. One man is smiling, and the other is wearing a cowboy hat. A prominent red banner in the center-right area says 'GANHE MAIS' and includes the text 'Suas milhas valem mais com o recebimento agendado.' At the bottom of the page, there's a 'SITE BLINDADO' seal and a link to 'Veja nossa Política de Privacidade.'





# Raspagem de Dados

- Alguns crawlers que fiz

The screenshot shows the homepage of the 123milhas website. At the top, there's a navigation bar with links for 'Passagens aéreas', 'Melhores destinos', 'Pacotes Promocionais', 'Carros', 'Seguro Viagem', and a green button with a dollar sign and the text 'Venda suas milhas!'. The main banner features the 123milhas logo and three smiling men wearing 123milhas t-shirts. The banner text reads: 'Compre passagens aéreas com até 50% de desconto' and 'Novo'. Below the banner is a search form for flights, including fields for 'Voar de' (origin), 'Para' (destination), 'Ida' (departure date), 'Volta' (return date), 'Passageiros' (number of passengers), and a 'Buscar' (Search) button. A '12x sem juros' payment option is shown at the bottom left, and a note at the bottom right says: 'Você não precisa ter milhas para comprar suas passagens com desconto!'. At the very bottom, it says 'Viaje com os nossos melhores pacotes' and 'Saindo de: Florianópolis'.



# Raspagem de Dados

- **Desafio**

[https://github.com/joaopauloaramuni/python/tree/master/desafio\\_python](https://github.com/joaopauloaramuni/python/tree/master/desafio_python)

**(Desafio de uma vaga antiga em BH, Dev Sr, Salário R\$ 6.000,00)**

- **Meus crawlers**

[https://github.com/joaopauloaramuni/python/tree/master/desafio\\_bots](https://github.com/joaopauloaramuni/python/tree/master/desafio_bots)

**(Alguns dos crawlers que fiz no mercado)**

# 10 motivos para aprender raspagem de dados



- 1** - Liberdade para capturar qualquer dado disponível na Web que tenha valor para você.
- 2** - Te ajuda a pensar em segurança da informação.
- 3** - Te ajuda a compreender os limites da privacidade dos dados, LGPD e afins.
- 4** - Faz você pensar como um hacker. Onde vivem, do que se alimentam.
- 5** - Te ajuda a entender a importância dos tokens, JWT, etc, entre outras iniciativas de segurança.
- 6** - Te ajuda a ser um testador melhor, uma vez que o ambiente de crawlers é sensível e exige que você repita os mesmos fluxos várias vezes e identifique problemas.
- 7** - Te torna expert em requisições HTTP, conhecendo a fundo como são montados os cabeçalhos das requisições e como o ambiente web funciona por trás dos panos.
- 8** - Torna o seu currículo mais atrativo para as empresas, uma vez que é muito comum precisar de um crawler em ambientes de e-commerce, redes sociais, raspagem de dados públicos do governo, entre outros.
- 9** - Te ajuda a entender melhor como os browsers (chrome, firefox, internet explorer kk), funcionam.
- 10** - Te ajuda a entender as estruturas de front-end, como são pensadas pela pessoa desenvolvedora. (Reforça não somente seus conhecimentos de back-end - que é onde o python está inserido - mas também de front-end)



[betrybe.com](http://betrybe.com)

