



# Intro to Web Scraping

---

## What is Web Scraping?

Web scraping is a technique to access and extract information from a website.

---

## Why Web Scraping?

- Collect data that is not available through an API
  - Automate the process of collecting data (Save time and effort)
  - Collect data that is updated dynamically
- 

## Is Web Scraping Legal?

- \* Web scraping is legal as long as you're not violating the website terms of service
- \* If the website doesn't have a terms of service, you should be fine  
(web scraping consumes server resources for the host website)



# Intro to Web Scrapping

---

## The Components of a Web Page

- \* HTML: The markup language that defines the structure of a web page.
- \* CSS: The language that defines the style of a web page.
- \* JavaScript: The language that defines the behavior of a web page.

## HTML Tree Structure

```
<html>
  <head>
    <title>Page Title</title>
  </head>
  <body>
    <h1>Heading 1</h1>
    <p>Paragraph 1</p>
    <p>Paragraph 2</p>
    <p>Paragraph 3</p>
  </body>
</html>
```



# Intro to Web Scraping

---

## CSS Classes and IDs

- \* CSS classes and IDs are used to identify specific elements on a web page.
- \* CSS classes are used to identify multiple elements on a web page.
- \* CSS IDs are used to identify a single element on a web page.

- Class:

```
<p class="class1"> Paragraph 1 </p>  
<h2 class="class1"> Heading 2 </h2>
```

- ID:

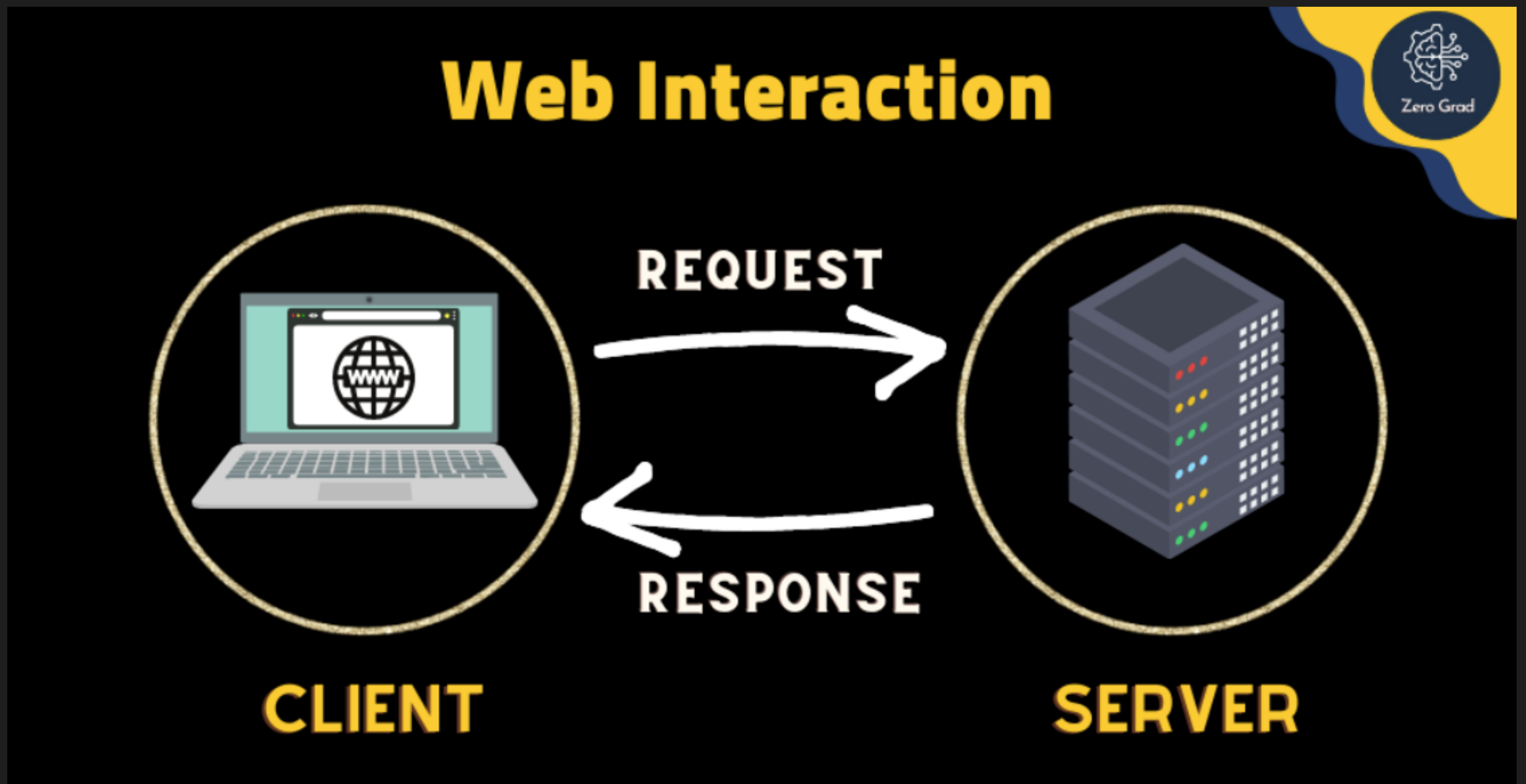
```
<p id="id1">Paragraph 1</p>
```



# Intro to Web Scrapping

---

How Does Web Scrapping Work?



\* We're essentially doing the same thing a web browser does – sending a server request with a specific URL and asking the server to return the code for that page.



# Intro to Web Scrapping

## Actual vs. Visual Response



```
<div class="css-laomuu"> -- $0
  <h2 class="css-m604qf">
    <a target="_blank" rel="noreferrer" class="css-o171kl" href="/jobs/p/63GD
    VXe1mip9-Machine-Learning-Engineer-Edentech-Giza-Egypt?o=2&l=sp&t=sj&a=ma
    chine_learning|search-v3|hpb">Machine Learning Engineer</a>
  </h2>
  <div class="css-d7j1kk">
    <a href="https://wuzzuf.net/jobs/careers/Edentech-Egypt-78060" target="_b
    lank" rel="noreferrer" class="css-17s97q8">Edentech -</a>
    <span class="css-5wys0k">...</span>
    <div class="css-do6t5g">...</div>
  </div>
</div>
```



**RESPONSE**

**Machine Learning Engineer**  
Edentech · Dokki, Giza, Egypt  
1 month ago  
**Full Time**

Experienced · 3 - 7 Yrs of Exp · IT/Software Development · Engineering · Telecom/Technology  
· Computer Engineering · Engineering · Computer Science · Information Technology (IT) · Machine Learning · SQL



**BROWSER**

- \* But unlike a web browser, our web scraping code won't interpret the page's source code and display the page visually.
- \* We write a code that sends a request to the server that's hosting the page we specified.
- \* The server will return the source code – HTML – for the page
- \* We can then parse the HTML to extract the data we want.

# Contact Us



[zero-grad.com](https://zero-grad.com)



[+201018486146](https://wa.me/201018486146)



[Zero Grad](https://www.facebook.com/ZeroGrad)



[Zero Grad](https://www.youtube.com/ZeroGrad)



[Zero Grad](https://www.linkedin.com/company/ZeroGrad)