

Test Overview and Objective for Web Automation Engineer

Extract structured data from a given website, handle potential challenges such as pagination or dynamic content, and present findings clearly. This test evaluates practical skills in **data extraction, anti-bot mitigation, and handling complex web automation challenges** that reflect real-world scenarios. Use appropriate tools and communicate your approach and results effectively.

Estimated Time: 2-4 hours

Task Description

1. **Core Web Automation Implementation**
 - a. Basic data extraction - build a scraper to extract product information from a website, handling at least 50 products while implementing proper error handling.
 - b. Pagination Handling - the scraper must navigate through multiple pages automatically, handling different pagination styles including next buttons and page numbers.
2. **Anti-bot Migration**
 - a. Rate Limiting bypass - implement intelligent delays between requests, exponential backoff for retries, user agent rotation, and proper handling of 429 Too Many Requests responses.
 - b. Advanced Anti-bot handling - Implement proxy rotation, handling JavaScript-rendered content, Captcha detection, and session management.
3. **Dynamic Content Handling**
 - a. JavaScript-Heavy Sites - Use headless browser automation tools like Selenium or Playwright to scrape JavaScript-heavy content, properly waiting for dynamic content to load and handling AJAX requests.

Deliverables

- Code/Scripts used for scraping and data extraction, with clear instructions on how to run them
- README with set-up and usage instructions
- Extracted data file in JSON or CSV format
- Written report (1-2 pages) including your overall approach and design