

Web Scraping

Beginner's Guide To Data Mining



A circular profile picture of a man with dark hair and a beard, wearing a white t-shirt and a dark jacket.

hossam-eldin



@codewello

Web Scraping in AI Era

In today's age of AI, the most important thing isn't the AI itself, but the data you use to train it. This data is the first step! But how do you get your own data, instead of using someone else's? That's where web scraping comes in!

A circular profile picture of a man with dark hair and a beard, wearing a white t-shirt and a dark jacket.

hossam-eldin



@codewello

Step One

Locate the website you want to collect data from. It will really help if you can read HTML, as it will assist you in identifying which elements contain the data you need. However, don't worry if you're not familiar with HTML, as this part of the process can be automated if you know the next steps.



hossam-eldin



@codewello

Step Two

Using Python and BeautifulSoup, an awesome beginner-friendly library, it helps you extract the data you want from HTML and lets you do what you want with it

```
response = requests.get(url)
soup = BeautifulSoup(response.content, "lxml")
```

Requests will call the page URL, and BeautifulSoup with the lxml parser will handle dealing with the HTML.



hossam-eldin



@codewello

Step Three

With Soup up and running, we can find what you want inside our HTML.

There are two main functions: find and find_all.

The find function will get you the first element it finds and then stop, but the find_all function will search the entire HTML for this element and list it for you.

```
recipes_box = soup.find("div", class_="no-right-padding")
recipes = recipes_box.find_all("li")
```



hossam-eldin



@codewello

In the final step, you can put the data you collected in any useful format, such as JSON.

```
with open("recipes.json", "w") as json_file:  
    json.dump(recipes, json_file, indent=4)
```

Or CSV Format

```
with open("recipes.csv", "w", newline="", encoding="utf-8") as csvfile:  
    fieldnames = recipe_info.keys()  
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)  
    writer.writeheader()  
    writer.writerows(recipes_data)
```

A circular profile picture of a man with dark hair and a beard, wearing a white t-shirt and a dark jacket.

hossam-eldin

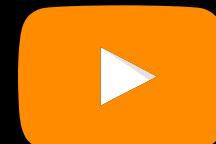


You can find the full guide link in the post, along with a detailed explanation in the video.



A circular profile picture of a man with dark hair and a beard, wearing a dark jacket over a white shirt.

hossam-eldin



@codewello

Thanks !!

if you enjoyed this, please:

- **Follow me** for more badass content
- **Share** this post so others can learn as well
- **Hit the notification** so you don't miss a thing

And most importantly...

Keep on learning



Happy Coding