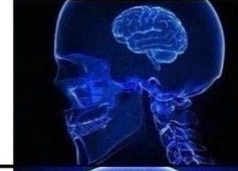


Usando Web Scrapping a nosso favor: o caçador de cupons da Udemey

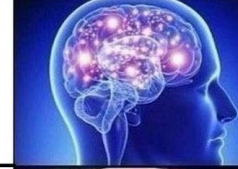
Por Ícaro Marley

Opções de ação para uma tarefa repetitiva:

Pagar alguém
para fazer



"Não sou pago o
suficiente para isso"



Começar e
desistir



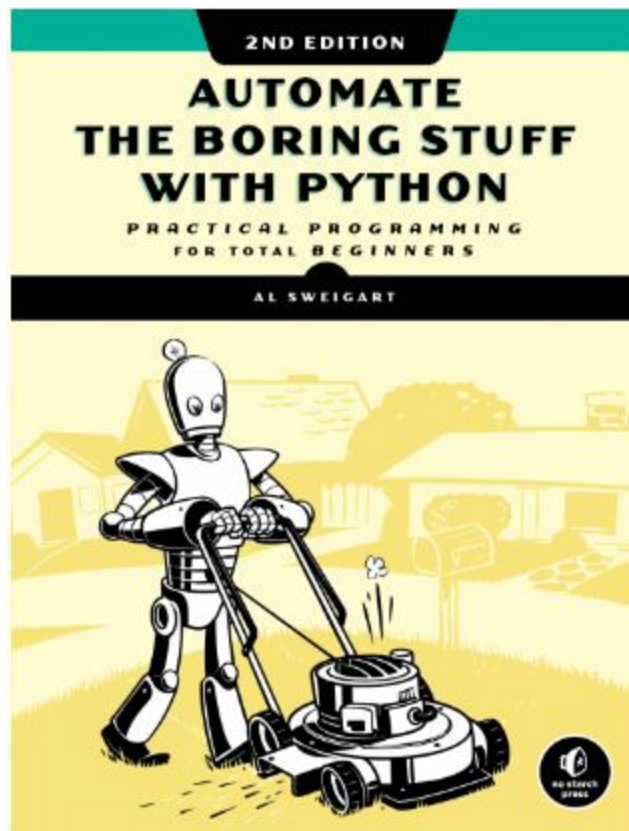
Fazer sem
questionar



**"EU SEI
PROGRAMAR"**



Inspiração



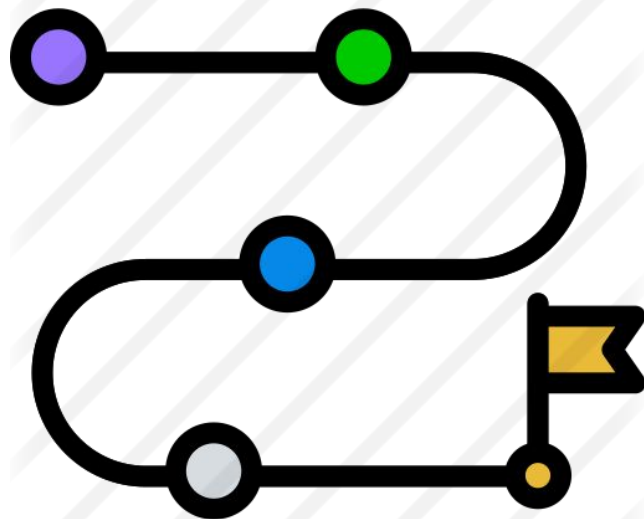
Movitação

- Testar os conhecimentos do livro
 - automação em sites
- Automatizar a busca de cursos online gratuitos



Roteiro

1. A plataforma Udemy
2. Ideia geral do projeto
3. Tecnologias utilizadas
4. A evolução do projeto
5. O projeto atual
6. Conclusão



1- A plataforma Udemey

- Plataforma EAD
- Modelo de ensino e aprendizagem descentralizado
- Em 2019:
 - 30 trilhões de alunos
 - 100 mil cursos
 - 2 mil instrutores
 - 50 países



Business

Design

Photography

Development

Marketing

IT &
Software

Personal
Development

The world's largest selection of courses

Choose from over
100,000 online video
courses with new
additions published
every month



An Entire MBA in 1 Course: Award Winning...

Chris Haroun, Complete Busines...

★★★★★ 4.5 (31,183)

R\$579.99 **R\$21.99**



The Complete SQL Bootcamp

Jose Portilla

★★★★★ 4.6 (53,923)

R\$564.99 **R\$21.99**



The Complete Financial Analyst Course 2020

365 Careers

★★★★★ 4.5 (28,192)

R\$564.99 **R\$28.99**



Development > Programming Languages > Python


Complete Python Bootcamp: Go from zero to hero in Python 3

Learn Python like a Professional! Start from the basics and go all the way to creating your own applications and games!

BESTSELLER

4.5 (236,697 ratings) 868,833 students enrolled

Created by Jose Portilla Last updated 9/2019

 English  English [Auto-generated], French [Auto-generated], [9 more](#)

+ Course Overview	4 lectures	15:28
+ Python Setup	5 lectures	39:34
+ Python Object and Data Structure Basics	21 lectures	02:02:22
+ Python Comparison Operators	2 lectures	09:08
+ Python Statements	7 lectures	01:15:18
+ Methods and Functions	14 lectures	02:02:20
+ Milestone Project - 1	5 lectures	55:51
+ Object Oriented Programming	9 lectures	01:20:49
+ Modules and Packages	3 lectures	29:10
+ Errors and Exceptions Handling	5 lectures	45:14
+ Milestone Project - 2	5 lectures	55:09
+ Python Decorators	2 lectures	23:25
+ Python Generators	3 lectures	17:17
+ Final Capstone Python Project	1 lecture	03:15
+ Advanced Python Modules	10 lectures	01:02:40
+ Advanced Python Objects and Data Structures	7 lectures	40:46
+ Bonus Material - Introduction to GUIs	7 lectures	45:15
+ APPENDIX: OLDER PYTHON 2 MATERIAL	74 lectures	10:06:14
+ BONUS SECTION: THANK YOU!	1 lecture	00:10



2 weeks ago

Carlos Eduardo Mayer d...



O curso é legal, mas acho que é para iniciantes, dizer que é "zero to hero" é muito para este conteúdo. Mas para iniciantes é um excelente ponto de entrada.

Was this review helpful?

Yes

No

Report



3 weeks ago

Sarawoot Jitsophaphun



Great for beginning start with not no anything to know more than before

in this course **If you get to know 50 %** of this course

I think it's ok

and you can relearn anytime



R\$21.99 ~~R\$564.99~~

96% off

🕒 **1 day** left at this price!

Add to cart

Buy now

30-Day Money-Back Guarantee

This course includes

- 📺 24 hours on-demand video
- 📄 18 articles
- 💻 19 coding exercises
- 🕒 Full lifetime access
- 📱 Access on mobile and TV
- 📜 Certificate of Completion

Apply Coupon



Preview this course

Free ~~R\$579.99~~ 100% off

🕒 **12 hours** left at this price!

Add to cart

Enroll now

30-Day Money-Back Guarantee

This course includes

- 📺 6 hours on-demand video
- 🕒 Full lifetime access
- 📺 Access on mobile and TV
- 📜 Certificate of Completion

Enter Coupon

Apply

✖ **FREEWEBDEV** is applied

Complete Responsive Web Development: 4 courses in 1

Master yourself in HTML5, CSS3, Bootstrap and responsive web development

★★★★☆ 3.9 (358 ratings) 44,546 students enrolled

Created by Creative Online School Last updated 3/2018 🗨 English 🎧 English [Auto-generated]

Udemy Free Coupon Codes



The Organized Schedule - Reclaim your time and live life on your own terms



Scrivener Discount Coupon 50% Off code for Windows & Scrivener 3 for Mac OS
For Mac OS & iOS



Java Programming: Complete Beginner to Advanced Udemy course 100% OFF



Python for Machine Learning Bootcamp 100% Off



Managerial Accounting / Cost Accounting Course 100% Off



Xero Accounting - Bank Feeds & Credit Card Feeds 2020 Course 100% Off



Bank Feeds - QuickBooks Online, Xero, Sage, Wave (Comparison) Course 100% Off



Sage 50cloud Accounting 2020 Course 100% Off

A tortura de procurar descontos manualmente

- Novos surgem todo o tempo
- Antigos expiram com o uso/tempo
- A maior parte já expirou
- Cupons válidos por muito tempo são de cursos menos requisitados

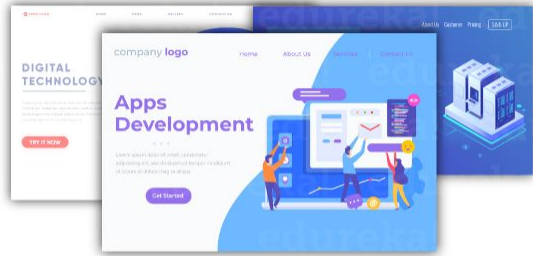


2- Ideia geral do projeto

- Automatizar as ações de alguém que procura cupons:
 - coleta → análise → compra

```
para cada site que possui cupons de desconto:  
    extrair cupons  
    para cada cupom:  
        se cupom for de 100%:  
            comprar na Udemy
```

Web Scraping



Webpages



Web Scraping



edureka!



Structured Data

Legalidade do Web Scraping

- Web Scraping (e Web Crawling) sem permissão é **ilegal**:
 - sobrecarga dos servidores web
 - viola os termos de serviço
 - é considerado **roubo de dados**



Legalidade do Web Scraping

- Existem milhares de robôs atuando na internet
 - Isso dificulta processos jurídicos
- A solução mais utilizada é bloquear o acesso
- O arquivo “robots.txt”



Em relação a esse projeto

- O Web Scraping ocorre nos sites de cupons, onde:
 - os mesmos cupons estão disponíveis em outros sites
 - não assinei um contrato ou termo de serviço



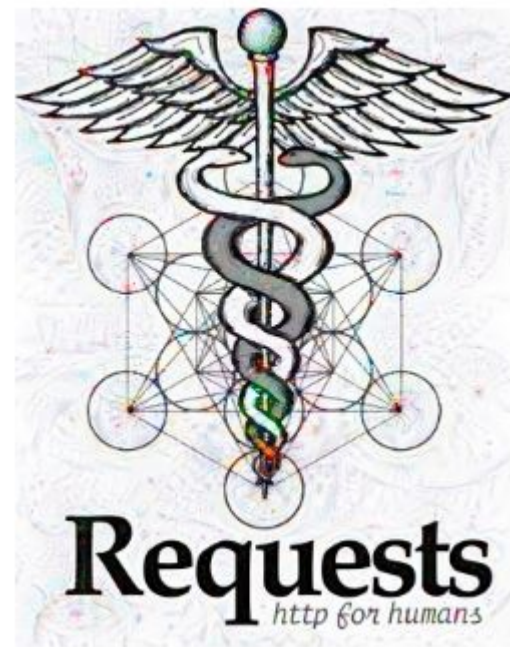
Em relação a esse projeto

- Na Udemy há um contrato assinado mas...
 - eu apenas automatizo a compra de cursos
 - não há extração de dados



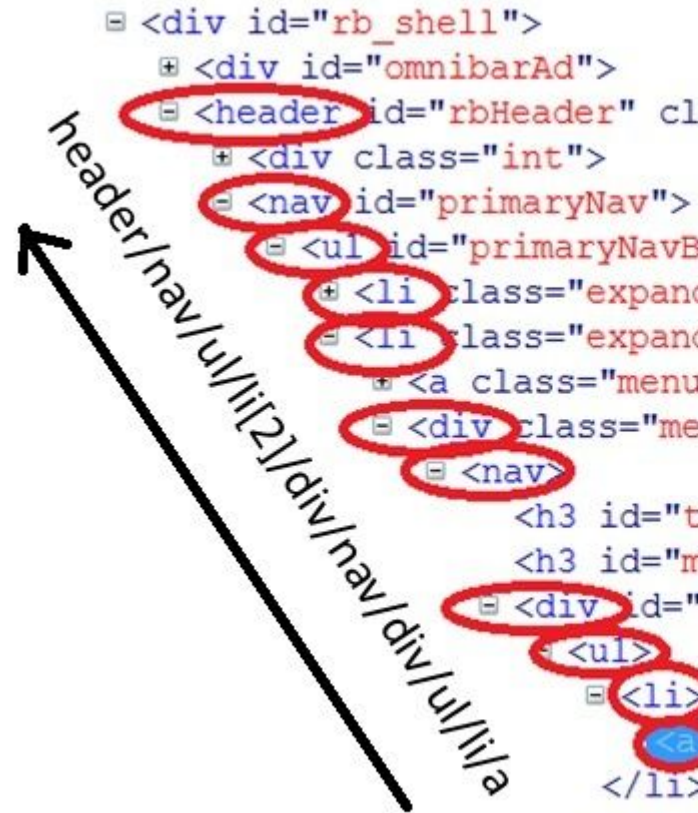
3- Tecnologias utilizadas: Requests

- Permite enviar requisições HTTP
- Coleta respostas e status
- Não executa JS
- Feita em Python



Xpath

- Linguagem de consulta para selecionar nós de documentos XML
- É utilizada em algumas bibliotecas que lidam com HTML



BeautifulSoup

- Biblioteca para processamento de HTML e XML
- Permite a navegação, busca e modificação da árvore XML (alto nível)
- Muito utilizada com a Requests
- Feita em Python



Selenium

- Biblioteca que controla navegadores
- Precisa de um webdriver e seu navegador
- Oferece suporte a navegação do código fonte (XPath)
- Voltada para testes automatizados
- Lida com HTML, CSS e JS
- Possui implementações em Python e Java



4- A evolução do projeto

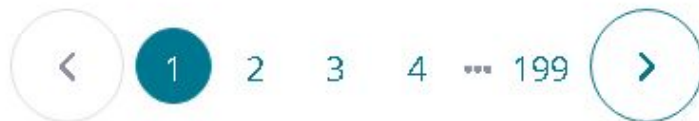
- Em meados de 2015/2016
 - Fiz uma aplicação que coletava cupons e comprava automaticamente
 - Testei versões com Selenium e Pyautogui
 - Decidi usar Selenium
- Coletei cupons dos sites
 - Profundidade: 1000

RSE

100% Complete 
Your rating

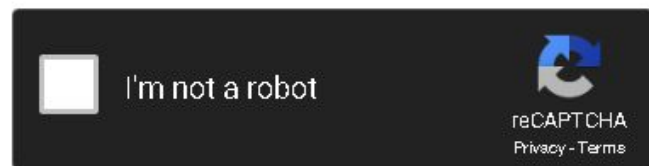
START COURSE

START COURSE



1-12 of 2377 courses

Please verify you are a human



Access to this page has been denied because we believe you are using automation tools to browse the website.

This may happen as a result of the following:

- Javascript is disabled or blocked by an extension (ad blockers for example)
- Your browser does not support cookies

Please make sure that Javascript and cookies are enabled on your browser and that you are not blocking them from loading.

[Report an issue](#)

Porque eu desisti do projeto por alguns anos?

- Evitar/resolver o CAPTCHA prejudica bastante o tempo de execução
- Eu não queria ter que forçar o acompanhamento do usuário

No meio tempo...

- Pequenas iterações ao longo dos anos
 - Simplifiquei o projeto (removi Selenium)
 - Adicionei a coleta de cupons ao GDGAju Bot
 - Comando /udemy

O que me fez reviver completamente o projeto?

Automate the Boring Stuff with Python Programming

A practical programming course for office workers, academics, and administrators who want to improve their productivity.

4.6 ★★★★★ (42,025 ratings) 402,567 students

Created by [Al Sweigart](#)

🕒 Last updated 1/2020

🌐 English

📄 English, Indonesian [Auto-generated], [2 more](#)



**You purchased this course
on Feb. 02, 2020**

30-Day Money-Back Guarantee

5- O projeto atual

- Módulos, classes e funções
- Lidando com o CAPTCHA da Udemey
- Adicionando um novo site
- Execução

```
attachEvent("onreadystatechange",H),e.attachE
boolean Number String Function Array Date RegE
_={};function F(e){var t=_[e]={};return b.ea
t[1])===!1&&e.stopOnFalse){r=!1;break}n=!1,u&
?o=u.length:r&&(s=t,c(r))}return this},remove
nction(){return u=[],this},disable:function()
re:function(){return p.fireWith(this,argument
ending",r={state:function(){return n},always:
romise)?e.promise().done(n.resolve).fail(n.re
dd(function(){n=s},t[1^e][2].disable,t[2][2].
=0,n=h.call(arguments),r=n.length,i=1!==r|e&
(r),l=Array(r);r>t;t++)n[t]&&b.isFunction(n[t
/><table></table><a href='/a'>a</a><input typ
/TagName("input")[0],r.style.cssText="top:1px
test(r.getAttribute("style")),hrefNormalized:
```

Uma conta recém criada: antes



Healthy living with "No Oil" Recipes - Non Vegetarian


Susmita Dutta, Bestselling
Udemy Instructor, Top...

[START COURSE](#)




Susmita Dutta, Udemy

Depois




Dynamic Linking Master Class - Premiere Pro & After Effects
Joe Parys, Bestselling Instructor, 80 Courses, Serving 500,000...

[START COURSE](#)




Your Mind-Body Journey Towards Wellness
Lance Rooney, Content Creator for Mind-Body Healing...

[START COURSE](#)




Machine Learning Masterclass A-Z: Beginner to Advance
Marufa Yasmin, Web Developer

[START COURSE](#)




Complete Data Science Course: Go from zero to hero
Marufa Yasmin, Web Developer

[START COURSE](#)




Compass - powerful SASS library that makes your life...
Arkadiusz Włodarczyk, Excellent teacher, Expert in Programming

[START COURSE](#)




Artificial Intelligence Music Creation & Remixing 2020
Srinidhi Ranganathan, Founder and CEO of FirstlookAi Trainin...

[START COURSE](#)



Shallow foundation
Vipin Joshi, B.Tech, M.Tech, IFERP, Editorial Board Membe...

[START COURSE](#)




Healthy living with "No Oil" Recipes - Non Vegetarian
Susmita Dutta, Bestselling UdeMy Instructor, Top...

[START COURSE](#)

[<](#) [>](#)

1-12 of 22 courses



H

6- Conclusão

- O projeto possui um longo tempo de execução (20+ mins)
- O acompanhamento humano ainda é necessário
 - CAPTCHA não pode ser completamente evitado
 - CAPTCHA precisa ser solucionado pelo usuário
- O projeto precisa ser acompanhado ao longo dos meses
 - Mudanças no HTML ocorrem de tempos em tempos

Nas próximas iterações...

- Adicionar mais sites de cupons
- Diminuir tempo de execução
- Manter bases de dados para cada usuário
- Adicionar mecanismo para superar o CAPTCHA



Links importantes

- Github do projeto:

<https://github.com/icaromarley5/Udemy-Coupon-Hunter>

🔗 Udemy-Coupon-Hunter

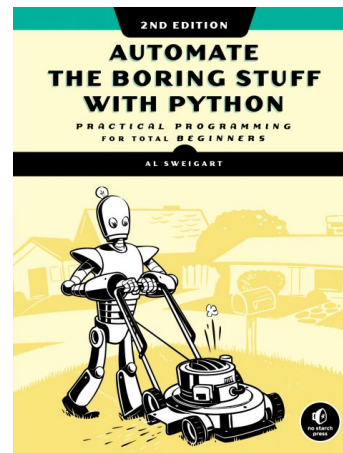
Projeto completo para a apresentação do Décimo encontro do Python Python User Group Sergipe (PUG-SE).

O script coleta cupons de 100% de desconto da plataforma Udemy (<https://www.udemy.com/>), checa a validade dos mesmos e executa a compra automática, dado um login e senha de usuário.

Projeto original em https://bitbucket.org/icaro_marley/udemy-coupon-hunter/

Links importantes

- Livro: <https://automatetheboringstuff.com/>
- Curso: <https://www.udemy.com/course/automate/>



Automate the Boring Stuff with Python Programming

A practical programming course for office workers, academics, and administrators who want to improve their productivity.

4.6 (42,025 ratings) 402,567 students enrolled

Created by Al Sweigart Last updated 1/2020

🗣 English 🌐 English, Indonesian [Auto-generated], [2 more](#)

Contato e dúvidas

 <https://www.linkedin.com/in/icaro-marley/>

 @icaromarley5

 <https://www.facebook.com/icaro.marley.9>



Referências

- <https://www.promocoupons24.com/>
- <https://www.uol.com.br/tilt/noticias/redacao/2019/11/08/hide-the-pain-harold-sorriso-nervoso-meme.htm>
- <https://www.edureka.co/blog/web-scraping-with-python/>
- <https://requests.readthedocs.io/en/master/>
- <http://datascienceacademy.com.br/blog/web-scraping-e-web-crawling-sao-legais-ou-ilegais/>
- <https://www.selenium.dev/>
- <http://www.zettaomnis.net.br/portal/a2d-mei/articles-a2d-mei/132-manipulando-xmles-com-xpath.html>
- <https://www.crummy.com/software/BeautifulSoup/>
- <https://pt.wikipedia.org/wiki/XPath>
- <https://www.udemy.com/course/responsive-web-development/?couponCode=FREEWEBDEV>
- <https://pt.wikipedia.org/wiki/Udemy>
- <https://automatetheboringstuff.com/>
- <https://www.stickpng.com/img/memes/mixed-memes/success-kid>
- https://www.flaticon.com/br/icone-premium/roteiro_2117023
- <https://www.businesswire.com/news/home/20190916005038/pt/>