

Web Scraping with Python

اهداف دوره

- یادگیری استخراج داده از وبسایت‌ها با ابزارهای مختلف پایتون
- تسلط بر BeautifulSoup ، Requests و Selenium
- مدیریت داده‌های ساختاریافته (JSON, CSV, DB)
- کار با وبسایت‌های پویا (JavaScript-heavy)
- آماده‌سازی برای پروژه‌های واقعی در تحلیل داده و هوش تجاری

فصل ۱: مقدمات Web Scraping

- معرفی Web Scraping و کاربردهای آن
- مفاهیم قانونی و اخلاقی در وب‌اسکرپینگ (Robots.txt, Rate Limiting)
- ساختار صفحات وب (HTML, CSS, DOM)
- مروری بر ابزارها و کتابخانه‌ها (requests, BeautifulSoup, Selenium, Scrapy)
- محیط توسعه و آماده‌سازی ابزارها (Anaconda/PyCharm, Virtualenv)

فصل ۲: کار با Requests و BeautifulSoup

- ارسال درخواست HTTP با requests (GET, POST, Headers, Cookies)
- واکشی محتوای HTML از وبسایت‌ها
- معرفی BeautifulSoup و نحوه پارس داده‌ها
 - انتخاب تگ‌ها، کلاس‌ها، IDها
 - جستجو با select() و find_all()

○ استخراج متن، لینک، تصاویر

- ذخیره داده‌ها در CSV و JSON
- پروژه عملی ۱: جمع‌آوری عناوین اخبار از یک وبسایت خبری

فصل ۳: Web Scraping پیشرفته با BeautifulSoup و Regex

- کار با Attribute ها و Data-Attribute ها
- استفاده از Regex برای جستجوی الگوهای متنی
- صفحه‌بندی (Pagination) و اسکرپینگ چندصفحه‌ای
- مدیریت خطاها و Timeout در requests
- پروژه عملی ۲: جمع‌آوری اطلاعات محصولات (نام، قیمت، لینک) از یک فروشگاه اینترنتی

فصل ۴: وبسایت‌های پویا و Selenium

- معرفی Selenium و WebDriver (ChromeDriver/GeckoDriver)
- کار با المان‌ها: find_element, send_keys, click
- مدیریت صفحات جاوااسکریپتی (Lazy Loading, Infinite Scroll)
- شبیه‌سازی کاربر (Login, Form Filling)
- گرفتن Screenshot و ذخیره HTML
- پروژه عملی ۳: ورود به یک سایت، استخراج داده‌های جدول و ذخیره در Excel

فصل ۵: مدیریت داده‌های Scraped

- ذخیره داده‌ها در فایل‌های CSV و Excel (pandas)
- ذخیره‌سازی داده در پایگاه داده SQLite/MySQL
- مدیریت داده‌های حجیم

- پاکسازی داده (Data Cleaning)
 - **پروژه عملی ۴:** جمع‌آوری لیست مشاغل از یک Job Board و ذخیره در پایگاه‌داده
-

فصل ۶: Scraping در مقیاس بزرگ

- کار با Scrapy Framework : معماری و اجزای اصلی
 - ساخت اسپایدر ساده
 - مدیریت Pipeline ها برای ذخیره داده
 - Rate Limiting و جلوگیری از بلاک شدن
 - استفاده از Proxy و User-Agent Rotation
 - **پروژه عملی ۵:** ساخت اسپایدر برای جمع‌آوری چند هزار داده از یک وب‌سایت
-

فصل ۷: پروژه پایانی

- اجرای یک پروژه واقعی انتخابی مثلا:
 - جمع‌آوری اطلاعات سهام از یک وب‌سایت مالی
 - استخراج داده‌های گردشگری (هتل‌ها، پروازها)
 - ساخت پایگاه‌داده قیمت کالاها از فروشگاه‌های آنلاین
- ارائه و تحلیل داده‌ها در قالب گزارش یا داشبورد ساده (pandas + matplotlib)