# Top 18 Web Scraping Applications & Use Cases in 2022

**Data Analytics & Data science**

**Machine learning training data collection**

Machine learning algorithms require a large volume of data to improve the accuracy of outputs. However, collecting a large amount of accurate training data is a big pain. Web scraping can help data scientists acquire the required training dataset to train ML models. For example, GPT-3 which impressed the computer science community with its realistic text generation was built on textual content on the web.

To learn more about web crawler use cases in data science, feel free to read our in-depth guide to web scraping for machine learning.

**Marketing & sales**

**Price intelligence data collection**

For every price elastic product in the market, setting optimal prices is one of the most effective ways to improve revenues. However, competitor pricing needs to be known to determine the most optimal prices. Companies can also use these insights in setting dynamic prices.

**Sponsored:**

Bright Data's Data Collector is a web scraper that can be used to extract competitors' pricing data and this is the most common web scraping use case mentioned by most companies in the space.

A web crawler can be programmed to make requests on various competitor websites' product pages and then gather the price, shipping information, and availability data from the competitor website.

Another price intelligence use case is ensuring Minimum Advertised Price (MAP) compliance. Manufacturers can scrape retailers' digital properties to ensure that retailers follow their pricing guidelines.

**Fetching product data**

Specifically, in e-commerce, businesses need to prepare thousands of product images, features, and descriptions that have already been written by different suppliers for the same product. Web scraping can automate the entire process and provide images and product descriptions faster than humans. Below is an example of extracted product data from an e-commerce company website.

Source: Scrapinghub

**Brand protection**

Using web scraping brands can swiftly identify online content (e.g. counterfeit products) which can hurt your brand. Once this content are identified, brands can take legal action against those responsible:

- **Counterfeiting**: Counterfeiters need to market their products and scrapers allow businesses to identify those products before actual users and protect users from buying fake products.
- **Copyright infringement** is the use of copyrighted works without permission. Web scrapers can help identify whether copyrighted intellectual property is used illegally.
- **Patent theft** is the unlawful manufacturing or selling of licensed products.
- **Trademark infringement** is the illegal use of a logotype, pattern, phrases, or any other elements that are associated with the brand.

**Competition research**

**Lead generation**

Lead generation efforts can help businesses reach additional customers. In this process, the marketer starts communicating with relevant leads by sending out messages. Web scraping helps reach out to leads by scraping contact details such as email, phone, and social media accounts.

**Lead prioritization**

In Account-Based Marketing (ABM) crawlers are used to scrape firmographic and technographic data. These data can be used to prioritize leads based on their likelihood to buy.

In addition, signals (e.g. promotions, new hires, new investments, M&A) that are

likely to trigger purchasing can be scraped from news or company announcements. This can help companies further prioritize their marketing efforts.

**Marketing communication verification**

Companies invest billions in spreading their message and especially large brands need to be careful about how their marketing messages are delivered. For example, Youtube got in trouble in 2017 by displaying Fortune 500 links in hateful and offensive videos.

**Monitoring consumer sentiment**

Analyzing consumer feedback and reviews can help businesses understand what is missing in their products & services and identify how competitors differentiate themselves. However, there are dozens of software review aggregator websites that contain hundreds of reviews in every solution category. Web scraping tools and open-source frameworks can be used to extract all these reviews and generate insights to improve services and products.

For example, AIMultiple solution pages include a summary of insights from all online sources, helping businesses identify different products' strengths and weaknesses.

**SEO Audit & Keyword research**

Search engines like Google consider numerous factors while ranking websites. However, search engines provide limited visibility into how they rank websites. This led to an industry of companies that offer insights on how companies can improve their online presence and rank higher on search engines.

Most SEO tools such as Moz and Ubersuggest crawl websites on-demand to analyze a website's domain. SEO tools utilize web crawlers for SEO monitoring to

- run SEO audits: Scrape their customers' websites to identify technical SEO issues (e.g. slow load times, broken links) and recommend improvements
- analyze inbound and outbound links, identifying new backlinks

- scrape search engines to identify different companies' web traffic and their competition in search engines. This scraping can also help generate new content ideas and content optimization opportunities supporting companies' keyword research efforts.

- scrape competitors to identify their successful strategies taking into account factors like the word count of the different pages etc.

- scrape the rank of your website weekly/ annually in keywords you are competing. This enables SEO team to take immediate action if any unpredicted rank decrease happens.

**Website testing**

Webmasters may use web scraping tools to test the website's front-end performance and functionality after maintenance. This enables them to make sure all parts of the web interface are functioning as expected. A series of tests can help identify new bugs. For example, tests can be run every time the tech team adds a new website feature or changes an element's position.

**Public Relations**

**Brand monitoring**

Brand monitoring includes crawling various channels to identify who mentioned your company so that you can respond and act on these mentions to serve them better. This can involve news, complaints & praises on social media.

**Trading**

**Data-driven portfolio management**

Hedge funds rely on data to develop better investment strategies for their clients. According to Greenwich Associates, an average hedge fund spends roughly $900,000 per year on alternative data source. Web scraping is listed as the largest source of alternative data:

Source: Institutional InvestorOne web scraping example is extracting and aggregating news articles for predictive analysis. This data can be used to feed into their own machine learning algorithms to make data-driven decisions.

**Strategy**

**Building a product**

The goal of Minimum Viable Products (MVPs) is to avoid lengthy and unnecessary work to develop a product with just enough features to be usable by early customers. However, MVPs may require a large scale of data to be useful to their users, and web scraping is the best way to acquire data quickly.

**Market research**

No research can be done without data. Whether it is academic research of a professor or commercial research on a specific market, web scraping can help researchers enhance their articles with insights uncovered by scraped data. This leads to better decisions such as entering a new market or a new partnership.

**Support functions**

**Procurement**

Health of a company's suppliers' is important to a company's success. Companies rely on software or services providers like Dunn & Bradstreet to understand supplier health. These companies use various approaches to collect company data and web data is another valuable data source for them.

**HR: Fetching candidate data**

There are various job portals such as Indeed and Times Jobs where candidates share their business experience or CVs. A web scraping tool could be utilized to scrape potential candidates' data so that HR professionals can screen resumes and contact candidates that fit the job description well. However, as usual, companies need to ensure that they do not violate T&Cs of job portals and only use public information on candidates, not their non-public personal information (NPPI).

AI has significant use cases in HR, for example by automating CV screening tasks and frees up a significant amount of the HR team's time. For example, candidates' career progression after joining a new company can be correlated with their educational background and previous experience to train AI models on identifying the right candidates. For example, if those with engineering

backgrounds and with a few years of marketing experience in a marketing agency, end up getting promoted fast in a marketing role in a certain industry, that could be a valuable information for predicting the success of similar candidates in similar roles. However, this approach has significant limitations, for example Amazon's recruiting tool was identified to be biased since it relied on such historical data.

Explore web crawling use cases in recruiting in our in-depth article.

## Technology

### Website transition

For companies that operate on a legacy website and transfer their data to a new platform, it is important to ensure that all their relevant data is transfered to the new website. Companies operating legacy websites may not have access to all their website data in an easy to transfer format. Web scraping can extract all relevant information in legacy websites.