

## IMPORTANCE OF WEB SCRAPING IN E- COMMERCE BUSINESS

---

**Sandeep Shreekumar**

Student, Sterling Institute of  
Management Studies, Nerul,  
Navi Mumbai

[sandeepshreekumar2@gmail.com](mailto:sandeepshreekumar2@gmail.com)

**Satyan Mundke**

Student, Sterling Institute of  
Management Studies , Nerul,  
Navi Mumbai

[satyanrm88@gmail.com](mailto:satyanrm88@gmail.com)

**Dr. Murlidhar Dhanawade**

Professor, Sterling Institute  
of Management Studies,  
Nerul, Navi Mumbai

[dr.murlidhar.dhanawade@gmail.com](mailto:dr.murlidhar.dhanawade@gmail.com)

---

### Abstract

Web scraping has become an indispensable tool for E-commerce businesses in recent years. With the proliferation of e-commerce businesses, the amount of data available on the internet has exploded, and Data extraction is a method of collecting data from various sources for dynamic websites, it's very complicated to extract data due to its dynamic nature. web scraping is an essential tool for e-commerce businesses to gather competitive intelligence, optimize pricing strategies, gather product information, analyze customer behaviour, and conduct market research. By scraping data from competitor websites, businesses can gain valuable insights into pricing strategies, product descriptions, and customer reviews. Web scraping is an automatic method to obtain big amounts of records from websites. Most of these records are unstructured statistics in an HTML layout which are then transformed into established statistics in a spreadsheet or a database. In this way, web scraping can help E-commerce businesses stay competitive and make data-driven decisions that ultimately drive growth and success.

**Keyword:** Web Scraping, E-Marketing, E- Commerce , Python, Beautiful Soup

### 1. INTRODUCTION

Web scraping has emerged as an important strategy for e-commerce businesses, especially in delivering rich data-based insights. For instance, e-commerce web scrapers help in the identification of consumer preferences and choices. They help in assessing trends of purchase behavior in online circles. The process of web scraping involves the use of automated crawlers across visible content of different e-commerce websites. These crawlers then extract desired data and compile the information into systematic reports. Automated web scraping service for e-commerce website portals generates information effortlessly.

Web scraping can be leveraged to generate all kinds of data right from pricing information to sentiments analysis from social media. For online retailers, web scraping is an exceptionally powerful business tool. Here are some ways to leverage web scraping to benefit online stores. E-commerce delivers profits to everyone, from independent freelancers to large corporations. According to Forbes, Global E-Commerce Market To Expand By \$1 Trillion By 2025.

## **2. LITERATURE REVIEW**

Web scraping refers to the process of collecting data from websites using automated tools, also known as web crawlers or spiders. In e-commerce, web scraping plays a crucial role in gathering competitive intelligence, product research, and price monitoring, which enables businesses to make data-driven decisions. This literature review provides an overview of the importance of web scraping in e-commerce business, including its benefits, challenges, and ethical concerns.

**1. Competitive Intelligence** - Web scraping is an important tool for e-commerce businesses to gather competitive intelligence. By analyzing the content and data on competitors' websites, e-commerce businesses can gain insights into their pricing strategies, product offerings, marketing campaigns, and customer engagement tactics. This information can be used to make informed decisions about product pricing, promotions, and marketing strategies, which can give them a competitive edge.

**2. Pricing Analysis** - Web scraping can help e-commerce businesses track changes in prices for their products and services in real-time. By monitoring competitor prices and analyzing customer behavior, e-commerce businesses can optimize their pricing strategies to increase sales and revenue. Web scraping also enables e-commerce businesses to track discounts, coupons, and promotional offers, which can help them stay competitive in the market.

**3. Product Information** - Web scraping is an effective way for e-commerce businesses to gather data on product information, such as product descriptions, specifications, and reviews. By monitoring product information on competitors' websites, e-commerce businesses can gain insights into market trends and consumer preferences, which can help them tailor their product offerings to meet customer needs.

**4. Customer Behavior** - Web scraping can help e-commerce businesses analyze customer behavior, such as browsing patterns, purchase history, and product preferences. This

information can be used to personalize product recommendations, marketing campaigns, and promotions, which can lead to increased customer loyalty and sales.

**5. Market Research** - Web scraping can provide e-commerce businesses with valuable insights into market trends and consumer preferences. By monitoring industry publications, blogs, and social media platforms, e-commerce businesses can gather data on customer sentiment, new product launches, and emerging trends. This information can be used to develop new product lines, enter new markets, and improve marketing strategies.

Challenges of Web Scraping in E-commerce Business:

- **Legal Issues:** Web scraping is not always legal and can potentially result in legal action if done without permission or in violation of terms of service agreements.
- **Data Quality:** The quality of data collected through web scraping can vary, and businesses need to ensure that they are collecting accurate and relevant data for their needs.
- **Technical Challenges:** Web scraping requires technical expertise and resources, and businesses need to invest in the right tools and personnel to ensure that they can scrape data effectively.

### **Ethical Concerns of Web Scraping in E-commerce Business**

**Privacy Concerns:** Web scraping can potentially collect personal data without users' consent, raising privacy concerns.

- **Fair Competition:** Web scraping can also be used to gain an unfair advantage over competitors, which can harm fair competition in the market.

Web scraping is a valuable tool for e-commerce businesses, providing them with access to data that can inform their decisions and drive growth. However, businesses need to be aware of the legal and ethical concerns associated with web scraping and ensure that they are collecting data in a responsible and ethical manner. With the right approach and tools, web scraping can be a powerful tool for driving growth and success in the e-commerce industry.

## **3. PROBLEM DEFINITION**

Web scraping offers insight into price data, market dynamics, prevailing trends, practices employed by the competitors and the challenges face in the business. This is readily available

data if marketers know how to get it. What many marketers don't know is how useful it can be to them. Web scraping has emerged as an important strategy for e-commerce businesses, especially in delivering rich data-based insights.

Practices of web scraping in e-commerce

- Price-monitoring and Product Research
- Online price comparison
- Better Customer analysis
- Market Analysis
- Better advertisements
- Influences Marketing and Sales Strategy
- Brand monitoring
- Extract business details from business directory
- Helps in Future analysis Etc.

#### **4. OBJECTIVE**

When it comes to E-Commerce, Web scraping can be used for a variety of purposes. For example, it can use scraping to gather product data, competitor data, or pricing information. It can also use scraping to find new leads and identify potential suppliers.

How to Effectively Optimize Prices Using Web Scraping Data:

The price-monitoring method is one of the most valuable advantages of web scraping data for e-commerce platforms. Whether you are an eBay or Amazon seller or startup founder, use of this approach can make your life much easier. By analyzing various proposals represented on the market and can quickly settle on the most effective price for the offering.

What are the advantages of web scraping for e-commerce?

**1. Optimized revenue:** Competitor monitoring is difficult when online businesses keep changing their strategies. Adapting products and lowering prices can take customers and, ultimately, revenue away from marketer site. By using a web scraper to extract data, marketer can monitor competitor prices and products. This detailed product analysis will allow Marketer to adjust own stock and prices accordingly. Thus, it helps to keep the revenue flowing into the business rather than market competitors.

**2. Improved efficiency:** Uploading products can be time-consuming when the business relies on it. However, harvested data can automate these manual uploads and other workflows. The automation of these processes saves valuable time and improves the overall efficiency of business.

**3. Track product performance:** Better product analysis is yet another perk of website scraping. By monitoring listings and sales data, it allows marketer to see how well different products are performing.

**4. New content and ideas:** New products and images pop up all over the internet, but they can be hard to get your hands on. Web scraping audits entire websites for both of the above, and much more. The company doesn't need to miss out on new material if marketer use this method of competitor monitoring.

**5. Market analysis and machine learning:** Data extraction is used for more than just monitoring competitor activity. The text, images, and other data gathered through website scraping can be used for market research or machine learning. This allows you to use historical data more effectively for model training and improve the accuracy of your market analysis so that you can make better decisions.

**6. Keep your listings up to date:** Manually keeping your listings up to date can be a long process. Constant, automated updates are far more efficient. Website scraping provides this through data extraction for app-based catalog updates. This automation means that your listings will never be left behind the times.

**7. Show customers what your competitors are doing:** Marketer can also automate your product customization process based on relevant market research. This type of competitor monitoring gives customers an accurate overview of market rivals' inventory and pricing tiers. When customers can monitor competitors, they can see the benefits of buying from the brand.

**8. Constant improvements:** The only constant in business is change, and making improvements to existing model requires real feedback. Fortunately, this data extraction is very straightforward—web scrapers monitor products to collect feedback. Marketer can use this to keep developing and bettering their online business

## 5. RESEARCH METHODOLOGY

How does a web scraper function?

The exact method differs depending on the software or tools that are using, all web scraping bots follow three basic principles:

Step 1: Making an HTTP request to a server

Step 2: Extracting and parsing (or breaking down) the website's code

Step 3: Saving the relevant data locally

First step, robots.txt: One of the most important and overlooked step is to check the robots.txt file to ensure that we have the permission to access the web page without violating any terms or conditions. Secondly, the web scraper will be given one or more URLs to load before scraping. The scraper then loads the entire HTML code for the page in question. More advanced scrapers will render the entire website, including CSS and JavaScript elements. Then the scraper will either extract all the data on the page or specific data selected by the user before the project is run. Ideally, the user will go through the process of selecting the specific data they want from the page. Lastly, the web scraper will output all the data that has been collected into a format that is more useful to the user.

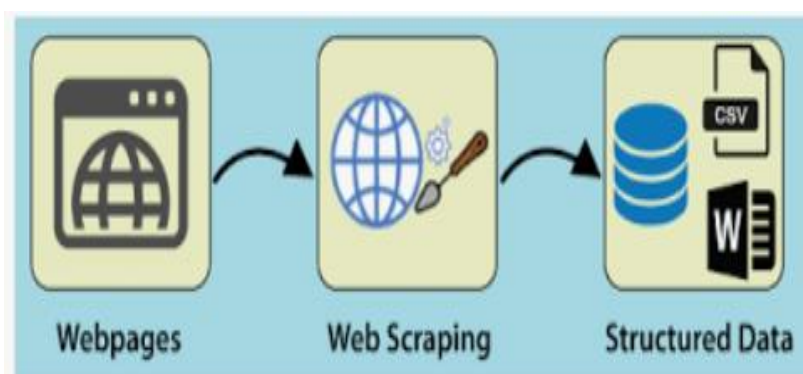


Figure1: Architecture of web scraping (Source: <https://www.javatpoint.com/web-scraping-using-python>)

The web scraper will output all the data that has been collected into a format that is more useful to the user. Most web scrapers will output data to a CSV or Excel spreadsheet, while more advanced scrapers will support other formats such as JSON which can be used for an API.

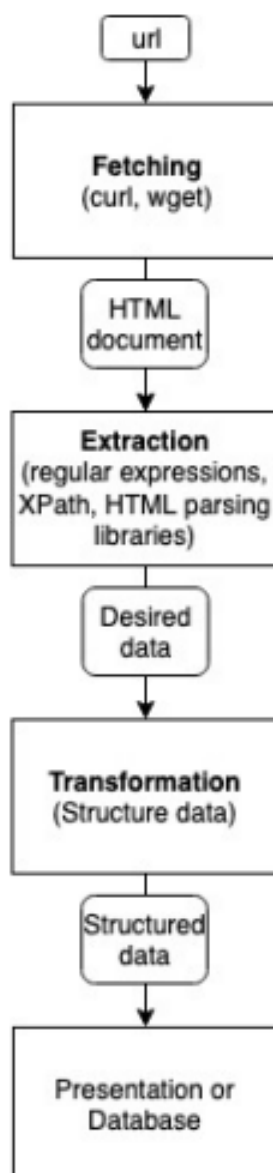


Figure 2: Web Scraping System (Source: [https://www.researchgate.net/figure/Web-Scraping-process-Persson-2019\\_fig2\\_357401723](https://www.researchgate.net/figure/Web-Scraping-process-Persson-2019_fig2_357401723))

## Web Scraping tools

### Scrapy

Scrapy is an open source Python framework originally designed solely for web scraping, but now also supports web crawling and extracting data via APIs.

XPath or CSS selectors is the built in way to do the data extraction, but external libraries such as BeautifulSoup and lxml can be imported and used. It allows for storing data in the cloud.

## **Selenium**

A way of automating and simulating a human browsing with a web browser can be accomplished by using a tool called Selenium. It is primarily used and intended for testing of web applications, but is a relevant choice for webscraping. Using the Selenium WebDriver API in conjunction with a browser driver (such as Chrome Driver for the Google Chrome browser) will act the same way as if a user manually opened up the browser to do the desired actions. Because of this, loading and scraping web pages that makes use of JavaScript to update the DOM is not a problem. The Selenium WebDriver can be used in Java, Python, C#, JavaScript, Haskell, Ruby.

## **Jaunt**

Jaunt is a web scraping library for Java that, compared to Selenium, uses a headless browser. It allows for controlling HTTP headers and accessing the DOM, but does not support JavaScript. Because of lacking the ability to work with JavaScript it is more lightweight, faster and can be easily scaled for larger projects. If JavaScript is a must, a crossover between Jaunt and Selenium called Jauntium can be used.

## **BeautifulSoup**

BeautifulSoup is a Python HTML extracting library. While not being a complete web scraping tool, it can be used in conjunction with the requests package, which allows for doing HTTP calls in Python. BeautifulSoup is a simple, pythonic way to navigate, search and modify parse trees, such as an HTML tree. It is intended to be easy to use and provides traversal functionality such as finding all links or all tables matching some condition.

## **6. CODE FOR WEB SCRAPING**

The most well-known language use in web scraping include:

- R programming (R-selenium)
- Python (BeautifulSoup, Selenium)
- Ruby
- NodeJS
- Java



Choosing Python for webscraping because Python is popular, high-level programming language. Python can run on many different platforms and has a simple syntax similar to the English language, therefore, it is easy to code.

The goal of web scraping is to collect web data that would be in HTML format or Json. Python provides a library called Requests, which is a simple HTML library that allows to integrate Python programs with web services.

To find the relevant data Python provides another library called BeautifulSoup, which helps retrieve particular content from a webpage, remove HTML tags, and save the information. The last step in web scraping is to save the collected data in a structured form. With the help of Python Pandas Library, the data can store in the desired format.

### **Web Scraping Example –**

The advantages of web scraping in ecommerce is market analysis. The section that will interest us will be to obtain the Amazon Best Sellers data.

Amazon is one of the largest online business enterprises which sells millions of varieties of goods all over the world through the platform called Amazon.com. They have different categories of products including Fashion, Books, Electronics, Toys, Jewellery, and more. There are some popular products that are listed in the bestseller category of Amazon that is helpful for sellers to find the bestselling products as well as the customers to find the bestselling quality products to buy. Here we are going to scrape those bestselling products of different categories using python and Beautiful Soup.

Steps -

- Install and import the required libraries
- Parse a specific bestseller category page using requests and Beautiful Soup
- Fetch information about the item including name, price, reviews, rating, and URL
- Store all the information's into a CSV file using pandas

### **Code Implementation –**

```
$ pip install requests lxml bs4 pandas
import requests
from bs4 import BeautifulSoup
import pandas as pd
```

```

HEADERS = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:66.0)
Gecko/20100101 Firefox/66.0", "Accept-Encoding": "gzip, deflate",
"Accept": "text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8",
"DNT": "1", "Connection": "close", "Upgrade-Insecure-Requests": "1"
def scrape_product_details(url):
    resp = requests.get(url, headers=HEADERS)
    content = BeautifulSoup(resp.content, 'lxml')
    item_details = []
for item in content.select('.zg-item-immersion'):
    try:
        data = {
            "item": item.select('.p13n-sc-truncate')[0].get_text().strip(),
            "price": item.select('.p13n-sc-price')[0].get_text().strip(),
            "rating": item.select('.a-icon-row i')[0].get_text().strip(),
            'reviews': item.select('.a-icon-row a')[1].get_text().strip(),
            'url': "https://www.amazon.in/" + item.select('.a-icon-row a')[1]['href']
        }
    except IndexError:
        continue
    item_details.append(data)
dataframe = pd.DataFrame(item_details)
return dataframe

```

the requests.get() (2nd line) function returns a response of data of the entire web page. Next, converting the received data into an HTML format using Beautiful Soup and lxml module. Now use the CSS selectors to fetch the needed data.

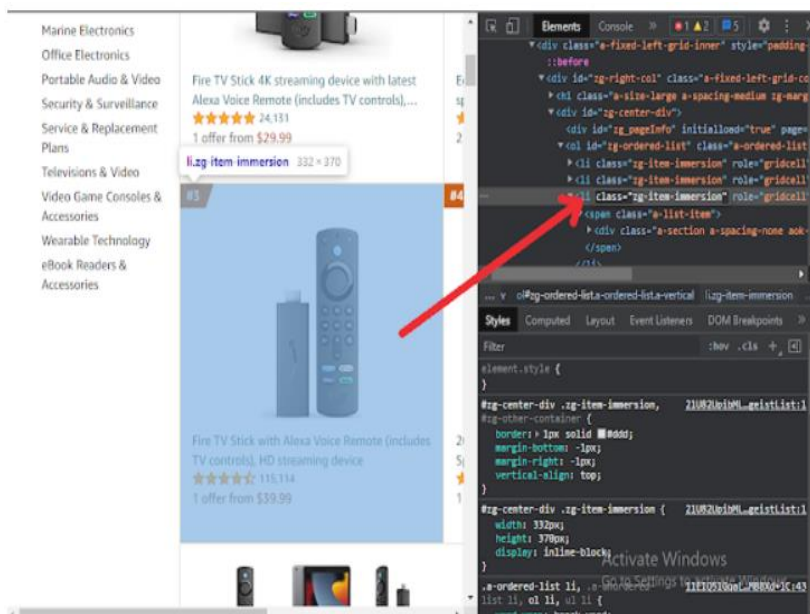


Figure 3: CAPTION (Source - <https://www.amazon.in/> product html code)

After getting all the selectors required, loop through each of the products on the web page for finding the name, price, reviews, etc, and store all those details in a python dictionary we named as data in the code. Then the entire dictionary consisting of product details is appended to a list named item details. After that, the item details is converted into a pandas data frame and simply returned it.

By calling the function with best sellers products URL as the parameter

```
df = scrape_product_details('https://www.amazon.com/Best-Sellers-Electronics/zgbs/electronics/ref-zg_bs...')
df
```

	item	price	rating	reviews	url
0	Fire TV Stick 4K streaming device with latest ...	\$29.99	4.8 out of 5 stars	24,132	https://www.amazon.in/product-reviews/B08XVYZ...
1	Echo Dot (3rd Gen, 2018 release) - Smart speak...	\$19.99	4.7 out of 5 stars	809,313	https://www.amazon.in/product-reviews/B07FZBS...
2	Fire TV Stick with Alexa Voice Remote (include...	\$39.99	4.7 out of 5 stars	115,116	https://www.amazon.in/product-reviews/B08C1W5...
3	2021 Apple 10.2-inch iPad (Wi-Fi, 64GB) - Spac...	\$299.00	4.8 out of 5 stars	2,438	https://www.amazon.in/product-reviews/B09G9FP...
4	Amazon Smart Plug, Works with Alexa – A Certif...	\$24.99	4.7 out of 5 stars	467,468	https://www.amazon.in/product-reviews/B07KTYJ...
5	Introducing Fire TV Stick 4K Max streaming dev...	\$54.99	4.7 out of 5 stars	13,522	https://www.amazon.in/product-reviews/B08MQZX...
6	Roku Express 4K+ 2021   Streaming Media Player	\$24.00	4.7 out of 5 stars	14,226	https://www.amazon.in/product-reviews/B0916TK...

Figure 4: CAPTION (Source: <https://www.amazon.in/> product URL list )

This will get the products listed with their names, price, rating, reviews, and the URL as pandas Data Frame. Now you can simply save it as a CSV file using: `df.to_csv('amazon_scrape.csv')`.

## 7. ANALYSIS FINDINGS

Web Scraping is a smart technique that can be utilized to extract vast amounts of information from the target websites. Web scraping advantages and processes can automate the processes of extracting data from the website using scripts. Data is the new differentiator. The market analysis always concludes that market leaders know the tactics to play with the current data trends and how to use them in their favor. Data is the core of market research and business strategies. This is where web scraping comes in. Web scraping can tremendously help E-Commerce help them build strategies to improve their sales, automate some of their processes

like updating product information, and ultimately grow their business. Some Other Business Use Cases Of Web Scraping Gathering data from multiple sources for

- Market analysis and Lead generation
- In-depth Research
- Data Integration

Monitoring

- Competitor's inventory information
- Stock prices
- Order status from e-commerce portals
- Opportunities
- Automation of repetitive tasks
- Procuring inventory
- Getting product reviews

## **8. LIMITATION & FUTURE SCOPE**

It is not always easy to know which site to scrape. Sometimes choose between scraping the data owner or an aggregator site.

- The internet is dynamic. Each web site has a particular structure, which may be subject to changes anytime.
- Data is volatile.
- Legal issues and informing web site owners.

## **9. FUTURE SCOPE**

Web scraping has been a big part of the internet since the beginning, and it's still going strong. In fact, web scraping is one of the most widely used tools in today's digital landscape. Web scraping may be changing is the growth of artificial intelligence (AI) and machine learning (ML). These technologies allow for more accurate and real-time processing of information than ever before. This means that web scrapers will have to adapt to these new tools and quickly.

## 10. CONCLUSION

Summing up, web data scraping can save the time it takes to examine large amounts of information. This groundbreaking method can also automate tasks like data collecting and analysis, which can significantly speed up the process of E-commerce development.

Using a web scraping application will help you to:

- generate new leads,
- find new ways to stand out in the market,
- optimize pricing,
- increase income,
- get reliable partners.

Web Scraping really works for the benefit of the business.

## 11. REFERENCES

- [1] Task - Specific Information Retrieval System for Software Engineers. Journal of Computer and System Sciences, 78, Adam, R. Tony Paul (Jul 7, 2022) [blog.datahut.co Web Scraping in Python – How to Scrape an eCommerce Website with 20 lines of Python code https://www.blog.datahut.co/post/ecommerce-web-scraping](https://www.blog.datahut.co/post/ecommerce-web-scraping)
- [2] Web scraping from Wikipedia using Python – A Complete Guide , By Garin gh (garingh128) <https://www.geeksforgeeks.org/web-scraping-from-wikipedia-using-python-a-complete-guide/>
- [3] Web Scraping Wikipedia with BeautifulSoup , Muktha Sai Ajay (Mar 2, 2020) , Data Driven Investor <https://medium.datadriveninvestor.com/web-scraping-wikipedia-with-beautifulsoup-a5d5fe6454ee>
- [4] Web Scraping in Python – How to Scrape an eCommerce Website Using Beautiful Soup and Pandas, Manthan Koolwal (March 23, 2021), freeCodeCamp.org <https://www.freecodecamp.org/news/scraping-ecommerce-website-with-python/>
- [5] Mitchell, R. (2015). Web Scrapping Using Python, First Edition. Orilley.
- [6] How-to-web-scrape-data-from-ecommerce-websites-like-amazon By Admin (August, 3 2022) <https://webautomation.io/blog/how-to-web-scrape-data-from-ecommerce-websites-like-amazon/>
- [7] Wikipedia (21 Feb 2023) E-commerce - Retrieved from wikipedia:[https://en.wikipedia.org/wiki/EcommerceWeb\\_scraping](https://en.wikipedia.org/wiki/EcommerceWeb_scraping) - Retrieved from Wikipedia Free Encycl:[https://wikipedia/web\\_scraping](https://wikipedia/web_scraping)
- [8] Judit, B.-I. Data collection methods on the web for. Springer Web Scraping , Bo Zhao University of Washington Seattle (May 2017) Encyclopedia of Big Data (pp.1-3) [https://www.researchgate.net/publication/317177787\\_Web\\_Scraping](https://www.researchgate.net/publication/317177787_Web_Scraping)
- [9] A study on Web Scraping Niranjana Krishna, Anvitha Nayak, Sana Badagan, Chethan Jetty, Dr. Sandhya N (Vol 9, Issue 12, December 2022) , (IJERCSE) [https://www.technoarete.org/common\\_abstract/pdf/IJERCSE/v9/i12/Ext\\_75106.pdf](https://www.technoarete.org/common_abstract/pdf/IJERCSE/v9/i12/Ext_75106.pdf)

- [10] Web Scraping: What It Is And How Companies Can Leverage It Sandro Shubladze, Forbes Councils Member (Jan 3, 2023) <https://www.forbes.com/sites/forbestechcouncil/2023/01/03/web-scraping-what-it-is-and-how-companies-can-leverage-it/?sh=2e02c13655a4>
- [11] Slideshare. (2021, 1 10). the history of webscraping. Retrieved from slideshare:<https://fr.slideshare.net/promptcloud/the-history-of-web-scraping>
- [12] The Upwork Team The 10 Best Web Scraper Tools (Dec 13, 2021) <https://www.upwork.com/resources/best-web-scraper>
- [13] Devin Pickell (February 17, 2023) What Is Web Scraping? How to Automate Web Data Collection <https://www.g2.com/articles/web-scraping>
- [14] xbyte. (2021, 1 10). the history of webscraping. Retrieved from xbyte: <https://www.xbyte.io/the-history-of-web-scraping.php>.