HACKTIV8

//01



Hacktiv8 DS Curriculum Team Phase 0 Learning Materials Hacktiv8 DS Curriculum Team

Definition	04
Basic HTML Structure	06
Tools	07

Contents

HACKTIV8 objectives

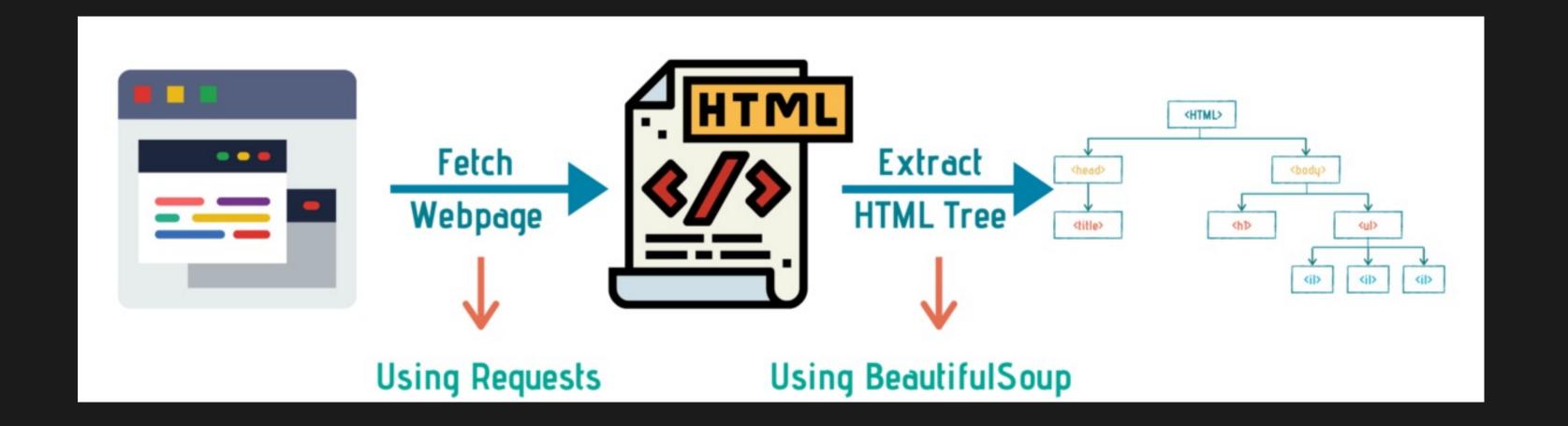
grasp the concept of web scraping able to analyze html structure used for scraping able to extract data using Beautifulsoup

HACKTIV8 Definition

Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites. The web scraping software may directly access the World Wide Web using the Hypertext Transfer Protocol or a web browser.

Hacktiv8 DS Curriculum Team

Definition



HACKTIV8 HTML Structure

```
<HTML>
        <head>
                   Inner HTML text
            <title>My cool title </title>
         </head>
tags
         <body>
            <h1>This is a Header</h1>
            ul id="list" class="coolList"
              item 1
              item 2
                                        attribute value
                                attributes
              item 3
            </body>
  </HTML>
```

HACKTIV8 Tools

Beautifulsoup is a Python library for pulling data out of HTML and XML files. It works with your favorite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. It has limited that cannot be used for dynamic websites that include Javascript. However, we will use Selenium to cover this drawback.

Hacktiv8 DS Curriculum Team Hacktiv8 DS Curriculum Team

//8

External References

Colab Link _____ <u>Visit Here</u>