

Planning for success with web data

The checklist for successful web scraping

As you delve into the complexity of a web scraping project, it can be overwhelming to know where to start. This checklist accompanying our [Web Data 101: Planning for success webinar](#) is a step-by-step guide to help you answer the questions that matter.

How to use this checklist

This checklist outlines the key steps in designing a successful web scraping project.

- Step 1: Define your business requirements
- Step 2: Data discovery
- Step 3: Selecting the right data attributes
- Step 4: Write the goal statement of your web scraping project

Follow this checklist to equip yourself with the necessary tools to effectively plan and execute a successful web scraping project.

Take a deep breath, grab a cup of coffee, and let's get started!

Step 1: Define your business requirements



Identify the problem

Clearly define the business problem you are trying to solve



Determine the role of web data

Evaluate how web data can play a role in achieving your business goals



Define the data utilization

Who will use it

How will they use it

How will you get the data to its destination

Step 2: Data discovery



Identify data sources

Research and identify target websites or sources that contain the data you need to collect



Data filtering

Determine if you need to input any data to filter the data on the website before extraction



Geolocation access

Consider if accessing the website from a specific location is required to view the correct data

Step 3: Selecting the right data attributes

Identify desired data fields to define the minimal viable schema

Determine which data fields you want to extract, such as product name, price, and description, etc to design the minimal viable schema (MVP) of the data that you need to collect to achieve your business goals

Download files

Consider if you want to download any files, such as PDFs or CSVs

*Be mindful of the copyright implications

Screenshot

Determine if you want to take a screenshot of the page

Data Formatting

Evaluate if you need the data to be formatted in a specific way, such as removing currency signs from product prices

Step 4: Write the goal statement of your web scraping project

The goal of the project is to build that will help us achieve for which we need to extract data from of in time and refresh it in days.

Example: The goal of the project is to build a Product Pricing Comparison that will help us achieve informed business decision-making such as setting prices, identifying gaps in product offerings, tracking market changes, and identifying the most popular products and trends in the market for which we need to extract product name, product price, product details data from competitor's websites in real-time and refresh it in regular intervals of time (e.g. daily, weekly).

That's it! Answering these questions will help you resolve a lot of your queries and easily set the foundation of your project on solid ground!

At Zyte, we understand that web scraping projects can be complex and challenging. That's why we're here to help. Don't shy away from reaching out to us for assistance. Our team of experienced developers is here to support you and help you succeed in your web scraping projects. If you have any questions or concerns, feel free to [reach out to us](#).

We're here to help you win.

[Talk to us](#)
zyte.com/data-extraction