

```
<html>
```

```
  <head>
```

```
    <h1>Intro to Web Scraping</h1>
```

```
  </head>
```

```
  <body>
```

```
    <a>By Anderson Lam & Carlos E. Ortega</a>
```

```
  </body>
```

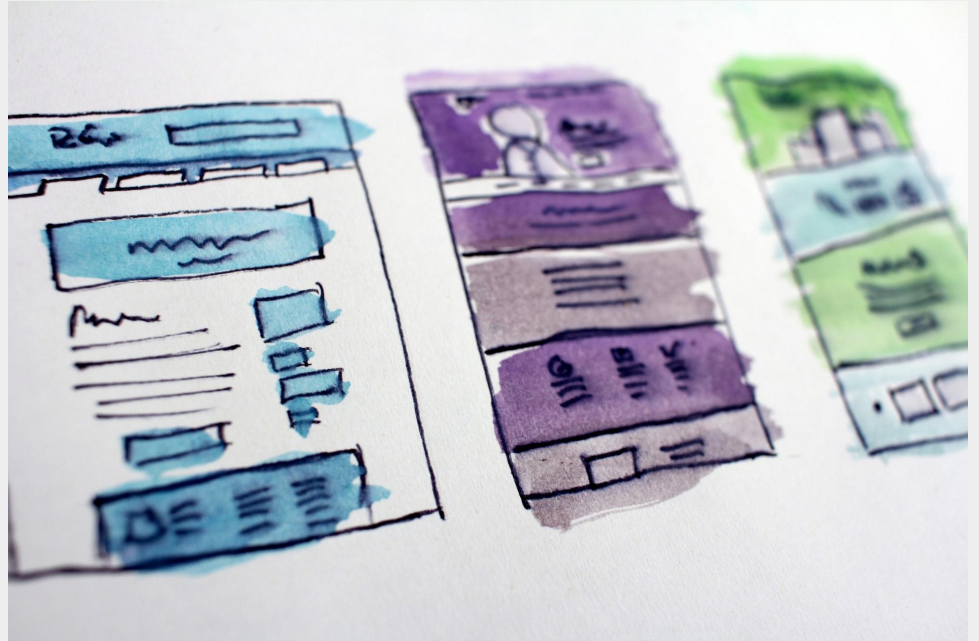
```
</html>
```

What is Web Scraping?

Websites are built on HTML.

Web Scraping gives us the ability extract elements from these pages:

- Text
- Images
- Tables



Web Scraping Use Cases



Nicole Bengiveno/The New York Times

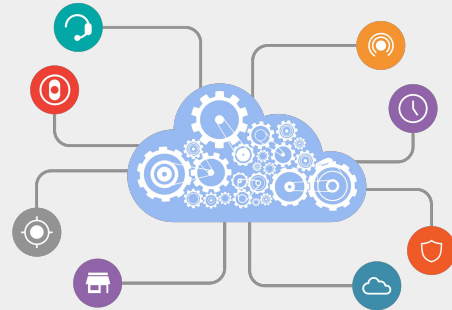


reddit

Web Scraping Technologies



Scrapy



HTML Basics


```
<element attribute="value">element content</element>
```

```
<a href="/academics/resources/peer-consulting">Data Peer Consulting</a>
```

`<a>` HTML Element

`href=""` Attribute

`Data Peer Consulting` Element Content

Other Common HTML Elements

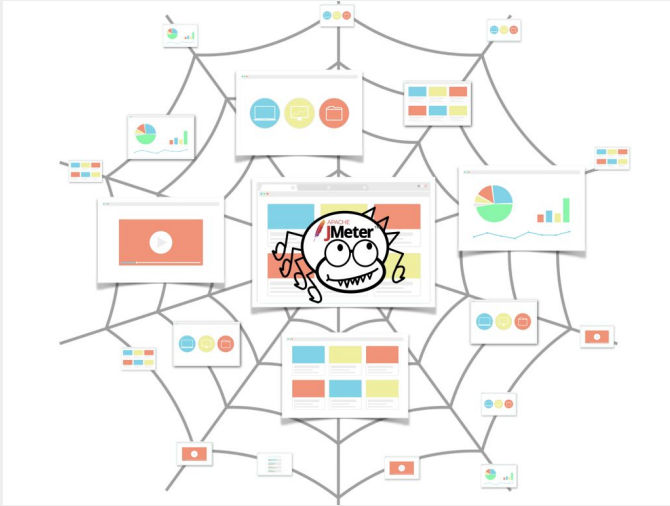
`div, span, p`

Hands-On

bit.ly/webscraping-pc

Letting The Scraper Loose:

Web Crawling



Scrapy