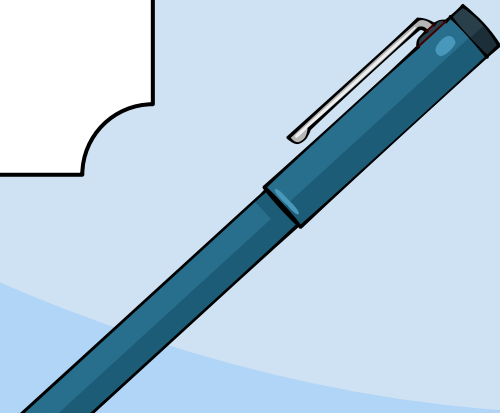


# Web Scrapping

(مقدمات)

سارا خرّم



## سرفصل مطالب



- تعریف Web Scraping
- اهمیت و کاربرد Web Scraping
- نحوه استخراج دیتا در Web Scraping
- تعریف Web Crawling
- کاربرد Web Crawling
- مقایسه دو مفهوم Web Scraping و Web Crawling

## سرفصل مطالب



- سوال: آیا Web Scrapping قانونی است؟
- فایل robots.txt

# تعریف Web Scrapping

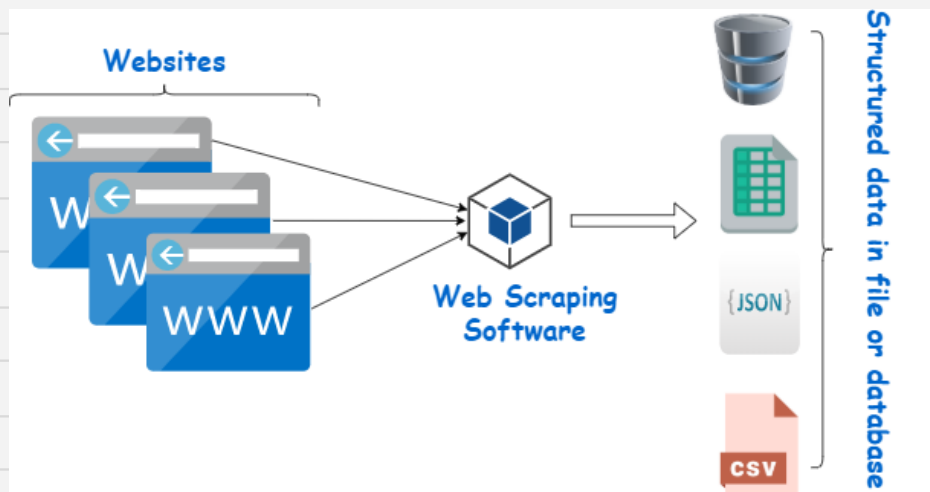
• Process دانلود کردن دیتا از وب سایتها

• عبارتهای معادل Web Scrapping

• Screen Scrapping

• Web Harvesting

• Web Data Extraction





**اهمیت Web Scrapping**

**DATA**

**DATA COLLECTION**

# کاربرد Web Scraping در حوزه‌های مختلف

Comparison & Competition Monitoring

Academic Research



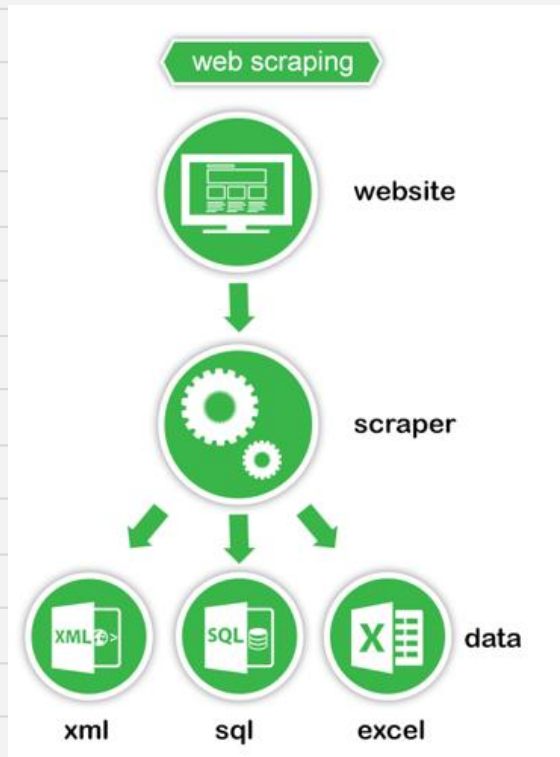
Web Scraping

Marketing

Data Analysis

Machine Learning Projects

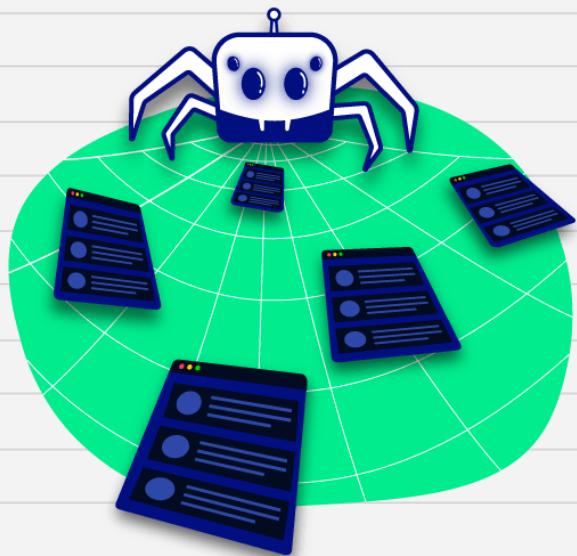
# نحوه استخراج دیتا در Web Scrapping



- دریافت URL
- Load کردن کد HTML
- استخراج دیتای مورد نظر کاربر
- ذخیره دیتا به فرمت‌های مختلف

# تعریف Web Crawling

- Web Crawler: Crawl کردن اینترنت و index کردن صفحات مربوطه برای موتورهای جستجو



- عبارتهای معادل Web Crawler

- Web Robot

- Web Spider



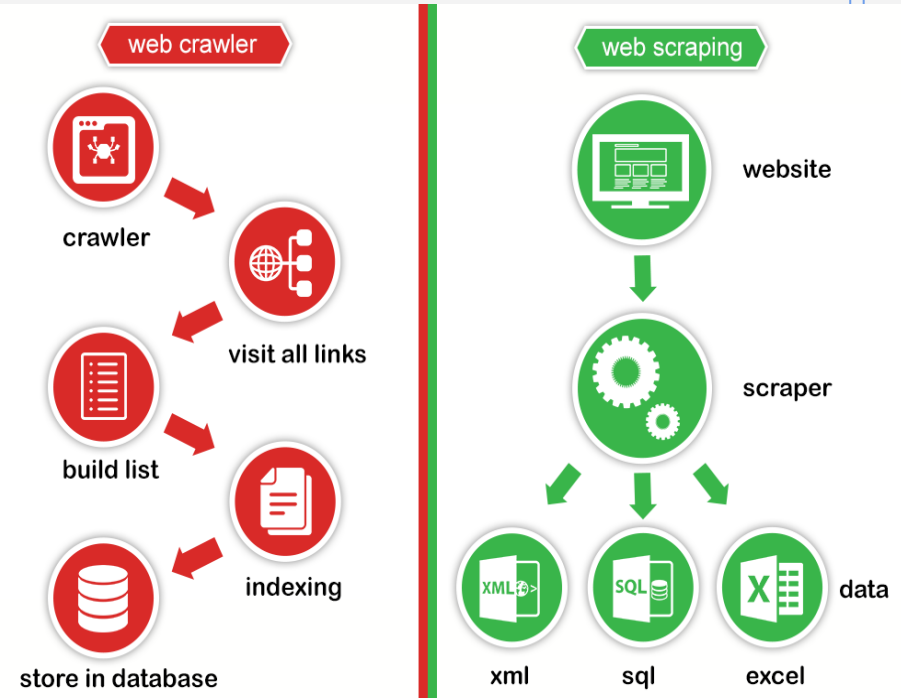
# کاربرد Web Crawling

- ارائه داده‌های به روز و up-to-date
- خودکارسازی کارهای maintenance یک وب سایت
- جمع‌آوری انواع مختلف دیتا



# مقایسه دو مفهوم Scraping

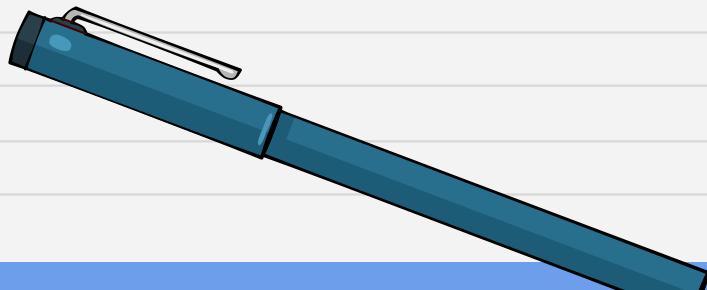
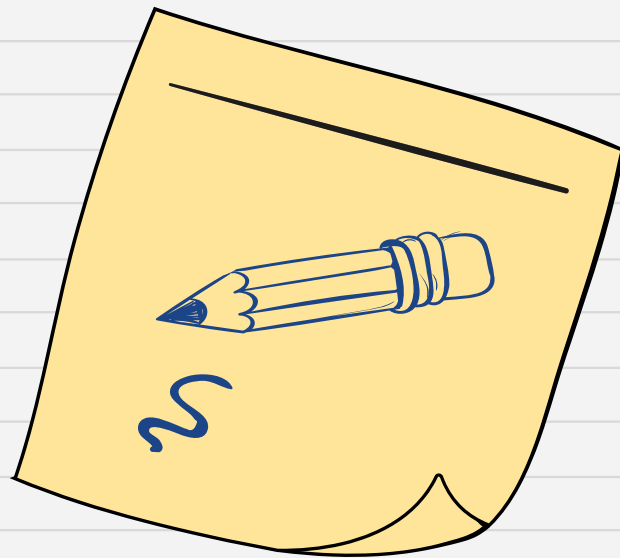
## و Web Crawling



Scraping	Web Crawling
شامل استخراج دیتا از منابع مختلف از جمله وب می شود.	به دانلود صفحات از وب اشاره دارد
کپی برداری از دیتا لازم نیست	کپی برداری از دیتا یک بخش اساسی است
در هر مقیاسی قابل انجام است	اکثرا در مقیاس بزرگ انجام می شود

# سوال؟؟؟

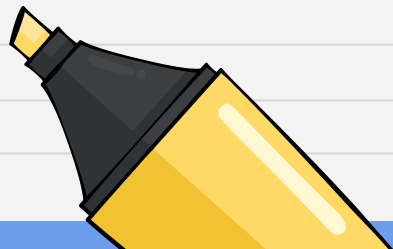
آیا Web Scrapping قانونی است؟



## Public Available Data

- کاربر برای دسترسی به دیتا، نیازی به ایجاد حساب کاربری یا ورود به سیستم را نداشته باشد.
- باتوجه به robots.txt وب سایت، Scraper ها block یا محدود نشده باشند.

جواب:



# robots.txt

• مثال : <https://medium.com/>

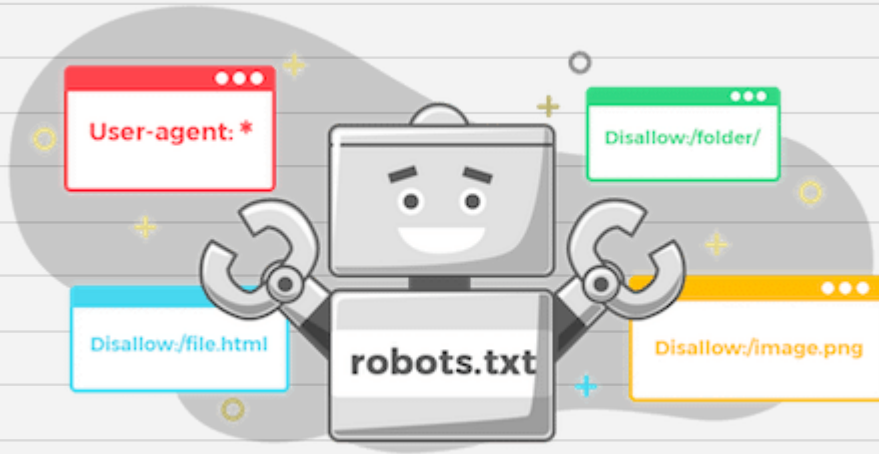
• مولفه‌های robots.txt

• User-Agent

• Disallow

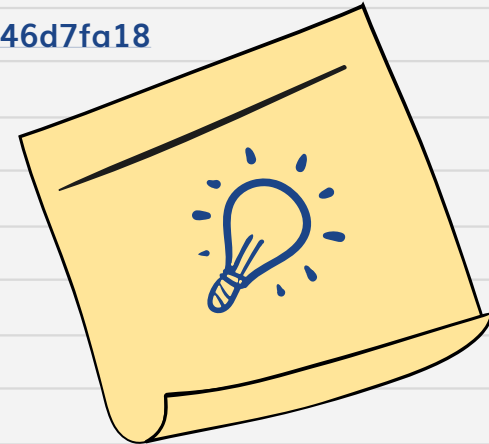
• Allow

• Sitemap



# مراجع

- [https://www.youtube.com/watch?v=aUECp\\_LS\\_1g](https://www.youtube.com/watch?v=aUECp_LS_1g)
- <https://www.webharvy.com/articles/what-is-web-scraping.html>
- <https://www.webharvy.com/articles/web-scraper-use-cases.html>
- <https://www.parsehub.com/blog/what-is-web-scraping/>
- <https://www.webfx.com/blog/internet/what-is-a-web-crawler/>
- <https://www.cloudflare.com/learning/bots/what-is-a-web-crawler/>
- [https://www.sciencedaily.com/terms/web\\_crawler.htm](https://www.sciencedaily.com/terms/web_crawler.htm)
- <https://itsvit.com/blog/big-data-scraping-vs-web-data-crawling/>
- <https://levelup.gitconnected.com/how-to-understand-a-robots-txt-file-667246d7fa18>



# Thanks!

آموزش های  
رایگان بیشتر

[www.data-hub.ir](http://www.data-hub.ir)

[www.t.me/data\\_hub\\_ir](https://www.t.me/data_hub_ir)

[www.linkedin.com/company/data-hub-ir](https://www.linkedin.com/company/data-hub-ir)