Web Scraping and Data Extraction with Python Training Course

Professional Development Training Course Outline

Category	Research and Data Analysis	**Duration**	5 Days
Base Fee	$1,500 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

In today’s data-driven world, extracting and analyzing online information is crucial, especially when dealing with sensitive, complex, or controversial topics. Web Scraping and Data Extraction with Python Training Course empowers participants to ethically harness the power of Python-based web scraping tools for collecting, cleaning, and analyzing data from diverse online sources. Whether exploring human rights issues, misinformation, political movements, or social justice topics, this course emphasizes ethical frameworks, privacy compliance, and data integrity in research.

Leveraging trending technologies such as BeautifulSoup, Selenium, and Scrapy, participants will gain hands-on experience in automated data collection, HTML parsing, API integration, and data anonymization techniques. This course is ideal for journalists, researchers, data scientists, and social activists aiming to responsibly investigate and document real-world issues from digital footprints. Equip yourself with the skills to turn chaotic online information into meaningful insights.

Programme Curriculum

Web Scraping and Data Extraction with Python Training Course

Introduction

In today’s data-driven world, extracting and analyzing online information is crucial, especially when dealing with sensitive, complex, or controversial topics. Web Scraping and Data Extraction with Python Training Course empowers participants to ethically harness the power of Python-based web scraping tools for collecting, cleaning, and analyzing data from diverse online sources. Whether exploring human rights issues, misinformation, political movements, or social justice topics, this course emphasizes ethical frameworks, privacy compliance, and data integrity in research.

Leveraging trending technologies such as BeautifulSoup, Selenium, and Scrapy, participants will gain hands-on experience in automated data collection, HTML parsing, API integration, and data anonymization techniques. This course is ideal for

journalists, researchers, data scientists, and social activists aiming to responsibly investigate and document real-world issues from digital footprints. Equip yourself with the skills to turn chaotic online information into meaningful insights.

Course Objectives

1. Understand the ethical considerations of scraping sensitive data
2. Master Python tools like BeautifulSoup, Selenium, and Scrapy
3. Learn to identify and access data sources relevant to sensitive topics
4. Develop dynamic scraping scripts for real-time data collection
5. Clean, structure, and store web data for analysis
6. Scrape social media and dark web data responsibly
7. Apply Natural Language Processing (NLP) on extracted text
8. Use APIs and automation for large-scale data extraction
9. Handle CAPTCHA, JavaScript, and anti-scraping mechanisms
10. Practice data anonymization and privacy preservation techniques
11. Visualize extracted data for storytelling and policy analysis
12. Integrate scraping workflows into research pipelines
13. Conduct risk assessments for working with vulnerable populations

Target Audience

1. Investigative Journalists
2. Academic Researchers
3. Human Rights Activists
4. Data Analysts
5. OSINT Professionals
6. Social Scientists
7. Policy Researchers
8. Digital Forensics Experts

Course Duration: 5 days

Course Modules

Module 1: Introduction to Sensitive Data and Ethical Scraping

- Defining sensitive data in digital research
- Legal and ethical frameworks (GDPR, consent, privacy)
- Risk factors in scraping sensitive topics
- Understanding digital traces and data sources
- Tools for ethical decision-making
- **Case Study:** Scraping COVID-19 misinformation responsibly

Module 2: Web Scraping Fundamentals with Python

- Setting up Python, pip, and virtual environments
- HTML structure and DOM navigation
- Intro to BeautifulSoup and Requests
- Data extraction from static web pages
- Writing modular and reusable scripts
- **Case Study:** Extracting policy updates from NGO websites

Module 3: Advanced Scraping with Selenium and Scrapy

- Handling dynamic content with Selenium

- Building spiders using Scrapy
- Navigating pagination, forms, and AJAX
- Bypassing basic anti-bot protections
- Rate-limiting and concurrency control
- **Case Study:** Tracking gender-based violence reports

Module 4: Social Media and Forum Scraping

- Scraping Twitter, Reddit, and Facebook (legally)
- Sentiment and trend analysis of user-generated content
- API-based scraping (e.g., Twitter API, Pushshift)
- Data structuring and metadata extraction
- Anonymizing user data and ensuring consent
- **Case Study:** Monitoring online hate speech trends

Module 5: Data Cleaning, Storage & Management

- Removing noise and duplicates from scraped data
- Structured storage: JSON, CSV, and databases
- Handling missing or inconsistent data
- Working with SQLite and MongoDB
- Versioning and data integrity
- **Case Study:** Building a database of environmental protest reports

Module 6: Natural Language Processing on Extracted Data

- Text tokenization and entity recognition
- Sentiment analysis and topic modeling
- Keyword extraction and classification
- Handling multilingual data
- Visualizing text data with word clouds and charts
- **Case Study:** Analyzing political discourse on refugee rights

Module 7: Security, Privacy & Anonymization

- Understanding data sensitivity and exposure risks
- Techniques for anonymizing personal identifiers
- Safe storage and encrypted transmission
- Ethical reporting of sensitive findings
- Mitigating harm to vulnerable communities
- **Case Study:** Scraping and anonymizing whistleblower testimonies

Module 8: End-to-End Research Pipeline & Project Showcase

- Planning research objectives and timelines
- Integrating scraping into full research workflow
- Automating recurring data collection tasks
- Reporting and visualization tools (e.g., Jupyter, Power BI)
- Final project design and feedback
- **Case Study:** Multi-country analysis of online protests using scraping and NLP

Training Methodology

- Hands-on coding sessions with live examples
- Guided walkthroughs for tool installation and usage
- Ethical scenario analysis and debates

- Real-world case study breakdowns
- Peer reviews and collaborative debugging
- Final capstone project presentation

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	27 Nov 2026	Online	$1,000
23 Nov 2026	27 Nov 2026	Physical	$1,500
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$6,000
23 Nov 2026	27 Nov 2026	Physical	$4,500
23 Nov 2026	27 Nov 2026	Physical	$6,000

Start Date	End Date	Location	Price
23 Nov 2026	27 Nov 2026	Physical	$6,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com