

A primeira Escola presencial gratuita de Inteligência Artificial do Brasil



Hello
my name is

Wagner Santos



accenture



Digital
House>
Coding School

Escola
Livre de IA

Escola
Livre de IA

Aula Especial 28/03 – Covid-19

Data Repository by Johns Hopkins CSSE

Wagner Sant...

Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)

Total Confirmed: 591,802

Confirmed Cases by Country/Region/Sovereignty:

- 101,802 US
- 84,899 Italy
- 81,892 China
- 64,835 Spain
- 16,871 Germany
- 13,450 France
- 12,532 Iran
- 12,743 United Kingdom
- 12,938 Switzerland
- 9,339 Korea, South
- 8,662 Netherlands
- 7,857 Austria
- 7,388 Belgium
- 5,998 Turkey
- 4,493 Canada
- 4,208 Portugal
- 3,955 Norway
- 3,817 Brazil
- 3,163 Australia
- 2,948 Sweden
- 2,020 Israel

Last Updated at (M/D/YYYY)
3/27/2020, 7:19:54 PM

176 countries/regions

Total Deaths: 26,996

Deaths by Country:

- 9,134 deaths Italy
- 4,940 deaths Spain
- 2,174 deaths Hubei China
- 2,378 deaths Iran
- 1,995 deaths France
- 759 deaths United Kingdom
- 546 deaths Netherlands
- 366 deaths New York City New York US
- 347 deaths Germany
- 269 deaths Belgium
- 231 deaths

Total Recovered: 130,915

Recovered by Country:

- 61,752 recovered Hubei China
- 11,132 recovered Iran
- 10,059 recovered Italy
- 9,357 recovered Spain
- 6,658 recovered Germany
- 5,705 recovered France
- 4,528 recovered Korea, South
- 3,530 recovered Switzerland
- 1,337 recovered Guangdong China
- 1,269 recovered Henan China
- 1,222 recovered

Confirmed Daily Increase

<https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>

5:37 / 2:19:37

Zoom

Aula especial 03/2020 - Covid-19, a Matemática por trás da pandemia - Desafio #IAcontraCOVID19

741 views • Streamed live on Mar 28, 2020

143K subscribers

SUBSCRIBED

Escola Livre de Inteligência Artificial

Nessa aula, lançamos o desafio #IAcontraCOVID19. Conheça todos os detalhes no link <http://www.escolalivre-ia.com.br/IAco...>

Escola Livre de IA

Desafio #IAContraCOVID19

Inscrição para o desafio #IAcontraCOVID19

Esse formulário se destina a todas as pessoas que queiram participar do desafio #IAcontraCOVID19 organizado pela Escola Livre de IA (<http://escolalivre-ia.com.br/>)

*Obrigatório

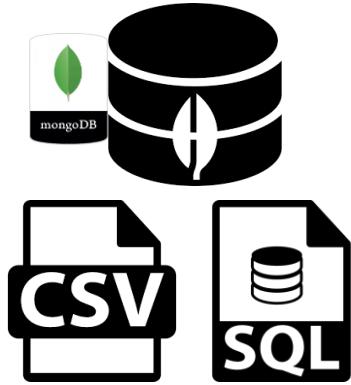
Qual é o nome do seu time? (É necessário apenas uma inscrição para todo o time) *

Sua resposta

Você tem um time completo? *

- Sim
- Não, e vou participar individualmente
- Não e gostaria de ajuda para formar um time

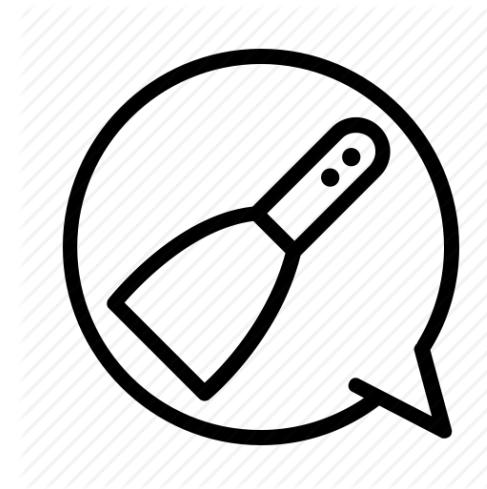
Como obtemos dados para nossos projetos de IA?



Bases “Prontas”



APIs



Web Scraping



ACESSO A DADOS ATRAVÉS DE APIs

API – Application Programming Interface



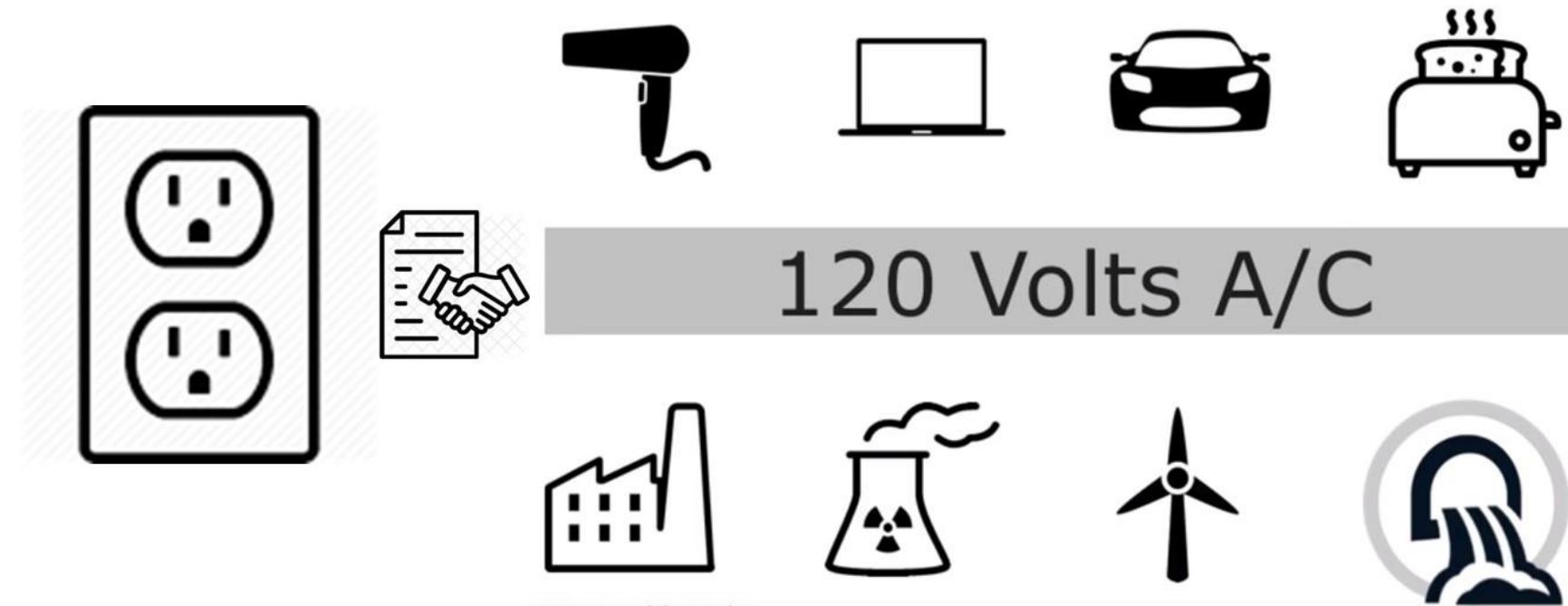
The screenshot shows a Wikipedia page with the following details:

- Header:** WIKIPÉDIA A encyclopédia livre
- Page Title:** Interface de programação de aplicações
- Page Content Summary:** Origem: Wikipédia, a encyclopédia livre.
Nota: Para outros significados de Api, veja [Api \(desambiguação\)](#).
Interface de Programação de Aplicações ([pt](#)) ou Interface de Programação de Aplicação ([pt-BR](#)), cujo acrônimo API provém do Inglês **Application Programming Interface**, é um conjunto de rotinas e padrões estabelecidos por um software para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação do software, mas apenas usar seus serviços.^[1]
- Page Footer:** Ler Editar Editar código-fonte Ver histórico Pesquisar na Wikipédia [ocultar]

https://pt.wikipedia.org/wiki/Interface_de_programa%C3%A7%C3%A3o_de_aplica%C3%A7%C3%A3o

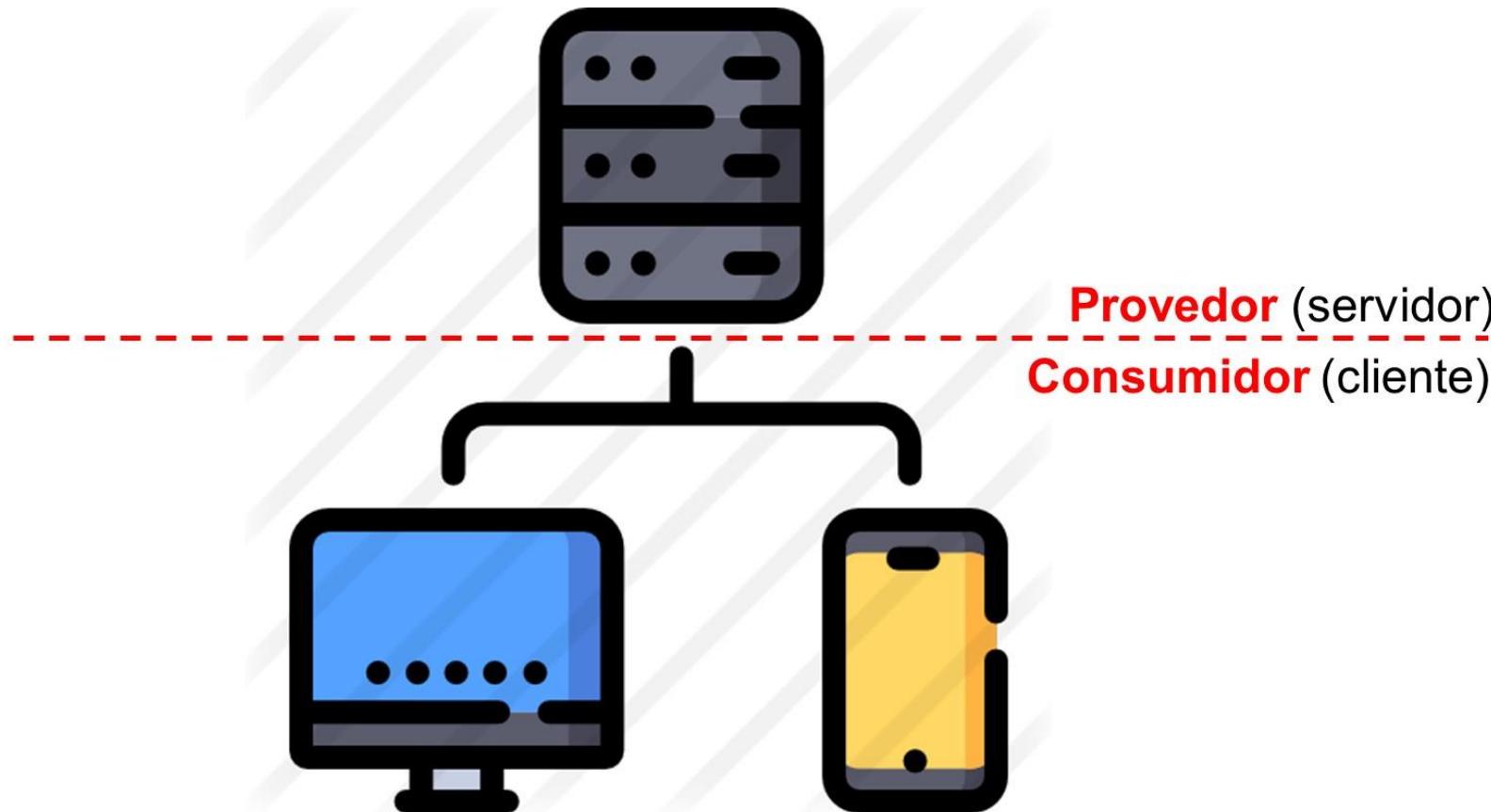
“Interface de Programação Online ou Interface de Programação de Aplicação cujo provém do inglês “Application Programming Interface”, é um conjunto de **rotinas** e **padrões** estabelecidos por um **software** para a utilização das suas **funcionalidades** por aplicativos que **não pretendem envolver-se em detalhes da implementação do software**, mas apenas usar seus serviços.”

API – Application Programming Interface



Podem se entender as APIs como uma espécie de **Contrato** que define regras e comportamentos esperados entre duas partes na utilização de um serviço.

API – Application Programming Interface



Pode-se entender a utilização de APIs sob dois pontos de vista:
do **Provedor** (servidor) e do **Consumidor** (cliente) do serviço
em questão.

API – Application Programming Interface



APIs são amplamente utilizadas em todos os contextos, não somente na obtenção de dados. Iremos explorar alguns **Casos de Uso** para que o aluno tenha uma **Visão Abrangente** de APIs, porém com uma ênfase **Prática** na **Coleta de Dados**.

API – Externalizar Serviços



amazon Rekognition



IBM Watson



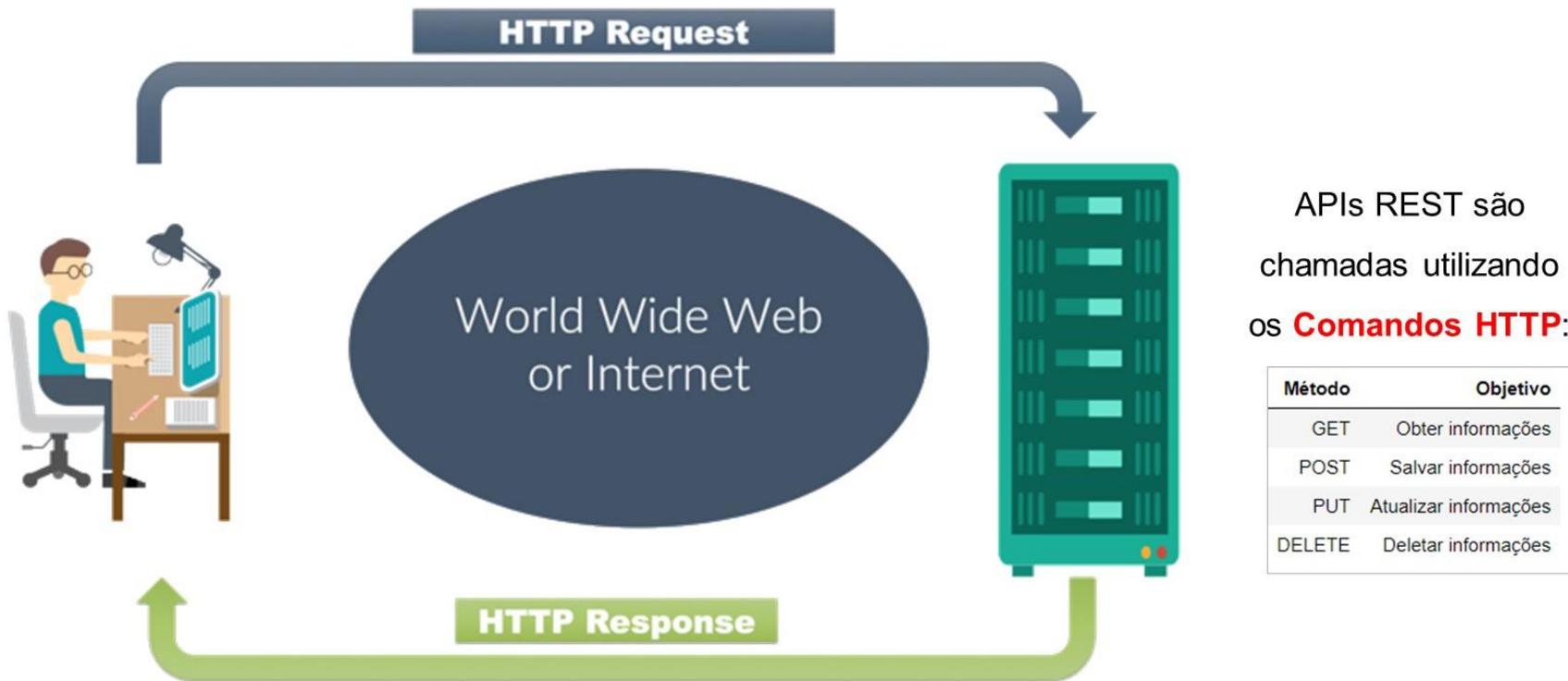
Sign up with Facebook

Sign up with Google



APIs também são utilizadas com frequência para externalizar/terceirizar serviços que não fazem parte do core da sua aplicação – em que existem provedores reconhecidos e competentes para a realização dos mesmos.

API – REST



Códigos de Retorno

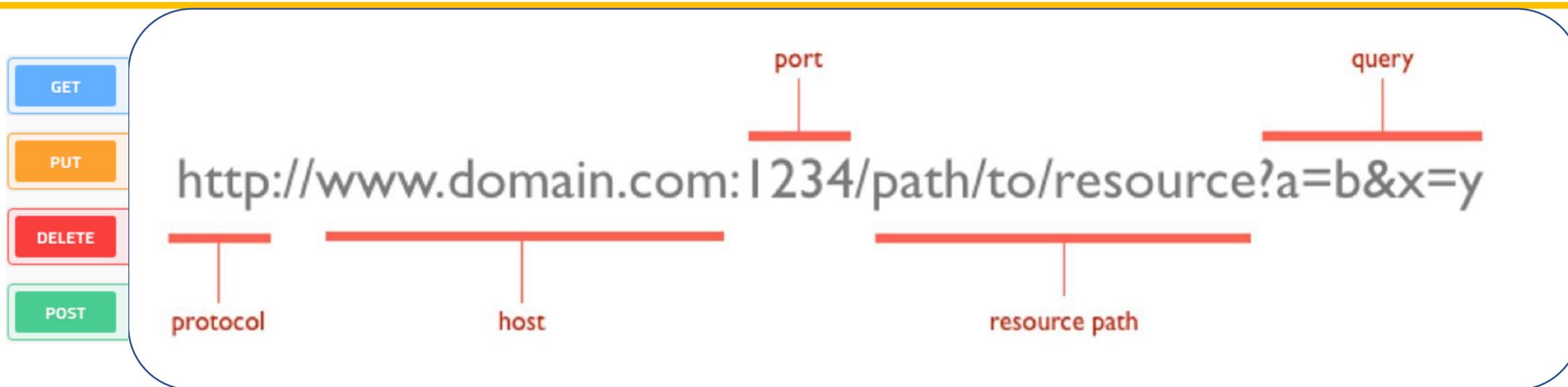
- 200 - tudo correu bem, e o resultado foi devolvido (se houver)
- 301- o servidor está redirecionando você para um terminal diferente. Isso pode acontecer quando uma empresa troca nomes de domínio ou um nome de terminal é alterado.
- 401- o servidor acha que você não está autenticado. Isso acontece quando você não envia as credenciais corretas para acessar uma API (falaremos sobre autenticação em uma postagem posterior).
- 400- o servidor acha que você fez um pedido incorreto. Isso pode acontecer quando você não envia os dados corretos, entre outras coisas.
- 403 - o recurso que você está tentando acessar é proibido - você não tem as permissões certas para visualizá-lo.
- 404 - o recurso que você tentou acessar não foi encontrado no servidor.

API REST – Comandos http

Método	Objetivo
GET	Obter Informações
POST	Inserir Informações
PUT	Atualizar Informações
DELETE	Apagar Informações



API REST – Comando + *Endpoint*



Como o servidor faz para saber o que está sendo pedido na solicitação?

Isso é especificado na URL (Uniform Resource Locator), uma espécie de caminho que indica onde um recurso pode ser encontrado.

protocol: Indica o protocolo que será utilizado para acessar o HTTP, FTP, HTTPS

host: Indica como encontrar na rede o servidor que tem o recurso.

port: Indica em qual porta TCP/IP está escutando o servidor
path: Indica o caminho para localizar o recurso dentro do servidor

query: Indica qual é a consulta que está sendo realizada

API – Application Programming Interface



Requests
http for humans

No Python vamos utilizar a biblioteca **Requests** para realizar as requisições HTTP. Ela fornece uma série de funcionalidades para facilitar a construção e entendimento dos comandos HTTP envolvidos. Por exemplo, ela permite que os parâmetros de requisições sejam informados através de dicionário.

Formatos Troca de Dados

XML (ontem)

```
1 <EmpRecord>
2   <Employee><-id>emp01
3   </-id>
4   <name>Alex</name>
5   <job>Developer</job>
6   <skills>python, C/C++, paskal</skills>
7 </Employee>
8 <Employee><-id>emp02
9 </-id>
10 <name>Bob</name>
11 <job>Tester</job>
12 <skills>lips, forton, REST
 APIs</skills>undefined</Employee>undefined</EmpRe
cord>
```

JSON (hoje)

```
1 {
2   "EmpRecord": {
3     "Employee": [
4       {
5         "-id": "emp01",
6         "name": "Alex",
7         "job": "Developer",
8         "skills": "python, C/C++, paskal"
9       },
10      {
11        "-id": "emp02",
12        "name": "Bob",
13        "job": "Tester",
14        "skills": "lips, forton, REST APIs"
15      }
16    ]
17  }
18 }
```

YAML (amanhã)

```
1 ---
2 EmpRecord:
3   Employee:
4     - "-id": emp01
5       name: Alex
6       job: Developer
7       skills: python, C/C++, paskal
8     - "-id": emp02
9       name: Bob
10      job: Tester
11      skills: lips, forton, REST APIs
12
```

Anos atrás a comunicação entre serviços se dava por XML usando o padrão SOAP. Na era da mobilidade **JSON** passou a ser largamente usado. Hoje, aos poucos está sendo substituído por YAML, que é similar mas mais legível pelos olhos humanos. A estrutura de um json se parece muito com os **Dicionários** em **Python**.

API – Application Programming Interface

The screenshot shows the ProgrammableWeb website. At the top, there's a navigation bar with links for 'API DIRECTORY', 'API NEWS', and 'ADD API & MORE'. Below the navigation is a search bar with the placeholder 'Search over 22,577 APIs and much more...'. The main content area features a large image of a human face overlaid with a circuit board pattern and the letters 'AI'. Below the image, the title '14 Top Artificial Intelligence APIs' is displayed, along with a brief description: 'Though there are concerns about AI, it is probably the hottest trend in computer science at the moment. Developers looking to add intelligence to applications can check out the ProgrammableWeb's Artificial Intelligence category for the best choices, including these fourteen very popular AI APIs.' To the right of the main content, there are sections for 'Today in APIs' (with a 'SUBSCRIBE' button), 'API UNIVERSITY' (with 'FEATURED' and 'LATEST' tabs), and 'FOR API PROVIDERS' (with links to 'What Are APIs and How Do They Work?', 'Guide to GraphQL: Understanding, Building and Using GraphQL APIs', and 'How Facebook Makes It Nearly Impossible for You To Out'). On the left side, there are two columns: 'LATEST APIs' (listing services like OKEx INDEX, OKEx Futures Trading, OKEx Websocket, OKEx Spot Trading, Uniswap, Lastline, Met Office Global Hourly Spot Data, REM REST, TalkPush, and kindred Spotsbook) and 'API NEWS' (listing stories like 'Spotify Testing Voice Controls Thanks to SiriKit', 'Smash Launches SDK for in-app Advertising', 'How to Transform an Air Traffic API to GeoJSON Is...', 'How to Build Twilio Appointment Reminders with...', 'How to Make a Heatmap with Mapbox GL JS', 'How to Build a Store Locator Using Mapbox GL JS', 'How to Switch from Google Maps to Mapbox', and 'How to Access Any RESTful API Using the R Language').

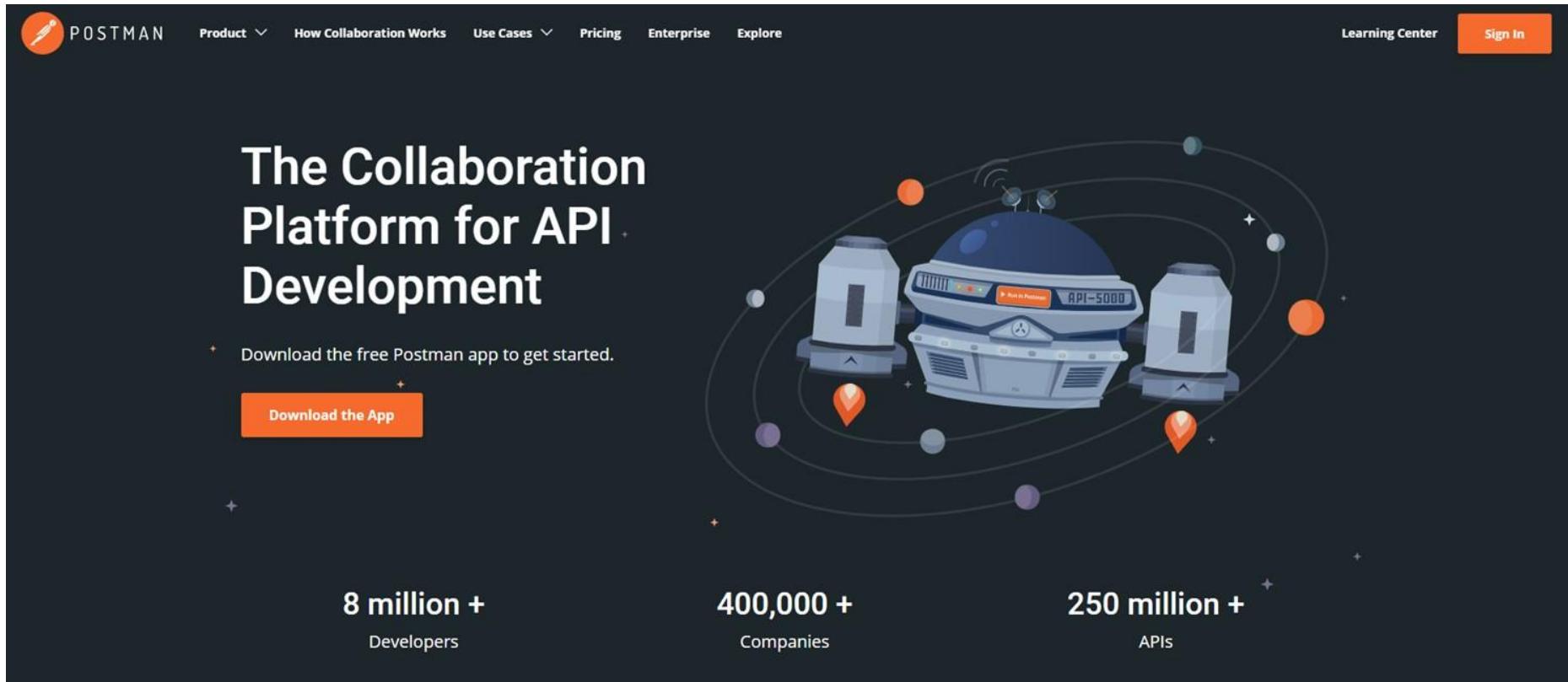
<https://www.programmableweb.com/>

The screenshot shows the RapidAPI website. At the top, there's a navigation bar with links for 'Create Organization', 'API Marketplace', 'My Apps', 'Add Your API', and 'Docs'. Below the navigation is a search bar with the placeholder 'Search APIs' and a 'All Categories' filter. The main content area features a section titled 'Find and Connect to Thousands of APIs' with the subtext 'One SDK. One API key. One dashboard.' Below this, there are sections for 'For Developers', 'For API Providers', and 'For Teams'. There are also buttons for 'Discover APIs', 'Test from the browser', 'Connect using code snippets', 'Manage from one dashboard', and 'Learn more'. A prominent section titled 'Unique APIs to Try' shows cards for various APIs: Chuck Norris (a free JSON API for hand curated Chuck Norris facts), Yoda Translator (a service that converts English to Yoda speak, where Yoda is a fictional character from the Star Wars movies), NasaAPI (a simple API to get NASA's collection of asteroids and space pictures), Opengraph.io (a simple API to extract URLs, descriptions, titles, images, products, from all sites without being blocked), Swapi (an API for Star Wars data you've ever wanted), Anilist (a platform to discover and share Anime and Manga), Open Weather Map (a weather forecast API for multiple cities), Weather (a current weather data API and Weather forecast API - Basic), National Weather Service API (a weather forecast API - NOAA), Visual Crossing Weather API (a visual weather forecast API that provides instant access to weather on the web), and US Weather by Zip (a service that provides current weather information for US cities by zip code).

<https://rapidapi.com/>

Vamos explorar agora alguns dos mais conhecidos repositórios / diretórios de APIs. O **ProgrammableWeb** e o **RapidAPI**.

Postman



The Collaboration Platform for API Development

Download the free Postman app to get started.

Download the App

8 million + Developers

400,000 + Companies

250 million + APIs

<https://www.getpostman.com/>

O **POSTMAN** é uma ferramenta *open source* muito útil em que você pode criar o seu próprio console e catalogar os serviços utilizados.

Acesso Dados - Portal da Transparência

The screenshot shows the Swagger UI interface for the API REST do Portal da Transparência do Governo Federal. The top navigation bar includes the Swagger logo, a dropdown menu for 'default (/v2/api-docs)', and a 'Explorar' button. The main title is 'API REST do Portal da Transparência do Governo Federal'. Below it, a subtitle reads 'API de serviços do Portal da Transparência do Governo Federal'. A note states 'Criado por Diretoria de Tecnologia da Informação - DTI' and provides a link 'Veja mais em <http://www.cgu.gov.br>'. It also includes links for 'Contate o desenvolvedor' and 'Decreto nº 8.777, de 11 de maio de 2016'. The main content area lists various API endpoints with options to 'Mostrar/Esconder', 'Listar operações', or 'Expandir operações'. These include: Acordos de Leniência, Benefício de Prestação Continuada (BPC), Bolsa Família, Cadastro de Expulsões da Administração Federal (CEAF), Cadastro Nacional de Empresas Inidôneas e Suspensas (CEIS), Cadastro Nacional de Empresas Punitas (CNEP), Contratos do Poder Executivo Federal, Convênios do Poder Executivo Federal, Despesas Públicas, Emendas parlamentares, and Entidades Privadas sem Fins Lucrativos Impedidas (CEPIM).

Endpoint	Mostrar/Esconder	Listar operações	Expandir operações
Acordos de Leniência			
Benefício de Prestação Continuada (BPC)			
Bolsa Família			
Cadastro de Expulsões da Administração Federal (CEAF)			
Cadastro Nacional de Empresas Inidôneas e Suspensas (CEIS)			
Cadastro Nacional de Empresas Punitas (CNEP)			
Contratos do Poder Executivo Federal			
Convênios do Poder Executivo Federal			
Despesas Públicas			
Emendas parlamentares			
Entidades Privadas sem Fins Lucrativos Impedidas (CEPIM)			

Primeiro iremos executar serviços desta API na **Console do Swagger** no Portal da Transparência.

Adicionalmente rodaremos o serviço no **Python**.

<http://www.portaltransparencia.gov.br/api-de-dados>

APIs – Rapid API

COVID-19

All APIs Collections More Options

Collections that match your search



Coronavirus COVID-19 APIs

[View more collections >](#)

COVID-19 Coronavirus API

Daily updated statistics from a leading university.

9.9 356ms 100%

COVID-19 data

Information about Coronavirus. Stats by country, collected from various sources.

9.6 612ms 99%

COVID-19

Get statistics for all countries about COVID-19 [FREE TO USE]

9.9 64ms 100%

Covid-19 - Dados Abertos

Acesso aos dados diariamente atualizados sobre o COVID e...

9.4 1015ms 100%

EndlessMedicalAPI

It is free to use in 2020. It already has COVID-19/SARS-CoV-2...

9.4 269ms 66%

Covid-19-CE

Informations about covid-19 in Ceará state, Brazil.

8.9 3ms 77%

COVID-19 Live Stats

Api to fetch live stats of covid-19 (corona virus) infection....

- - -

Coronavirus (COVID-19)

Regular coronavirus (COVID-19) global stats updated for...

- - -

Corona virus World API

COVID-19 World and India data

- - -

<https://rapidapi.com/search/COVID-19>

APIs – Programmable Web

APIs to Track Coronavirus COVID-19

Review, Coronavirus, API, Health, Medical

Mar 27, 2020 By [Wendell Santos](#), Editor



Editor's Note: Please check back as this article will continue to be updated as new APIs related to COVID-19 are made available.

Over the past month, as COVID-19 has proliferated, *ProgrammableWeb* has been tracking APIs that provide access to data related to the pandemic. The World Health Organization now reports that the virus has infected more than 460,000 people across more than 197 countries.

Promoted Listings

APIs can't help cure the disease but they can be used by developers to collect data about the outbreak, track its spread, and even produce data visualizations. In this article, we highlight a few APIs that let developers leverage the available data about the virus. We also included a couple of tools that use various APIs to track the outbreak.

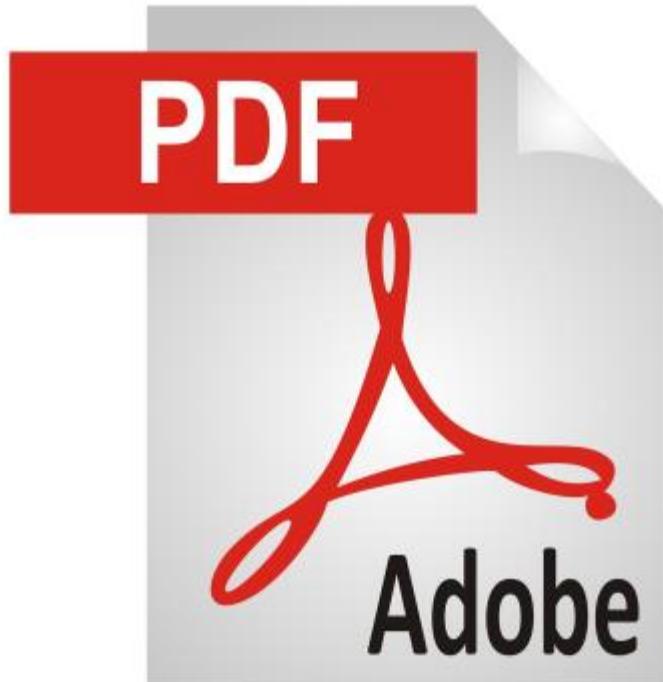
About Corona Covid-19 - The About Corona Covid-19 API provides statistics via REST API from The World Health Organization Situation Reports, Johns Hopkins CSSE, The U.S. Department of Health & Human Services, The National Health Commission of the People's Republic of China, The European Centre for Disease Prevention and Control, and China CDC Weekly.

This API retrieves data by country including population, number of cases confirmed, recovered, critical cases, deaths, recovered per death ratio, cases per million population, and more. The data is updated

<https://www.programmableweb.com/news/apis-to-track-coronavirus-covid-19/review/2020/03/27>

ACESSO A DADOS EM DOCUMENTOS PDF

PDF – *Portable Document Format*



A sigla inglesa **PDF** significa Portable Document Format (Formato Portátil de **Documento**), um formato de arquivo criado pela empresa Adobe Systems para que qualquer **documento** seja visualizado, independente de qual tenha sido o programa que o originou

<https://www.significados.com.br/pdf/>

Vídeo Explicativo – COVID-19



pdfrw
<https://github.com/pmaupin/pdfrw>

PyPDF2
<https://pythonhosted.org/PyPDF2/>

XPDF-PYTHON
<https://pypi.org/project/xpdf-python/>

PDFQuery
<https://github.com/jcushman/pdfquery>

Slate
<https://github.com/timClicks/slate>



Camelot & Excalibur

CC-BY-NC-SA 4.0

Extracting tabular data from PDFs with Camelot & Excalibur

Vinayak Mehta

pyconau Sydney 2019

"Extracting tabular data from PDFs with Camelot & Excalibur" - Vinayak Mehta (PyCon AU 2019)

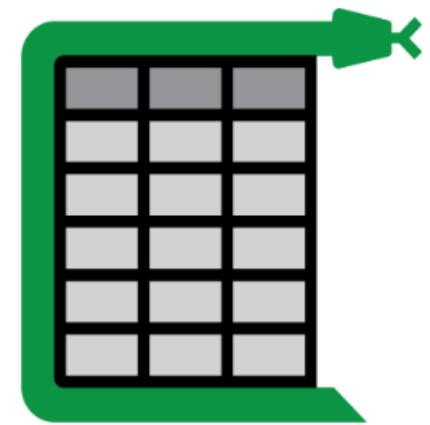
2,224 views • Aug 2, 2019

46 likes, 1 dislike, SHARE, SAVE, ...

Up next: Dimitar Naydenov - Extracting Tabular Data from PDFs with... EuroPython Conference

AUTOPLAY

<https://youtu.be/99A9Fz6uHAA>



<https://camelot-py.readthedocs.io/en/master/>

Camelot & Excalibur

Excalibur Files Rules Jobs [Support Excalibur on OpenCollective!](#)

Excalibur

A web interface to extract tabular data from PDFs

[Upload PDF](#)

Page numbers (example inputs: 1,3 or 5-8 or 1-end or all)
Comma-separated page numbers to extract tables from. [Upload](#)

Previous Uploads

#	Filename	Uploaded at	Actions
1	twotables_2.pdf	2018-11-25T20:13:26	Extract Again Download

<https://github.com/camelot-dev/excalibur>

Camelot & Excalibur

Using Excalibur

Note: You need to [install ghostscript](#) before moving forward.

After [installing Excalibur with pip](#), you need to initialize the metadata database using:

```
$ excalibur initdb
```

And then start the webserver using:

```
$ excalibur webserver
```

That's it! Now you can go to <http://localhost:5000> and start extracting tabular data from your PDFs.



ACESSO A DADOS ATRAVÉS DE WEB *SCRAPING*

Web Scraping



Scraping é uma forma de obter **Dados** através da **Varredura Automática** de páginas **Web**, utilizando o que se chama de *web crawlers* ou *spiders*.

Web Scraping



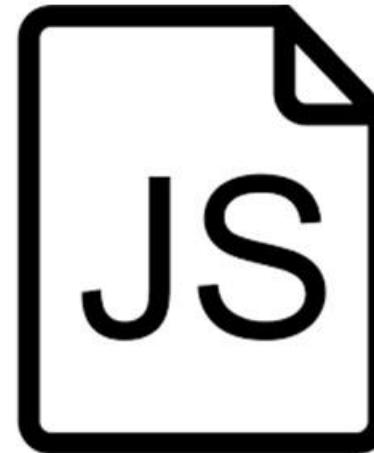
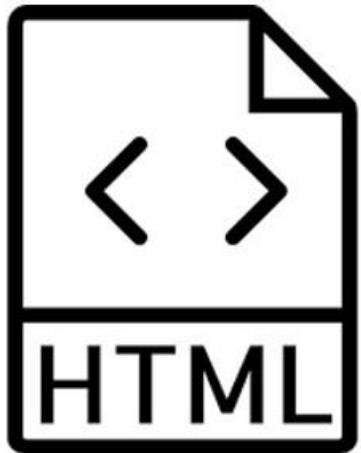
Embora a prática não seja ilegal, existem contextos em que empresas e indivíduos já foram condenados por realizar scraping. Além disso, deve-se levar em consideração que muitas empresas e sites buscam se defenderativamente da prática.

Web Scraping



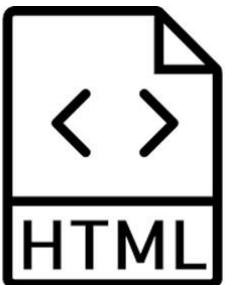
No entanto existem diversos contextos em que o scraping é realizado de maneira aberta e que os sites de fato desejam que os robôs/crawlers passem por suas páginas, como os grandes mecanismos de busca (Google, Bing, Baidu, etc). Inclusive a convenção do documento **ROBOTS.TXT** é uma maneira pela qual sites buscam melhorar seu **SEO** (Search Engine Optimization)

Como páginas Web funcionam?



Para entender scraping se faz necessário alguns conhecimentos básicos de **Como Páginas Web Funcionam** e como elas são apresentadas.

HTML



TAGS

- `<html>` - define o começo e fim do documento
- `<body>` - define a parte visível do documento
- `<h1>...<h6>` - Headings, cabeçalhos
- `<p>` - parágrafo
- `` - link
- `` - imagem
- `
` - quebra de linha

ATRIBUTOS

- `style, lang, width, height, title`

HTML

Abreviação para a expressão

inglesa HyperText Markup

Language, que significa Linguagem de Marcação de Hipertexto é uma linguagem de marcação utilizada na construção de páginas na Web.

Documentos HTML podem ser interpretados por navegadores.

CSS



CSS

Cascading Style Sheets (CSS) é um mecanismo para adicionar estilo (cores, fontes, espaçamento, etc.) a um documento web[1].

TAGS

- *<div> - divisão entre estilos na página e marca para a carga de objetos*
- *<style> - define estilos no documento*

ATRIBUTOS

- *Class, id, etc*

O código CSS pode ser aplicado diretamente nas tags ou ficar contido dentro das tags `<style>`.

Também é possível, em vez de colocar a formatação dentro do documento, criar um link para um arquivo CSS que contém os estilos. Assim, quando se quiser alterar a aparência dos documentos vinculados a este arquivo CSS, basta modifica-lo.

xpath

<xpath>

O XPath é uma linguagem de consulta, que é utilizada para identificar e percorrer tags/elementos específicos em documentos XML/HTML.

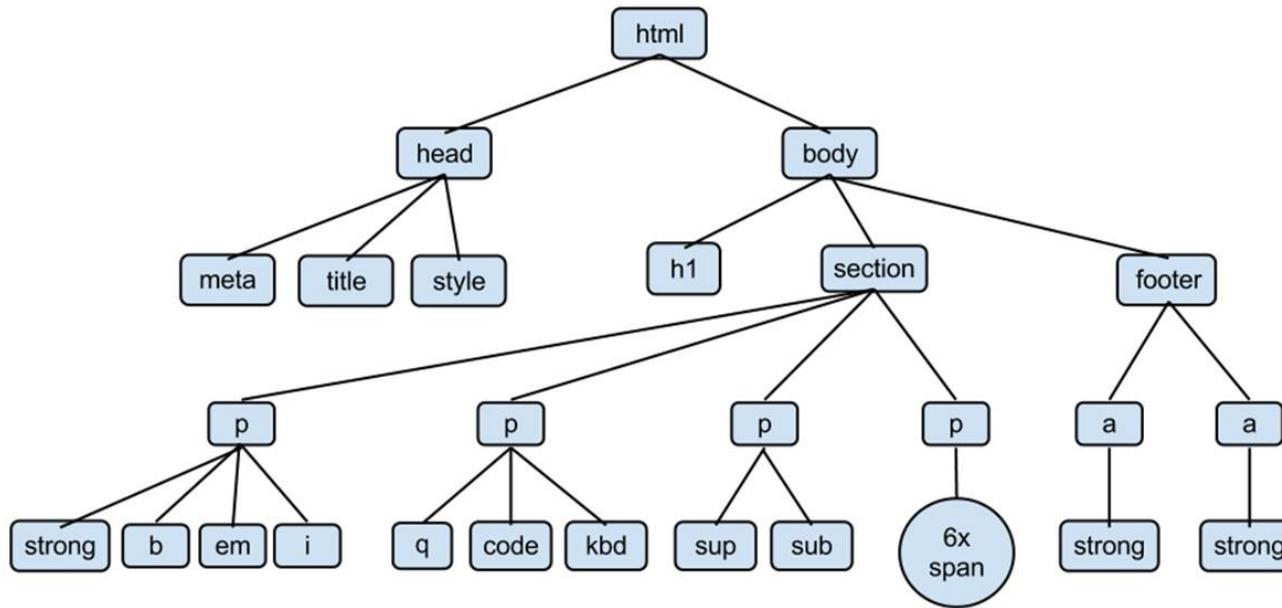
Comandos básicos:

/ - avança um nível

[n] - indica qual elemento utilizar se houver mais de 1 no mesmo nível

// - traz todo níveis abaixo

xpath



Uma página completa pode ser vista então como uma estrutura que se assemelha a uma “árvore” (em termos computacionais) – com nós que possuem filhos, que permitem uma busca em profundidade dos dados. O Xpath nos ajuda a definir estes caminhos de uma maneira assertiva.

Ferramentas *Scraping* Python

tesseract



Ferramenta que utiliza OCR (*Optical Character Recognition*) para a leitura dos dados, com extensões e aplicações em Machine Learning e Inteligência Artificial



É um scraper projetado para trabalhar em alta escala com um grande volume de crawlers. No entanto a curva de aprendizado para utilizá-lo é considerável.



Embora não tenha sido concebida especificamente para o scraping (é um automatizador de testes em páginas Web), é extremamente útil para capturar os dados. É o único que consegue pegar dados gerados por Javascript.

BeautifulSoup

Pode ser considerado o scraper “padrão” – o mais utilizado e conhecido pela comunidade. Permite implementar, com grande facilidade, a raspagem de dados.



KEEP
CALM
AND
stay
at home!



obrigado