

Course : Web Scraping, harvesting data from the web with Python

Practical course - 4d - 28h00 - Ref. WPY

Price : 2470 CHF E.T.

★★★★☆ 4,7 / 5

Do you need to extract, manipulate, verify or archive data from the web? To be more efficient, automate your harvesting and broaden the scope of these operations. Opt for web scraping with Python: its scraping libraries and simplicity will quickly industrialize your processes.

Teaching objectives

At the end of the training, the participant will be able to:

- ✓ Master the basics of the Python language
- ✓ Advanced programming in Python
- ✓ Get an overview of the main Python libraries available for managing all types of site data
- ✓ Select the right Python library for your web scraping project and be able to implement it
- ✓ Automate large-scale web scraping with scripts

Intended audience

Developers, consultants, analysts, project managers and anyone wishing to automate web data collection.

Prerequisites

Basic knowledge of algorithms or programming skills. Knowledge of HTML and CSS is recommended.

Practical details

Hands-on work

Numerous exercises are used to illustrate and assimilate the topics.

Teaching methods

The trainer uses active teaching methods, including feedback and demonstrations, to enable participants to quickly put the system into practice.

Course schedule

PARTICIPANTS

Developers, consultants, analysts, project managers and anyone wishing to automate web data collection.

PREREQUISITES

Basic knowledge of algorithms or programming skills. Knowledge of HTML and CSS is recommended.

TRAINER QUALIFICATIONS

The experts leading the training are specialists in the covered subjects. They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

ASSESSMENT TERMS

The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more. Participants also complete a placement test before and after the course to measure the skills they've developed.

1 The basics of the Python language

- The main types of variable.
- Perform operations and work on strings.
- Data structures in Python.
- Understand mutable and immutable types.
- Control structures.returnchariot
- Creating and using functions
- Read and write text or binary files.

Hands-on work

Familiarize yourself with the language, control structures and data manipulation. Use functions. Create a game.

2 Advanced programming in Python

- Lambda anonymous functions.
- Understand the usefulness of generators and know how to create them.
- Handle errors with exception handling.
- Create new data types with object programming.
- Inheritance concepts in object programming.
- Use a library.
- Select and evaluate open source libraries

Hands-on work

Write a generator. Create and manipulate an object with attributes and methods.

3 Extract data via Restful APIs

- Connect to a web API.
- Perform different types of HTTP requests.
- JSON format, analyze extracted data.
- Search for specific data in text using regular expressions.
- Manage connection errors.

Hands-on work

Extract the right data from text. Harvest and exploit geolocation information via a RESTful API. Evaluate the results of a real API connection.

4 Capture data tables from the Internet and process them with Pandas

- HTML and CSS basics.
- Pandas fundamentals.
- Import and export data in different formats.
- Handling data with Pandas.
- Scrape data tables from the web.

Hands-on work

Extract up-to-date digital data from the Internet. Process and archive collected data.

TEACHING AIDS AND TECHNICAL RESOURCES

- The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
- At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
- A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@orsys.fr to review your request and its feasibility.

5 Scrape websites with BeautifulSoup

- Easy Scraping: BeautifulSoup.
- Implement the parser.
- Search the parser tree.

Hands-on work

Scrape websites with BeautifulSoup. Quickly find useful data, save it with the relevant information.

6 Automate large-scale harvesting with Scrapy

- Basic operation of the Scrapy framework.
- Identify content to be scrapped.
- Structuring a spider.
- Automate Scrapy harvesting and record results.
- Evaluate campaign performance.

Hands-on work

Crawl web articles and collect relevant data with Scrapy.

Dates and locations

REMOTE CLASS

2026 : 9 June, 8 Sep.