

Forum:[Hack Forums](#)

Thread: [Data from website, web scraping not allowed?](#)

LiamGallagher - August 12, 2022, 3:24 pm

Hello,

In the future I would like to make a website which compares holidays from several dutch websites. I thought I could do this with web scraping, but I think this is not allowed for most websites (looked up the robots.txt file).

So now I have a few questions:

- Are there any other ways to get the data from the websites (data like price, date, insurance etc) if web scraping is not allowed? There is no API available.
- Is it true that I am not allowed to use web scraping from these websites?

Below I posted two pages I would like to receive some data from:

Quote:

https://www.vakantiediscounter.nl/last-minute?resultsperpage=18&origin=%2Flast-minute&holidaytype=sun&trip_duration_range=6-10&flexibility=2&sort=popularity_desc&special_deals=LastMinute&transporttype=VL&layout=list&room=2_0_0&live=true&facets=nodate

Quote:

[https://www.tui.nl/reizen/?_smstate=1\\$37_30005_4\\$16_30001_8\\$-17_30002_1\\$-18_30003_47\\$46_-1_-1](https://www.tui.nl/reizen/?_smstate=1$37_30005_4$16_30001_8-17_30002_1-18_30003_47$46_-1_-1)

Fresku - August 13, 2022, 2:15 am

Web scraping is legal if you scrape data publicly available on the internet. But you should avoid scraping personal data or intellectual property.

LiamGallagher - August 16, 2022, 3:07 pm

(Aug 13, 2022 02:15 AM)Fresku Wrote: [\[CITE\]](#) Web scraping is legal if you scrape data publicly available on the internet. But you should avoid scraping personal data or intellectual property.

Thank you for your reply, this is clear for me.

steelseries - August 21, 2022, 8:04 pm

Download the source html, then scrape it with regex.

Omniscient - August 23, 2022, 11:38 am

Scraping isn't illegal as far as I know. What might be is how you use the data. For example there could be a copyright violation for scraping certain content and posting it on your own website. Usually this does not apply to data though.

When you say you are comparing holidays, do you mean pricing? Because that should be fine.

Many sites take steps to block scraping. You may have to do it in a very sneaky way and even then you could find yourself blocked at some point having to make an adjustment. For example, HF gets

scraped constantly and when I find the bots I block their datacenters and take measures against them. But eventually they adjust and scrap again. I think it's China or Russia government agencies doing it mostly but sometimes it's probably other hackers. Hard to say.