



Data Science Academy (<http://datascienceacademy.com.br/blog>)

WEB SCRAPING E WEB CRAWLING SÃO LEGAIS OU ILEGAIS?

Analytics / (<http://datascienceacademy.com.br/blog/categoria/analytics/>)

Data Science / (<http://datascienceacademy.com.br/blog/categoria/data-science/>)

Web Scraping / (<http://datascienceacademy.com.br/blog/categoria/web-scraping/>)

Postado em julho 1, 2018

O que é Web Scraping? É essencialmente extrair e reunir conjuntos de dados da web (o que pode ser considerado [Big Data](http://datascienceacademy.com.br/blog/como-o-big-data-e-o-analytics-estao-transformando-o-recrutamento-e-o-rh/) (<http://datascienceacademy.com.br/blog/como-o-big-data-e-o-analytics-estao-transformando-o-recrutamento-e-o-rh/>) em alguns casos), dados esses que são a pedra angular do [Big Data Analytics](http://datascienceacademy.com.br/blog/formacao-cientista-de-dados/) (<http://datascienceacademy.com.br/blog/formacao-cientista-de-dados/>), [Machine Learning](http://datascienceacademy.com.br/blog/17-casos-de-uso-de-machine-learning/) (<http://datascienceacademy.com.br/blog/17-casos-de-uso-de-machine-learning/>) e [Inteligência Artificial](http://datascienceacademy.com.br/blog/se-a-sua-empresa-nao-e-bom-em-analytics-ela-nao-esta-pronta-para-ia/) (<http://datascienceacademy.com.br/blog/se-a-sua-empresa-nao-e-bom-em-analytics-ela-nao-esta-pronta-para-ia/>). Esses dados podem ser usados em projetos de Data Science para resolver problemas de negócio específicos e ajudar os tomadores de decisão, podendo trazer vantagem competitiva. Mas Web Scraping e Web Crawling são Legais ou Ilegais? É o que vamos discutir neste artigo.

O ponto complicado é que a informação é a mercadoria mais valiosa do mundo (depois do **tempo**, uma vez que você não pode comprar o tempo de volta), como Michael Douglas disse no famoso filme "Wall Street" muito antes da era da Internet.

Faça sua busca

Isso significa que aqueles que possuem as informações tomam todas as precauções possíveis para protegê-las contra a cópia. Nos tempos pré-Internet era fácil, pois a legislação sobre direitos autorais é bastante velha e não foi feita para o mundo da Web, onde tudo pode ser copiado por qualquer um. Mas a World Wide Web mudou tudo, pois qualquer um pode copiar o texto de uma página e colá-lo em outra página, e os web scrapers são simplesmente os algoritmos que podem fazer isso muito mais rápido que os humanos.

Curiosamente, temos visto mais e mais projetos de web scraping ultimamente. E ainda mais tutoriais que incentivam alguma forma de web scraping ou web crawling. **Mas o que incomoda é a espantosa ignorância generalizada sobre o aspecto legal disso.** Vamos primeiro definir esses termos para garantir que você entenda sobre o que estamos falando. [INÍCIO \(HTTP://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/NOSSA-HISTORIA\)](http://www.datascienceacademy.com.br/pages/nossa-historia)

Web Scraping: o ato de baixar automaticamente os dados de uma página web e extrair informações muito específicas dela. As informações extraídas podem ser armazenadas praticamente em qualquer lugar (banco de dados, arquivo, etc.). Web scraping, também conhecido como Extração de Dados da Web, é uma maneira automatizada de extrair informações/conteúdo usando bots, conhecidos como scrapers. Aqui, as informações podem ser usadas para replicar em algum outro site ou podem ser usadas para análise de dados. [TODOS OS CURSOS \(HTTP://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/TODOS-OS-CURSOS-DSA\)](http://www.datascienceacademy.com.br/pages/todos-os-cursos-dsa)

Web Crawling: o ato de baixar automaticamente os dados de uma página web, extrair os hiperlinks contidos nela e segui-los. Os dados baixados são geralmente armazenados em um índice ou banco de dados para facilitar sua busca. Web crawling, também conhecido como Indexação, é usado para indexar as informações em uma página web usando bots, também chamados de crawlers. Web Crawlers são basicamente utilizados pelos principais motores de busca como o Google, Bing e Yahoo.

Com Web Crawling obtemos informações genéricas e com Web Scraping, obtemos informações específicas.

Também é importante entender a diferença entre Web Scraping e a Mineração de Dados ([Data Mining](http://datascienceacademy.com.br/blog/4-principios-para-extrair-valor-do-big-data/) (<http://datascienceacademy.com.br/blog/4-principios-para-extrair-valor-do-big-data/>)). Resumindo, enquanto a mineração de dados pode acontecer em qualquer matriz de dados e pode ser feita manualmente, o Web Scraping ocorre apenas nas páginas web sendo executados por robôs especiais – scrapers/crawlers. [ENTRAR \(HTTP://WWW.DATASCIENCEACADEMY.COM.BR/?MSG=NOT-LOGGED-IN\)](http://www.datascienceacademy.com.br/?MSG=NOT-LOGGED-IN)

Muitos scrapers são descritos em linguagem Python para facilitar a etapa de processamento adicional dos dados coletados. Podemos escrever scrapers usando frameworks e bibliotecas para web crawling, como [Scrapy](https://scrapy.org/) (<https://scrapy.org/>), [Ghost](https://github.com/jeanphix/Ghost.py) (<https://github.com/jeanphix/Ghost.py>), [lxml](https://pypi.org/project/lxml/4.1.1/) (<https://pypi.org/project/lxml/4.1.1/>), [aiohttp](https://aiohttp.readthedocs.io/en/stable/) (<https://aiohttp.readthedocs.io/en/stable/>) ou [Selenium](https://www.seleniumhq.org/) (<https://www.seleniumhq.org/>).

No entanto, o web scraping tem se tornado uma prática maliciosa utilizada por criminosos para roubar conteúdos protegidos e cometer fraudes, repassando informações de produtos e serviços de uma empresa para a concorrência, o que pode causar grandes prejuízos aos negócios. Isso é ilegal, criminoso e pode levar para a cadeia.

Devido a esse problema, muitas organizações estão perdendo clientes e, consequentemente, diminuindo suas vendas, além de terem seus direitos autorais infringidos e dados sigilosos roubados. Por isso, é essencial que elas invistam em estratégias de segurança online para proteger seus dados e proporcionar um lugar seguro para seus clientes comprarem e disponibilizarem suas informações.

Por que web scraping é frequentemente visto de forma negativa?

A reputação do web scraping ficou muito pior nos últimos anos e por boas razões.

É cada vez mais usado para fins comerciais para obter uma vantagem competitiva e normalmente há um motivo financeiro por trás disso. Muitas vezes, é feito com total desconsideração das leis de direitos autorais e dos Termos de Serviço.

Outras vezes é feito de maneira abusiva. Por exemplo, os web scrapers podem enviar muito mais solicitações por segundo do que o que um ser humano faria, causando uma carga inesperada nos sites. Eles também podem optar por permanecer anônimos e não se identificarem. Finalmente, eles também podem realizar operações proibidas em sites, como contornar as medidas de segurança que são colocadas em prática para baixar automaticamente os dados, que de outra forma seriam inacessíveis.

Toneladas de indivíduos e empresas estão executando seus próprios web scrapers. Tanto que isso vem causando dores de cabeça para empresas cujos sites são “raspados”, como redes sociais (por exemplo, Facebook, LinkedIn, etc.) e lojas on-line (por exemplo, Amazon). Provavelmente, é por isso que o Facebook tem termos separados para coleta automatizada de dados.

Em contraste, o web crawling tem sido usado historicamente pelos mecanismos de pesquisa conhecidos (por exemplo, Google, Bing etc.) para fazer o download e indexar a web. Essas empresas construíram uma boa reputação ao longo dos anos, porque criaram ferramentas indispensáveis que agregam valor aos sites que rastreiam. Então, o web crawling é geralmente visto de forma mais favorável, embora às vezes possa ser usado de forma abusiva.

Mas afinal, é legal ou ilegal?

O web scraping e o web crawling não são ilegais por si mesmos. Afinal, você pode raspar ou rastrear seu próprio site, sem problemas.

O problema surge quando você raspa ou rastreia o site de outra pessoa, sem obter a permissão prévia por escrito ou em desconsideração dos Termos de Serviço ou Uso. Você está essencialmente se colocando em uma posição vulnerável.

Apenas pense nisso: você está usando a largura de banda de outra pessoa e está recuperando e usando livremente os dados. É razoável pensar que eles podem não gostar, porque o que você está fazendo pode prejudicá-los de alguma forma. Portanto, dependendo de muitos fatores, eles estão perfeitamente livres para entrar com uma ação legal contra você.

Eu sei o que você pode estar pensando. “Vamos lá! Isso é ridículo! Por que eles me processariam?” Claro, eles podem simplesmente ignorar você. Ou eles podem simplesmente usar medidas técnicas para bloquear você. Ou podem enviar uma carta solicitando que você pare com esta atividade. Mas tecnicamente, não há nada que os impeça de processá-lo. Este é o verdadeiro problema.

Precisa de provas?

O LinkedIn (<https://www.linkedin.com/company/data-science-academy>) está processando quase 100 pessoas que anonimamente raspam seu site (não existe anonimato).
 O Facebook (<https://facebook.com/dsacademybr>) o Twitter (<https://twitter.com/dsacademybr>) o Google+ (<http://google.com/+DataScienceAcademy>)
 o YouTube (<https://youtube.com/c/DataScienceAcademy>) o SoundCloud (<https://soundcloud.com/dsacademybr>) o Spotify (<https://open.spotify.com/track/dsacademybr>) o Instagram (<https://www.instagram.com/dsacademybr>) o LinkedIn (<https://www.linkedin.com/company/data-science-academy>)

- Violação da Lei de Fraude e Abuso de Computadores
- Violação do Código Penal da Califórnia
INÍCIO ([HTTP://WWW.DATASCIENCEACADEMY.COM.BR/](http://WWW.DATASCIENCEACADEMY.COM.BR/))
- Violação do Digital Millennium Copyright Act (DMCA)
- Quebra de contrato
NOSSA-HISTÓRIA ([HTTP://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/NOSSA-HISTORIA](http://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/NOSSA-HISTORIA))
- Trespasse
BLOG ([HTTP://DATASCIENCEACADEMY.COM.BR/BLOG](http://DATASCIENCEACADEMY.COM.BR/BLOG))
- Apropriação indevida
TODOS OS CURSOS ([HTTP://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/TODOS-OS-CURSOS](http://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/TODOS-OS-CURSOS))

Esse processo é bastante preocupante, porque não está claro o que acontecerá com essas pessoas “anônimas”

Considere que, se você for processado, não pode simplesmente desconsiderar o processo. Você precisa se defender e provar que não fez nada errado. Isso não tem nada a ver com se é justo ou não, ou se o que você fez é realmente ilegal. Uma vez processado, você terá que responder perante o juiz e comprovar sua inocência.

Outro problema é que passar por um processo pode ser algo bem estressante e custar muito dinheiro. Onde você usa lógica, bom senso e sua perícia técnica, eles usam o jargão legal e algumas áreas cinzentas da lei para provar que você fez algo errado. Este não é um campo de jogo nivelado e certamente não é uma boa situação para se estar. E você precisa contratar um advogado, o que pode lhe custar um bom dinheiro.

Além disso, com base no processo acima do LinkedIn, você pode ver que os casos podem, sem dúvida, se tornar bastante complexos e muito amplos no escopo, mesmo que você “apenas tenha raspado um site”.

Os contra-argumentos típicos

As pessoas geralmente tentam defender suas atividades de web scraping ou web crawling minimizando sua importância. E elas o fazem tipicamente usando os mesmos argumentos repetidamente. Então, vejamos os mais comuns: <http://datascienceacademy.com.br/blog>

1. "Eu posso fazer o que quiser com dados publicamente acessíveis."

Falso. O problema é que o "arranjo criativo" dos dados pode ser protegido por direitos autorais. Assim, um site – incluindo suas páginas, design, layout e banco de dados **(sem falar em conteúdos pagos, que são obviamente protegidos por lei)** – pode ser protegido por direitos autorais, porque é considerado como um trabalho criativo. E se você raspar esse site para extrair dados dele, o simples fato de copiar uma página web na memória com o seu web scraper pode ser considerado uma violação de direitos autorais.

2. "Este é o uso justo!"

Esta é uma área cinzenta:

No processo Kelly v. Arriba Soft Corp. nos EUA, a corte descobriu que o mecanismo de busca de imagens Ditto.com fez uso justo das fotos de um fotógrafo profissional exibindo miniaturas delas.

Em outro processo, Associated Press vs. Meltwater U.S. Holdings, Inc., o tribunal considerou que o serviço agregador de notícias da Meltwater não fazia uso justo dos artigos da Associated Press, embora os artigos copiados só fossem exibidos como trechos dos originais.

3. "É o mesmo que o meu navegador já faz! Raspagem de um site não é tecnicamente diferente de usar um navegador web. Eu poderia coletar dados manualmente, de qualquer maneira!"

Falso. Os Termos de Serviço geralmente contêm cláusulas que proíbem rastreamento/captura/extração e uso automatizado de seus serviços associados. Você está legalmente vinculado por esses termos. Não importa que você possa obter esses dados manualmente.

4. "O pior que pode acontecer se eu quebrar seus Termos de Serviço é que eu posso ser banido ou bloqueado."

Esta é uma outra área cinzenta:

No processo do Facebook contra Pete Warden, o advogado do Facebook ameaçou Warden processá-lo se ele publicasse seu conjunto de dados composto de centenas de milhões de perfis do Facebook.

No processo do LinkedIn Corporation v. Michael George Keating, o LinkedIn impediu Keating de acessar o LinkedIn porque ele criou uma ferramenta que eles achavam que foi criada para raspar seu site. Eles estavam errados. Mas, no entanto, ele nunca conseguiu restaurar sua conta.

No processo do LinkedIn versus Robocog Inc., a Robocog Inc. foi condenada a pagar US\$ 40.000 ao LinkedIn por sua extração não autorizada do site.

5. "Isso é completamente injusto! O Google vem rastreando/raspando toda a web desde sempre!"

Verdade. Mas a lei aparentemente não tem nada a ver com justiça. É baseada em regras, interpretadas por pessoas.

6. "Se algum dia eu for processado, vou procurar um bom advogado para me defender."

Ok. Boa sorte então!

7. "Mas eu usei um script automatizado, então não entrei em nenhum contrato com o site."

Esta é uma área cinzenta:

[f](https://facebook.com/dsacademybr) (<https://facebook.com/dsacademybr>) [t](https://twitter.com/dsacademybr) (<https://twitter.com/dsacademybr>) [G+](http://google.com/+DataScienceAcademy) (<http://google.com/+DataScienceAcademy>) [yt](https://youtube.com/c/DataScienceAcademy) (<https://youtube.com/c/DataScienceAcademy>) [p](https://br.pinterest.com/dsacademybr/) (<https://br.pinterest.com/dsacademybr/>) [in](https://www.linkedin.com/company/data-science-academy/) (<https://www.linkedin.com/company/data-science-academy/>) [ig](https://www.instagram.com/dsacademybr/) (<https://www.instagram.com/dsacademybr/>)
No processo Internet Archive v. Suzanne Shell, o Internet Archive foi considerado culpado de quebrar um contrato ao copiar e arquivar páginas do site da Sra. Shell usando um crawler. Em seu site, a Sra. Shell disse em uma página dizendo que, assim que você copia o conteúdo de seu site, você entra em um contrato e você deve US\$ 5.000 por página copiada (!!!). As duas partes aparentemente chegaram a uma resolução amigável.

[INÍCIO \(HTTP://WWW.DATASCIENCEACADEMY.COM.BR/\)](http://www.datascienceacademy.com.br/)
No processo Southwest Airlines Co. vs. BoardFirst, LLC, a BoardFirst foi considerada culpada de violar um contrato de busca exibido no site da Southwest Airlines. A BoardFirst criou uma ferramenta que transferia automaticamente os cartões de embarque dos clientes da Southwest para oferecer-lhes melhores assentos.

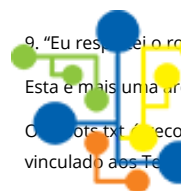
[BLOG \(HTTP://DATASCIENCEACADEMY.COM.BR/BLOG\)](http://datascienceacademy.com.br/blog)

8. "Termos de Serviço não são aplicáveis de qualquer maneira. Eles não têm valor legal." (DSA)

Falso: O escritório de advocacia Bingham McCutchen LLP publicou um artigo bastante extenso sobre esse assunto e afirma que:

Como é a regra geral com qualquer contrato, os termos de uso de um site geralmente serão considerados aplicáveis se forem mutuamente acordados pelas partes. **A falha do réu em ler esses termos é geralmente considerada irrelevante para a aplicabilidade de seus termos.** Um tribunal desconsiderou argumentos de que a conscientização dos termos de uso de um site não poderia ser imputada a uma parte que acessou o site usando uma ferramenta de rastreamento ou raspagem da Web que não é capaz de detectar, muito menos concordar com esses termos. Da mesma forma, um tribunal imputou o conhecimento dos termos de uso de um site a um réu que acessou repetidamente esse site usando essas ferramentas.

Em outras palavras, os Termos de Serviço serão legalmente aplicados dependendo do tribunal e se houver provas suficientes de que você estava ciente deles. Na Data Science Academy, por exemplo, os alunos só conseguem criar sua conta de acesso se aceitarem os Termos de Uso, logo, estão cientes das regras.



9. “Eu respeito os robots.txt e raspei o site a uma velocidade razoável, então não posso me meter em confusão, certo?”

Esta é mais uma área cinzenta.

Os robots.txt é reconhecido como uma “ferramenta tecnológica para impedir o rastreamento ou a raspagem indesejados”. Mas se você respeita ou não, você ainda está vinculado aos Termos de Serviço. (<http://datascienceacademy.com.br/blog>)

Os robots.txt é reconhecido como uma “ferramenta tecnológica para impedir o rastreamento ou a raspagem indesejados”. Mas se você respeita ou não, você ainda está vinculado aos Termos de Serviço.

10. “Tudo bem, mas isso é para uso pessoal. Apenas para minha pesquisa pessoal. Não vou republicá-lo, nem publicar qualquer conjunto de dados derivado, nem mesmo vendê-lo. Então, estou pronto para ir, certo?”

Esta é uma área cinzenta. Os Termos de Serviço muitas vezes proíbem a coleta automática de dados, para qualquer finalidade.

De acordo com o escritório de advocacia Bingham McCutchen LLP:

Os termos de uso dos sites incluem cláusulas que proíbem o acesso ou uso do site por rastreadores, raspadores ou outros robôs, inclusive para fins de coleta de dados. Os tribunais reconheceram as causas de ação por violações de contrato com base no uso de ferramentas de rastreamento ou raspagem da Web, violando tais disposições.

11. “Mas o site não tem robots.txt. Então eu posso fazer o que eu quero, certo?”

Falso. Você ainda está vinculado aos Termos de Serviço e o conteúdo é protegido por direitos autorais.

Conselho geral para seus projetos de web scraping ou web crawling

Com base no acima exposto, você pode certamente adivinhar que deve ser extremamente cauteloso com o web scraping e web crawling. Aqui estão alguns conselhos:

- Use uma API, se uma for fornecida, em vez de coletar dados.
- Respeite os Termos de Serviço.
- Respeite as regras do robots.txt.
- Use uma taxa de rastreamento razoável, ou seja, não bombardeie o site com solicitações. Respeite a configuração de tempo de acesso de rastreamento fornecida no robots.txt; se não houver nenhum, use uma taxa de rastreamento conservadora (por exemplo, 1 solicitação por 10 a 15 segundos).
- Identifique seu scraper ou crawler com uma string legítima do agente do usuário. Crie uma página que explique o que você está fazendo e porquê e faça um link para a página na string do agente (por exemplo, “MY-BOT (+ <https://seusite.com/mybot.html>)”)
- Se o Termo de Serviço ou o robots.txt impedirem que você rastreie ou raspe, peça uma permissão por escrito ao proprietário do site antes de fazer qualquer outra coisa.
- Não publique novamente seus dados rastreados ou copiados ou qualquer conjunto de dados derivado sem verificar a licença dos dados ou sem obter uma permissão por escrito do detentor dos direitos autorais.
- Se você duvida da legalidade do que está fazendo, não faça. Ou procure o conselho de um advogado.
- Não baseie toda a sua empresa na extração de dados. O site que você raspa pode, eventualmente, bloquear você.
- Finalmente, você deve suspeitar de qualquer conselho que você encontrar na internet (incluindo estes), então, por favor, consulte um advogado.
- Lembre-se de que empresas e indivíduos estão perfeitamente livres para processá-lo, por qualquer motivo que desejarem. Mas se você raspar/rastrear o site deles sem permissão e fizer algo que eles não gostem, você definitivamente se colocará em uma posição vulnerável.

Conclusão

Como vimos neste post, o web scraping e o web crawling não são ilegais por si mesmos. Eles podem se tornar problemáticos quando você joga no território de outra pessoa, em seus próprios termos, sem obter sua permissão prévia. O mesmo é verdade na vida real também, quando você pensa sobre isso.

Há muitas áreas cinzentas na lei em torno deste tópico, então o resultado é bastante imprevisível. Antes de se meter em problemas, certifique-se de que o que você está fazendo respeita as regras.

Faça sua busca

E, finalmente, a questão relevante não é “Isso é legal?” Em vez disso, você deve se perguntar “Estou fazendo algo que possa incomodar alguém? E estou disposto a assumir o risco financeiro de sua resposta?” Faça o que é certo. (<https://facebook.com/dsacademybr>) (<https://twitter.com/dsacademybr>) (<http://google.com/+DataScienceAcademy>) (<https://youtube.com/c/DataScienceAcademy>) (<https://br.pinterest.com/dsacademybr>) (<https://www.linkedin.com/company/data-science-academy>) (<https://www.instagram.com/dsacademybr>) Sinta-se livre para deixar seu comentário abaixo!

INÍCIO ([HTTP://WWW.DATASCIENCEACADEMY.COM.BR/](http://www.datascienceacademy.com.br/))

Equipe DSA

NOSSA HISTÓRIA ([HTTP://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/NOSSA-HISTORIA](http://www.datascienceacademy.com.br/pages/nossa-historia))

BLOG ([HTTP://DATASCIENCEACADEMY.COM.BR/BLOG](http://datascienceacademy.com.br/blog))

Referências:

TODOS OS CURSOS ([HTTP://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/TODOS-OS-CURSOS-DSA](http://www.datascienceacademy.com.br/pages/todos-os-cursos-dsa))

Capítulo 14 – Python Fundamentos Para Análise de Dados (<https://www.datascienceacademy.com.br/course?courseid=python-fundamentos>)

Is Web Scraping Legal? (<https://www.webharvy.com/articles/is-scraping-legal.html>)


Web scraping: legal, illegal, or does it depend on the circumstances? (<https://medium.com/enrique-dans/web-scraping-legal-illegal-or-does-it-depend-on-the-circumstances-3b4bc7b3fc3f>)

Web Scraping and Crawling Are Perfectly Legal, Right? (<https://benbernardblog.com/web-scraping-and-crawling-are-perfectly-legal-right/>)

LinkedIn: It's illegal to scrape our website without permission (<https://arstechnica.com/tech-policy/2017/07/linkedin-its-illegal-to-scrape-our-website-without-permission/>)

Beware of the Spiders: Web Crawling and Screen Scraping – the Legal Position (<https://parissmith.co.uk/blog/beware-spiders-web-crawling-screen-scraping-legal-position/>)

Compartilhe isso:



<https://twitter.com/dsacademybr> (<http://datascienceacademy.com.br/blog>)
<https://www.facebook.com/dsacademybr> (<http://datascienceacademy.com.br/blog/web-scraping-e-web-crawling-sao-legais-ou-ilegais/?share=facebook&nb=1>)
<https://plus.google.com/+DataScienceAcademy> (<http://datascienceacademy.com.br/blog/web-scraping-e-web-crawling-sao-legais-ou-ilegais/?share=google-plus-1&nb=1>)
<https://www.linkedin.com/company/data-science-academy> (<http://datascienceacademy.com.br/blog/web-scraping-e-web-crawling-sao-legais-ou-ilegais/?share=linkedin&nb=1>)
<https://tumblr.com/dsacademybr> (<http://datascienceacademy.com.br/blog/web-scraping-e-web-crawling-sao-legais-ou-ilegais/?share=tumblr&nb=1>)
<https://www.pinterest.com/dsacademybr> (<http://datascienceacademy.com.br/blog/web-scraping-e-web-crawling-sao-legais-ou-ilegais/?share=pinterest&nb=1>)

Relacionado

Bibliografia Big Data e Data Science
[\(http://datascienceacademy.com.br/blog/bibliografia-big-data-e-data-science/\)](http://datascienceacademy.com.br/blog/bibliografia-big-data-e-data-science/)
 outubro 3, 2017
 Em "Bibliografia"


10 Carreiras em Big Data e Data Science
[\(http://datascienceacademy.com.br/blog/10-carreiras-em-big-data-e-data-science/\)](http://datascienceacademy.com.br/blog/10-carreiras-em-big-data-e-data-science/)
 outubro 25, 2016
 Em "Big Data"

Fast Data - A Evolução do Big Data
[\(http://datascienceacademy.com.br/blog/fast-data-a-evolucao-do-big-data/\)](http://datascienceacademy.com.br/blog/fast-data-a-evolucao-do-big-data/)
 fevereiro 13, 2018
 Em "Analytics"

[\(/#facebook\)](https://www.facebook.com/dsacademybr)
[\(/#twitter\)](https://twitter.com/dsacademybr)
[\(/#google_plus\)](https://plus.google.com/+DataScienceAcademy)
[\(https://www.addtoany.com/share#url=http%3A%2F%2Fdatascienceacademy.com.br/blog/web-scraping-e-web-crawling-sao-legais-ou-ilegais%2F&title=Web%20Scraping%20e%20Web%20Crawling%20s%C3%A3o%20Legais%20ou%20Ilegais%2F\)](https://www.addtoany.com/share#url=http%3A%2F%2Fdatascienceacademy.com.br/blog/web-scraping-e-web-crawling-sao-legais-ou-ilegais%2F&title=Web%20Scraping%20e%20Web%20Crawling%20s%C3%A3o%20Legais%20ou%20Ilegais%2F)


NOTÍCIAS RELACIONADAS

4 Razões Porque Cientistas de Dados Continuam em Alta Demanda
[\(http://datascienceacademy.com.br/blog/4-razoes-porque-cientistas-de-dados-continuam-em-alta-demanda/\)](http://datascienceacademy.com.br/blog/4-razoes-porque-cientistas-de-dados-continuam-em-alta-demanda/)

 Carreira /
[\(http://datascienceacademy.com.br/blog/categoria/carreira/\)](http://datascienceacademy.com.br/blog/categoria/carreira/)
 Data Science /
[\(http://datascienceacademy.com.br/blog/categoria/data-science/\)](http://datascienceacademy.com.br/blog/categoria/data-science/)


Pensando em Migrar de Carreira? Por que não Gerente de Projetos de Big Data?

[\(http://datascienceacademy.com.br/blog/pensando-em-migrar-de-carreira-por-que-nao-gerente-de-projetos-de-big-data/\)](http://datascienceacademy.com.br/blog/pensando-em-migrar-de-carreira-por-que-nao-gerente-de-projetos-de-big-data/)

 Big Data /
[\(http://datascienceacademy.com.br/blog/categoria/big-data/\)](http://datascienceacademy.com.br/blog/categoria/big-data/)
 Carreira /
[\(http://datascienceacademy.com.br/blog/categoria/carreira/\)](http://datascienceacademy.com.br/blog/categoria/carreira/)
 Data Science /
[\(http://datascienceacademy.com.br/blog/categoria/data-science/\)](http://datascienceacademy.com.br/blog/categoria/data-science/)


Podcast DSA

[\(http://datascienceacademy.com.br/blog/podcast-dsa/\)](http://datascienceacademy.com.br/blog/podcast-dsa/)

 Carreira /
[\(http://datascienceacademy.com.br/blog/categoria/carreira/\)](http://datascienceacademy.com.br/blog/categoria/carreira/)
 Data Science /
[\(http://datascienceacademy.com.br/blog/categoria/data-science/\)](http://datascienceacademy.com.br/blog/categoria/data-science/)
 Podcast /
[\(http://datascienceacademy.com.br/blog/categoria/podcast/\)](http://datascienceacademy.com.br/blog/categoria/podcast/)

O Que é o TensorFlow Machine Intelligence Platform?

[\(http://datascienceacademy.com.br/blog/o-que-e-o-tensorflow-machine-intelligence-platform/\)](http://datascienceacademy.com.br/blog/o-que-e-o-tensorflow-machine-intelligence-platform/)

 Data Science /
[\(http://datascienceacademy.com.br/blog/categoria/data-science/\)](http://datascienceacademy.com.br/blog/categoria/data-science/)
 Deep Learning /
[\(http://datascienceacademy.com.br/blog/categoria/deep-learning/\)](http://datascienceacademy.com.br/blog/categoria/deep-learning/)
 Inteligência Artificial /
[\(http://datascienceacademy.com.br/blog/categoria/inteligencia-artificial/\)](http://datascienceacademy.com.br/blog/categoria/inteligencia-artificial/)
 Machine Learning /
[\(http://datascienceacademy.com.br/blog/categoria/machine-learning/\)](http://datascienceacademy.com.br/blog/categoria/machine-learning/)

[f \(https://facebook.com/dsacademybr\)](https://facebook.com/dsacademybr)
[t \(https://twitter.com/dsacademybr\)](https://twitter.com/dsacademybr)
[G+ \(https://plus.google.com/+DataScienceAcademy\)](https://plus.google.com/+DataScienceAcademy)
[yt \(https://www.youtube.com/c/DataScienceAcademy\)](https://www.youtube.com/c/DataScienceAcademy)
[p \(https://br.pinterest.com/dsacademybr\)](https://br.pinterest.com/dsacademybr)
[i \(https://www.instagram.com/dsacademybr\)](https://www.instagram.com/dsacademybr)
[in \(https://www.linkedin.com/company/data-science-academy\)](https://www.linkedin.com/company/data-science-academy)

INÍCIO ([HTTP://WWW.DATASCIENCEACADEMY.COM.BR/](http://www.datascienceacademy.com.br/))

NOSSA HISTÓRIA ([HTTP://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/NOSSA-HISTORIA](http://www.datascienceacademy.com.br/pages/nossa-historia))

BLOG ([HTTP://DATASCIENCEACADEMY.COM.BR/BLOG](http://datascienceacademy.com.br/blog))

TODOS OS CURSOS ([HTTP://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/TODOS-OS-CURSOS-DSA](http://www.datascienceacademy.com.br/pages/todos-os-cursos-dsa))

ENTRAR ([HTTP://WWW.DATASCIENCEACADEMY.COM.BR/?MSG=NOT-LOGGED-IN](http://www.datascienceacademy.com.br/?MSG=NOT-LOGGED-IN))

b i link b-quote u ul ol li code spoiler



Iniciar a discussão...

Subscribe ▼



Data Science Academy

(http://datascienceacademy.com.br/blog)



(https://facebook.com/dsacademybr)



(https://twitter.com/dsacademybr)



(http://google.com/+DataScienceAcademy)



(https://youtube.com/c/DataScienceAcademy)



(https://br.pinterest.com/dsacademybr)



(http://dsacademybr.tumblr.com/)



(https://www.linkedin.com/company/data-science-academy)



(https://www.instagram.com/dsacademybr)

RECEBA NOSSAS ATUALIZAÇÕES

ENVIAR

CATEGORIAS

[Analytics \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Analytics/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Analytics/)
[Banco De Dados \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Banco-De-Dados/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Banco-De-Dados/)
[Bibliografia \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Bibliografia/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Bibliografia/)
[Big Data \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Big-Data/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Big-Data/)
[Blockchain \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Blockchain/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Blockchain/)
[Business \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Business/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Business/)
[Carreira \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Carreira/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Carreira/)
[Chatbots \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Chatbots/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Chatbots/)
[Cloud Computing \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Cloud-Computing/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Cloud-Computing/)
[Data Lake \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Data-Lake/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Data-Lake/)
[Data Mining \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Data-Mining/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Data-Mining/)
[Data Science \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Data-Science/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Data-Science/)
[Deep Learning \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Deep-Learning/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Deep-Learning/)
[Engenharia De Dados \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Engenharia-De-Dados/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Engenharia-De-Dados/)
[f](https://facebook.com/dsacademybr) (https://facebook.com/dsacademybr) [t](https://twitter.com/dsacademybr) (https://twitter.com/dsacademybr) [G+](http://google.com/+DataScienceAcademy) (http://google.com/+DataScienceAcademy) [You](https://youtube.com/c/DataScienceAcademy) (https://youtube.com/c/DataScienceAcademy) [p](https://br.pinterest.com/dsacademybr) (https://br.pinterest.com/dsacademybr) [i](https://www.instagram.com/dsacademybr) (https://www.instagram.com/dsacademybr) [in](https://www.linkedin.com/company/data-science-academy) (https://www.linkedin.com/company/data-science-academy)

[Gestão De Projetos \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Gestao-De-Projetos/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Gestao-De-Projetos/)
[INÍCIO \(HTTP://WWW.DATASCIENCEACADEMY.COM.BR/\)](http://www.datascienceacademy.com.br/)
[IBM SPSS \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Ibm-Spss/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Ibm-Spss/)
[NOSSA HISTÓRIA \(HTTP://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/NOSSA-HISTORIA\)](http://www.datascienceacademy.com.br/pages/nossa-historia)
[BLOG \(HTTP://WWW.DATASCIENCEACADEMY.COM.BR/BLOG\)](http://www.datascienceacademy.com.br/blog)
[Inteligência Artificial \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Inteligencia-Artificial/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Inteligencia-Artificial/)
[TODOS OS CURSOS \(HTTP://WWW.DATASCIENCEACADEMY.COM.BR/PAGES/TODOS-OS-CURSOS-DSA\)](http://www.datascienceacademy.com.br/pages/todos-os-cursos-dsa)
[Java \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Java/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Java/)
[ENTRAR \(HTTP://WWW.DATASCIENCEACADEMY.COM.BR/?MSG=NOT-LOGGED-IN\)](http://www.datascienceacademy.com.br/?MSG=NOT-LOGGED-IN)
[Linguagem Python \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Linguagem-Python/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Linguagem-Python/)
[Linguagem R \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Linguagem-R/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Linguagem-R/)
[Linguagem SQL \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Linguagem-Sql/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Linguagem-Sql/)
[Machine Learning \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Machine-Learning/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Machine-Learning/)
[Podcast \(Http://Datascienceacademy.Com.Br/Blog/Categoria/Podcast/\)](http://Datascienceacademy.Com.Br/Blog/Categoria/Podcast/)



FACEBOOK

ARQUIVOS

outubro 2018 (<http://datascienceacademy.com.br/blog/2018/10/>) (4)

setembro 2018 (<http://datascienceacademy.com.br/blog/2018/09/>) (4)

agosto 2018 (<http://datascienceacademy.com.br/blog/2018/08/>) (5)

julho 2018 (<http://datascienceacademy.com.br/blog/2018/07/>) (10)

junho 2018 (<http://datascienceacademy.com.br/blog/2018/06/>) (6)

maio 2018 (<http://datascienceacademy.com.br/blog/2018/05/>) (6)

abril 2018 (<http://datascienceacademy.com.br/blog/2018/04/>) (3)

março 2018 (<http://datascienceacademy.com.br/blog/2018/03/>) (2)

fevereiro 2018 (<http://datascienceacademy.com.br/blog/2018/02/>) (6)

janeiro 2018 (<http://datascienceacademy.com.br/blog/2018/01/>) (5)

INÍCIO (<http://www.datascienceacademy.com.br/>)

NOSSA HISTÓRIA (<http://www.datascienceacademy.com.br/pages/nossa-historia>)

BLOG (<http://datascienceacademy.com.br/blog>)

TODOS OS CURSOS (<http://www.datascienceacademy.com.br/pages/todos-os-curso-dsa>)

ENTRAR (<http://www.datascienceacademy.com.br/?MSG=NOT-LOGGED-IN>)

SIGA NOSSAS REDES SOCIAIS

