

Interview Questions Answers

Q1.What is web scraping and how does BeautifulSoup help in it?

- Web scraping is the process of extracting data from websites. BeautifulSoup is a Python library that helps parse HTML and XML documents, making it easier to navigate and search through the document structure to extract the data you need.

Q2.What are the ethical concerns in web scraping?

- Respecting website terms of service, not overloading servers with requests, not scraping personal or sensitive data, checking robots.txt for scraping permissions, and using scraped data appropriately.

Q3.What's the difference between find() and find_all() in BeautifulSoup?

- find() returns the first matching element, while find_all() returns a list of all matching elements. find() is useful when you know there's only one match, while find_all() is for multiple matches.

Q4.How do you handle pagination in web scraping?

- By identifying the pattern in pagination URLs (often a page parameter) and looping through pages, or by locating and following "Next" buttons. Always include delays between page requests.

Q5.How would you extract information from a tag with multiple attributes?

- You can use CSS selectors or pass multiple attributes to find/find_all. For example:

```
soup.find('div', {'class': 'job', 'id': '123'})
```

```
# or
```

```
soup.select('div.job#123')
```