



web-scraping.dev

web-scraping.dev is a mock website for testing and learning about web scraping. It covers popular web patterns encountered in web scraping so take a look at the [scenarios](#) section for details.

This platform is used in:

- [ScrapFly Docs](#)
- [ScrapFly Web Scraping Academy](#)
- [ScrapFly Blog Highlights and Tutorials](#)

Refer to 🙌 for learning web scraping.

Scenarios

web-scraping.dev implements many web patterns encountered in modern web scraping:

Static Paging

/products

HTML-based server-side item paging where each page is it's own URL.

Endless Scroll Paging

/testimonials

Dynamic client side paging where new items appear as user scrolls.

Secret API token

/testimonials

The testimonial paging shows how X-Secret-Token is used to lock access to hidden APIs.

Endless Button Paging

/reviews

Dynamic client side paging where new items appear as presses Load More button.

GraphQL Background Requests

/reviews

Forced New Tab Links

/reviews

The review policy and code of conduct links use different

Data loaded using javascript through a backend graphql API.

techniques to force opening links in a new tab.

Product HTML Markup

/product/1

Basic e-commerce product structure and css class based markup.

Hidden Web Data

/product/1

Product review data is hidden in HTML as json, then loaded to HTML on page load.

Local Storage

/product/1

The cart system is powered by local storage to demo client-side website state store (alternative to cookies).

CSRF Token Locks

/product/1

The Load More reviews action uses a X-CSRF-Token header to block cross-site access.

Cookies Based Login

/login

User authentication based on form request and cookies.

PDF Downloads

/login

The login page features link and js based file download triggers.

Cookie Popup

/login?cookies

Cookie info modal pop up that blocks the entire screen

Example Block Page

/blocked

Valid 200-status response that redirects to block notification.

Blocking Redirect for Invalid Referer

/credentials

The credentials page is only accessed with valid Referer otherwise redirects to blocked. So this page can only be accessed by clicking link on /login.

Persistent Cookie-Based blocking

/blocked

Using cookies to mark blocked clients for persistent blocking.

Changelog

▶ **v1.3.1**

▶ **v1.3.0**

▶ **v1.2.0**

▶ **v1.1.0**

SCRAPFLY ACADEMY

v1.3.0 made by [ScrapFly](#). 