|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Technical  Documentation | |  | | |
|  |  |
|  |  |
| Inside this Report | |
| PG. 2  Why is Web Scrapping Used for?  PG. 3  Give us a blurb about a story on page 3. Give a quote from the story to get interest.  PG. 4  Give us a blurb about a story on page 4. Make this your own! | | |  | WebSites for ScrappingWhat is Web Scrapping? **Web scraping** is the best data-collection method if you are looking to grab data on web pages. As capital flows around the globe through the Internet, web scraping is widely used among businesses, freelancers, and researchers as it helps gather web data on a global basis, accurately and efficiently  Author: - Pandula Pallewatta  Data Analyst Consultant (Part – Time)  Table Contents  [Inside this Report 1](#_Toc109318640)  [WebSites for Scrapping 1](#_Toc109318641)  [What is Web Scrapping? 1](#_Toc109318642)  [Why People Be Online 3](#_Toc109318643)  [Engagement with Web Sites 3](#_Toc109318644)  [Main Domains we are focusing 4](#_Toc109318645)  [***Ecommerce sites*** 4](#_Toc109318646)  [***Social media*** 4](#_Toc109318647)  [***Directories sites*** 4](#_Toc109318648)  [1) Amazon 4](#_Toc109318649)  [2) eBay 5](#_Toc109318650)  [3) Walmart 5](#_Toc109318651)  [4) Google 6](#_Toc109318652)  [5) Twitter 6](#_Toc109318653)  [Demonstration for Web Scrapping in Popular Site 7](#_Toc109318654)  [Chormium Testing with Amazon Site with Evidance 7](#_Toc109318655) |

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  | Why People Be OnlineEngagement with Web Sites If you spend any time online, you are certainly aware with some of the most popular websites in the world. A few of internet titans currently have unrivaled domination.  **Top Three Websites (Monthly visits):**   * Google: 92.5 billion * YouTube: 34.6 billion * Facebook: 25.5 billion   The top three websites receive 152 billion monthly views, surpassing the next 47 websites combined. In addition, because of the pandemic's impact on how we work, study, communicate, and buy, the bulk of these activities have moved online.  Alphabet-owned Google ranks best with its **flagship domain, Google.com**, which yearly processes more than two trillion search requests across its network. Google generates around 80% of its revenue from advertising. |
|  |  | |  |
|  |  | |
| Figure ‑ Top Visited Websites | |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |
| Main Domains we are focusingEcommerce sites Consistently the most scraped websites in terms of frequency and amount. As online buying becomes an integral part of everyday living, ecommerce impacts people from all walks of life. Consumers, online vendors, and storefront stores all gather ecommerce data. Social media Consists of an abundance of data on human ideas, feelings, and daily behaviors. In general, scraping from social networking sites is more difficult than from other sites. This is because several social media sites implement robust anti-scraping mechanisms to preserve the privacy of their members. However, social media remains a vital source of data for sentiment analysis and other types of research. Directories sites It is not unexpected at all that he placed second in the race. Websites that categorize companies according to categories act as a functional information filter, making them an excellent choice for efficient data collecting. Many websites are scraped for contact information to increase sales leads.  **Here are the Most 5 popular websites using web scrapping to identify customer engagements** Amazon Yes, it is not unexpected that Amazon ranks the most scraped website. Amazon is grabbing the huge shares in the ecommerce sector which implies that Amazon data is the most representative for any form of market research. It has the largest database.  Obtaining ecommerce data, on the other hand, is difficult. The captcha may be the most difficult aspect of Amazon scraping, but we have it covered. Captcha prevents the site from failing since too many people want Amazon data and repeated scraping might overburden the system. Many web scrapping tools tend to use **cloud extraction** and **IP rotation** to completely nail it.  *Scraping from Amazon can give you data for all below purposes:*   * Price tracking * Competition analysis * MAP monitoring * Product selection * Sentiment analysis     Figure ‑ amazon E-Business eBay Ecommerce websites are usually the most popular for web scraping, and eBay is without a doubt one of them. We have many users that operate their own eBay companies and receiving data from eBay is a crucial method for them to keep track of their competition and market trends.  The client is an eBay seller who scrapes data from eBay and other ecommerce marketplaces on a daily basis, building up his own database over time for in-depth market research.  Figure ‑ Ebay Walmart If you're interested in the retail industry, this [Vox](https://www.vox.com/recode/21587779/shopping-deals-coupons-privacy-data-collection-retail) article depicts how merchants utilize data to watch every move of their consumers to encourage sales. While the truth is that data is also utilized to create a transparent market and serve the interests of consumers.  Web scraping is used to create price comparison websites. Walmart, with its tagline "Save Money, Live Better," might be one of the top targets. One of the reasons individuals are stealing from Walmart is because of this. Walmart is also a significant source of product data for merchants and grocery stores doing market research.    Figure ‑ Walmart Company Google **Google's** powerful machine learning algorithm has the potential to make it the robot that knows everyone better than their relatives and friends. It all comes down to data. What can we acquire from Google from the standpoint of an individual?  [SEO marketers](https://www.octoparse.com/blog/3-actionable-seo-hacks-through-content-scraping) can be the folks who are most interested in Google search. They scrape Google search results to monitor a set of keywords and collect TDK (Title, Description, Keywords: metadata of a web page that appears on the result list and has a significant effect on click-through rate) information for an SEO optimization approach.  Figure ‑ Google Twitter According to [Statistics](https://www.oberlo.com/blog/twitter-statistics#:~:text=Summary%3A%20Twitter%20Statistics,-Here's%20a%20summary&text=There%20are%20330%20million%20monthly,female%20and%2066%20percent%20male.), there are roughly 330 million monthly active users and 145 million daily active users on Twitter. With a significant number of users, Twitter is not only a platform for chatting and sharing, but also becomes a fantastic area for branding and marketing.  Figure ‑ Twitter Social Media  People are looking for data on Twitter for a variety of purposes, including industrial research, sentiment analysis, customer experience management, and so on. And, as you can see from this article about text mining Donald Trump's tweets, tweet data may be utilized in a variety of ways. | | |

|  |  |  |
| --- | --- | --- |
| Demonstration for Web Scrapping in Popular SiteChormium Testing with Amazon Site with Evidance |  |  |
|  |  |
|  |  |
| When doing the research, have gone through some of the expertise opinions. After thoroughly investigation find out that these are the main libraries used in web scrapping.    Figure ‑ Top Efficient Web Scrapping Libraries  Loading all the required library for web scrapping and saving the data into csv file.    Figure ‑ Using BeautifulSoup for webScrapping    Figure ‑ Web Scrapping From Input Keyword in Amazon Site  **For More Reference Use this Link to Get the Function: -**  **<https://github.com/Pandula1234/Data-Analyst-Consultancy/tree/master/Web%20Scrapping%20Using%20Amazon>** | | |
|  | | |
| **References**   1. Neufeld, D., Desjardins, J. & Routley, N. (2021) *Ranked: The 50 Most Visited Websites in the World*. 27 January 2021. Visual Capitalist. <https://www.visualcapitalist.com/the-50-most-visited-websites-in-the-world/> [Accessed: 18 July 2022]. 2. Anon (n.d.) *Web Scraping Services & Software | Octoparse*. Web Scraping Services & Software | Octoparse. <https://www.octoparse.com/product> [Accessed: 20 July 2022]. 3. Anon (2020) *What is Web Scraping and How to Use It? - GeeksforGeeks*. 22 June 2020. GeeksforGeeks. <https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/> [Accessed: 16 July 2022]. | | |
|  | | |