

Vincent Doh

Web Crawling Engineer

North Ridge, Osu, Greater Accra, Ghana | Phone: (+233) 20-884-1157 | Email: vincent.doh@hotmail.com LinkedIn: <https://www.linkedin.com/in/vincent-doh/> | GitHub: <https://github.com/vin-doh/> Portfolio: <https://vincentdoh.netlify.app>

PROFESSIONAL SUMMARY

Detail-oriented and ethical Web Crawling Engineer with over 4 years of experience building scalable and efficient crawlers using Python, Scrapy, Puppeteer, and Selenium. Proven ability to extract, clean, and manage large-scale datasets while complying with data privacy regulations like GDPR. Adept at using Docker for deployment, managing proxies and CAPTCHAs, and maintaining robust web scraping systems. Strong focus on automation, optimization, and cost-effective solutions in fast-paced startup environments.

EDUCATION

Harvard University
Data Science: Machine Learning

Harvard Online, USA. 2021-2023

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript, HTML, CSS

Tools: Scrapy, Puppeteer, Selenium, BeautifulSoup, MongoDB, PostgreSQL, MySQL, SQLite, Slack, Microsoft Office Suite Expert- Excel, Power BI, Microsoft Azure, Pandas, Scikit-learn, Git, Docker, Linux, Databases

SKILLS

- Communication skills, excellent interpersonal, oral, written and proven leadership skills.
- Strong collaboration skills, prioritizing in a fast-paced work environment, delivering to deadlines.
- Deep web scraping tools and libraries experience with Scrapy, Puppeteer, and Selenium.
- Experienced in Docker, data cleaning, proxy management, and ethical data extraction.

PROFESSIONAL EXPERIENCE

Freelance Web Crawling Engineer

Feb 2020 - Present

Agently AI

Remote, EMEA.

- Built and maintained 20+ scalable web crawlers using Scrapy and Puppeteer for clients across recruitment, e-commerce, and fintech sectors.
- Solved CAPTCHA challenges using anti-captcha APIs and manual fallback systems.
- Automated IP rotation using proxy pools and rate limiting to avoid blacklisting.
- Extracted and cleaned structured/unstructured data using pandas and regex.
- Used Docker to containerize scraping environments for deployment consistency.
- Stored and queried scraped data in MongoDB and PostgreSQL databases.
- Ensured compliance with GDPR and robots.txt to maintain ethical standards.

Business Intelligence Analyst

Jun 2022 - Dec 2024

Ingeo

Hybrid - Accra.

- Developed and optimized data models for BI and analytics, which led to a 20% improvement in reporting accuracy and timeliness for key stakeholders.
- Created and maintained playbooks and style guides to ensure scalable consistency across hundreds of assets.
- Contributed to the development of cloud-based data solutions on AWS and Microsoft Azure, enabling 30% faster access to integrated data across platforms.

PROJECTS

Web Scraping for Competitive Market Research

Developed a web scraping tool to extract competitor data on pricing, location trends, and product conditions, providing actionable insights for strategic decisions.

<https://github.com/vin-doh/jijiresearchdata>

Tools: Selenium, pandas, and openpyxl, with ChromeDriver for browser automation.