

IS WEB SCRAPING LEGAL OR NOT ?

Top 9 Aspects to Understand Whether Web Scraping is Legal Or Not

When legal

When Not Legal

1. Computer Fraud and Abuse Act (CFAA)

- When you don't access the data in an abusive fashion, you are legally safe. Secondly, as long as you don't use the same data for commercial purposes, you are not in violation of the CFAA.



- CFAA denotes how abusive access and use of web data violates the law, particularly for business purposes or financial gain. So when you indulge in web scraping in a way that violates the CFAA, your web scraping exercise can be termed illegal.

2. Copyright Infringement

- If you scrape the data but don't use the same data for publishing on the Internet or re-using it for commercial purposes, you are safe. Scraping is not illegal but re-using the copyrighted data for business purposes would be considered a violation of copyright laws.



- Companies may have data protected by copyright. Accessing and using this data for commercial purposes may invite legal trouble.

3. Trespass to Chattel

- As long as you don't enter the prohibited space and don't behave in a way that harms the website in any way, it would be mostly legal.



- It's like trespassing on somebody else's property. In this case, you are entering a prohibited digital space and behaving in an irresponsible and harmful way.
- Entering the prohibited space on a website and behaving in a way that harms the website can form a legal offense.

4. Robots.txt

- As long as you follow or respect the rules of Robots.txt, it is legal.
- If Robots.txt clearly prevents you from crawling or scraping, you need to ask for permission in writing from the owner of the site before you go ahead and scrape the data.



- When you openly and brazenly violate the norms laid down in Robots.txt, it becomes illegal. Not following or respecting Robots.txt can invite legal trouble.

5. Crawl Rate

- When you use a reasonable crawl rate and don't harass the site with repeated requests, your web scraping would be termed legal.
- You should also follow the crawl-delay settings provided in Robots.txt.
- If there's no specific setting or rate mentioned, you should follow a conservative crawl rate of 1 request per 10-15 seconds.



- Websites are made for human use. So they are designed to sustain a reasonable crawl rate. If you get aggressive and start using a fast crawl rate that brings down the server or harms the website, your actions become illegal.

6. API vs. scraping the data

- When you use an API if one is provided, instead of scraping data, it is perfectly legal.



- When you don't use the API given and behave in a way that harms either copyright laws or harm the website in any way, it becomes illegal.

7. Violating Terms of Service (ToS)

- When you follow and respect the Terms of Service (ToS), you are legally safe.
- If ToS clearly mentions that web scraping is not allowed, you should seek permission in writing and scrape only after you get the permission.



- When you go against the Terms of Service (ToS) and scrape web data in a way that violates its terms and harms the website or the business in any way, it becomes illegal.

8. Hitting the servers too frequently

- When you access the website with reasonable a time interval in between, it would be considered legal.
- You should also ensure that you don't send too many parallel requests and keep the number of parallel requests in control.



- When you don't follow a reasonable crawl rate and hit the server with frequent requests, leading to some sort of harm to the server or website, it can result in a legal problem.

9. Going beyond the Public Content

- As long as you scrape the data given in the public domain, you would be safe.
- If you don't re-use or re-publish it for financial gain, it would be legal.



- When you go beyond the public content and encroach upon the prohibited data and scrape it for business or financial gain.