Derrick Karanja
Software Developer (Python)
Nairobi, Kenya | 0795150677
derrick01.dk@gmail.com | www.linkedin.com/in/derks01/

SUMMARY

- Detail-oriented Python Web Scraper Specialist with extensive experience in developing, optimizing, and maintaining automated web scraping scripts.
- Skilled in extracting, processing, and managing large datasets from e-commerce, finance, and research platforms.
- Adept at leveraging Scrapy, BeautifulSoup, Selenium, Requests, and Pandas to ensure efficient and scalable data pipelines.
- Strong background in data cleaning, storage, and API integration, with a focus on performance optimization, legal compliance, and automation.

SKILLS

Proficiency in Python

- Strong understanding of Python programming for web scraping and data extraction.
- Experience with Scrapy, BeautifulSoup, Selenium, Requests, and lxml for structured/unstructured data extraction.

Web Technologies

- Knowledge of HTML, CSS, and JavaScript to analyze web structures and implement scraping strategies.
- Familiarity with JSON, XML, and REST APIs for data extraction and integration.
- Proficiency in browser automation tools (Selenium, Playwright) for handling dynamic content.

Data Handling & Storage

- Skilled in Pandas for data manipulation, cleaning, and transformation.
- Experience with SQL (MySQL, PostgreSQL) and NoSQL (MongoDB) databases for storing scraped data.
- Knowledge of data formats (CSV, JSON, Parquet) for structured data storage.

Problem-Solving & Optimization

- Strong analytical and debugging skills to troubleshoot scraping issues.
- Ability to optimize scrapers for speed, efficiency, and scalability.
- Experience implementing proxy rotation, CAPTCHA solving, and headless browsing to avoid bans.

Web Scraping Ethics & Compliance

- Understanding of legal and ethical considerations surrounding web scraping.
- Ability to design scrapers that respect robots.txt, user-agent rotation, and site terms of service.

Automation & Scheduling

- Proficiency in automating scraping tasks with Cron Jobs, Celery, and Airflow.
- Experience deploying scrapers using cloud services (AWS Lambda, GCP Cloud Functions, Azure).
- Knowledge of CI/CD pipelines (Docker, GitHub Actions, Jenkins) for automated deployments.

EDUCATION

BSc. Mathematics and Computer Science
Meru University of Science and Technology, Kenya (September 2021 – May 2025)

WORK EXPERIENCE

Python Web Scraper Developer
Freelance | Remote | May 2023 – Present
- Developed and optimized Python web scrapers for clients in e-commerce, research, and financial analytics.
- Improved scraping efficiency by 60% by implementing asynchronous processing with Scrapy and multiprocessing techniques.
- Automated data extraction workflows, reducing manual effort by 70% through scheduled scraping tasks.
- Extracted, cleaned, and stored over 5 million records in PostgreSQL, MongoDB, ensuring 99% data accuracy.
- Built a web dashboard (Flask + React) to visualize real-time scraped data for clients.
- Implemented proxy rotation and headless browsing, reducing the IP ban rate by 80%.

Data Annotation Specialist
CloudFactory | Nairobi, Kenya | July 2022 - February 2025
- Annotated over 20,000 structured and unstructured datasets to improve AI model performance.
- Increased NLP model accuracy by 15% through optimized text and sentiment analysis annotation.
- Developed Python scripts for data preprocessing, reducing annotation errors by 30%.
- Conducted data quality checks, ensuring 99.5% labeling accuracy in training datasets.

Prompt Engineer & Web Data Extraction
SkillCat | Remote | January 2023 - March 2024
- Created structured LLM prompts for AI-driven content generation, increasing text quality by 20%.
- Automated data extraction from public repositories using Scrapy & Selenium, reducing manual research efforts by 50%.
- Designed an AI-based chatbot, leveraging Python, OpenAI API, and scraped knowledge databases.

PROJECTS

1. Automated E-Commerce Price Monitoring Tool
Tech Stack: Python, Scrapy, Selenium, PostgreSQL, AWS Lambda
- Reduced manual price tracking time by 90% using automated scraping pipelines
- Increased scraping efficiency by 40% by switching to asynchronous requests in Scrapy.
- Implemented proxy rotation, reducing IP bans by 80%.
- Stored & processed over 3 million price records, enabling better market analysis for clients.

2. Job Posting Aggregator for Remote Work
Tech Stack: Python, Scrapy, Flask, MongoDB, Docker
- Built a scraper that collected 10,000+ job postings monthly from multiple sources.
- Designed a REST API to serve the scraped data dynamically via Flask.
- Automated scraper scheduling, reducing maintenance workload by 60%.
- Deployed using Docker & AWS Lambda for scalability and reliability.

3. Portfolio Website for a Freelance Writer
Tech Stack: React, Next.js, Tailwind CSS, Firebase (backend), Vercel (deployment)
- Designed and developed a fully responsive personal portfolio showcasing writing samples.
- Implemented dynamic content management so the client could update blogs/articles easily.
- Optimized site speed by 40%, ensuring fast loading and improved SEO ranking.

4.Academic Past Paper Scraper
Tech Stack: Python, Requests
- Scraped 100,000+ academic past papers from educational websites.
- Implemented PDF text extraction to convert scanned documents into searchable formats.

- Created a Flask-based API to allow users to query and download papers.
- Reduced manual search time by 95%, improving accessibility for students.

CERTIFICATIONS
- Python for Data Science – Cisco Networking Academy
- Web Scraping with Python & Scrapy – Udemy
- Cloud Computing Fundamentals (AWS, GCP) – ALX
- AI & Data Science Essentials – Microsoft Training
- Ethical hacking - Cisco Networking Academy
- Data Entry and Management - AJira Digital Kenya

TOOLS & TECHNOLOGIES
- Python | Scrapy | Selenium | BeautifulSoup | Requests | lxml
- Pandas | NumPy | MySQL | PostgreSQL | MongoDB | SQLite
- AWS Lambda | Google Cloud Functions | Docker | Jenkins
- APIs & JSON Parsing | Git & GitHub | Linux & Bash Scripting

PROFESSIONAL DEVELOPMENT
- Member of Data Science & AI Communities
- Contributor to Open-Source Scrapy & Web Scraping Projects
- Active Blogger on Medium (Web Scraping & Data Automation)

VOLUNTEER EXPERIENCE
Data Annotation Mentor
Youth for Tech | Nairobi, Kenya | April 2023 – July 2023
- Conducted hands-on workshops on web scraping & AI data annotation.
- Mentored students on Scrapy, Selenium, and API-based data collection.

INTERESTS
- AI & Data Science – Exploring AI-driven data extraction techniques.
- Cybersecurity – Investigating ethical scraping & CAPTCHA bypass solutions.
- Technology Blogging – Writing educational content on web scraping & automation.

LANGUAGES
- English – Fluent
- Swahili – Native