

Web Scraping – Chapter 2, Lesson 1: Using `find()` and `find_all()`

What You'll Learn

In this lesson, you'll explore how to navigate HTML using `find()` and `find_all()` to extract specific data from websites with precision.

What Are `find()` and `find_all()`?

These are core BeautifulSoup functions used to locate HTML tags within a web page.

- `find()` retrieves the **first match** for a specified tag.
- `find_all()` returns **all matching elements** as a list.

They are powerful tools for selecting and filtering HTML content based on tags and attributes like `class`, `id`, or `href`.

Key Concepts Covered

- How to inspect a webpage's HTML structure to target specific data
 - The difference between grabbing one item vs. many using `find()` vs `find_all()`
 - Filtering tags using attributes like class names
 - Extracting clean text from within HTML tags
 - Handling whitespace and formatting issues with basic cleanup methods
-





Real-World Applications

- Pulling product titles or prices from an online store
 - Extracting article snippets or author names from blog pages
 - Gathering structured data like names or categories from repetitive tag elements
-

Best Practices

- Use `find()` when expecting a single, specific result
 - Use `find_all()` when scraping multiple similar elements
 - Always preview your results to ensure accuracy
 - Clean up extracted text to remove excess formatting and whitespace
-

Recap

-  Use `find()` for one result and `find_all()` for many
-  Combine tag selection with attributes for precision
-  Extract and clean text for usable results
-  Lays the groundwork for turning HTML into clean data