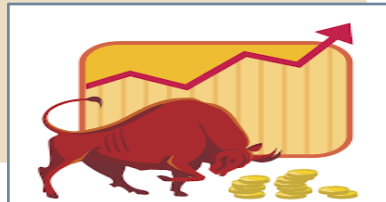




WEB SCRAPING OF NATIONAL STOCK EXCHANGE DATA

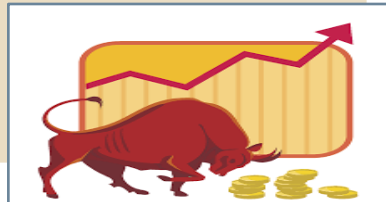
WITH CHANGE IN PRICE ALERTS

WHAT IS WEB SCRAPING?



METHODS OF WEB SCRAPING:

- Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites.
- Web scraping software may access the World Wide Web directly using the Hypertext Transfer Protocol, or through a web browser. While web scraping can be done manually by a software user, the term typically refers to automated processes implemented using a bot or web crawler.
- It is a form of copying, in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet, for later retrieval or analysis.
- Web scrapers typically take something out of a page, to make use of it for another purpose somewhere else. An example would be to find and copy names and phone numbers, or companies and their URLs, to a list (contact scraping).



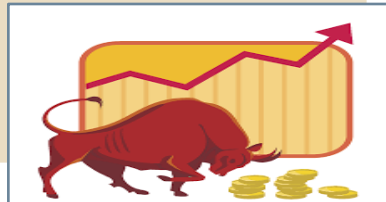
USES OF WEB SCRAPING

OFFLINE USE:

Used for contact scraping, and as a component of applications used for web indexing, web mining and data mining.

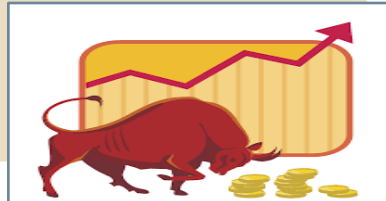
ONLINE USE:

- online price change monitoring and price comparison.
- product review scraping (to watch the competition).
- gathering real estate listings, weather data monitoring.
- website change detection, research.
- tracking online presence and reputation.
- web mashup and, web data integration.



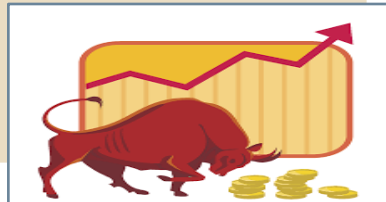
HISTORY

- The history of the web scraping is actually much longer, dating back significantly to the time when the World Wide Web, or colloquially “the Internet”, was born. After the birth of World Wide Web in 1989, the First web robot - World Wide Web Wanderer was created in 1993, June, which was intended only to measure the size of the web.
- In 1993, December, the First crawler-based web search engine - Jump Station. As there were not so many websites available on the web, search engines at that time used to rely on their human website administrators to collect and edit the links into a particular format
- In 2000, the first Web API and API crawler came. API stands for Application Programming Interface. In 2000, Salesforce and eBay launched their own API, with which programmers were enabled to access and download some of the data available to the public.
- In 2004, BeautifulSoup was released. It is a library designed for Python. As not all websites offer APIs, programmers were still working on developing an approach that could facilitate web scraping. With simple commands, BeautifulSoup could parse content from within the HTML container.



TECHNIQUES

- Human copy-and-paste
- Text pattern matching
- HTTP programming: Static and dynamic web pages can be retrieved by posting HTTP requests to the remote web server using socket programming.
- HTML parsing
- DOM parsing
- Vertical aggregation
- Semantic annotation recognizing
- Computer vision web-page analysis



SOFTWARE

EXAMPLE tools:

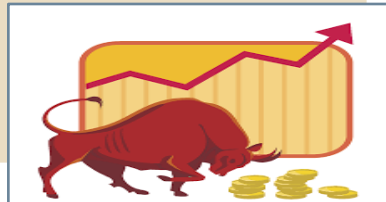
- BeautifulSoup – open source Python library for web scraping created in 2004.
- URL – command line tool and library for transferring (including getting) data with URLs supporting a wide range of HTTP methods (GET, POST, cookies, etc.).
- Data Toolbar – web scraping add-on for Internet Explorer, Mozilla Firefox, and Google Chrome Web browsers that collects and converts structured data from web pages into a tabular format that can be loaded into a spreadsheet or database management program.

JavaScript tools:

- Grease monkey
- Node.js
- PhantomJS – scripted, headless browser used for automating web page interaction.
- jQuery

Web crawling frameworks:

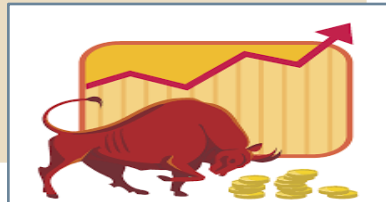
- These can be used to build web scrapers.
- Scrapy



Methods to prevent web scraping:

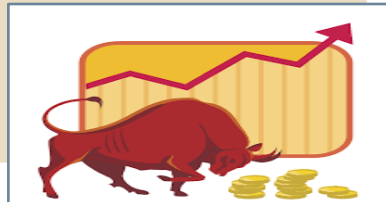
The administrator of a website can use various measures to stop or slow a bot. Some techniques include:

- Blocking an IP address either manually or based on criteria such as geolocation and DNSRBL. This will also block all browsing from that address.
- Websites can declare if crawling is allowed or not in the robots.txt file and allow partial access, limit the crawl rate, specify the optimal time to crawl and more.
- Bots sometimes declare who they are (using user agent strings) and can be blocked on that basis using robots.txt; 'googlebot' is an example. Other bots make no distinction between themselves and a human using a browser.
- Bots can sometimes be blocked with tools to verify that it is a real person accessing the site, like a CAPTCHA. Bots are sometimes coded to explicitly break specific CAPTCHA patterns or may employ third-party services that utilize human labor to read and respond in real-time.
- Commercial anti-bot services: Companies offer anti-bot and anti-scraping services for websites. A few web application firewalls have limited bot detection capabilities as well. However, many such solutions are not very effective.
- Locating bots with a honeypot or other method to identify the IP addresses of automated crawlers.
- Obfuscation using CSS sprites to display such data as phone numbers or email addresses, at the cost of accessibility to screen reader users..
- Disabling any web service API that the website's system might expose.



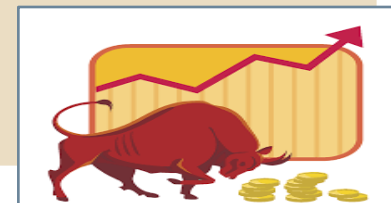
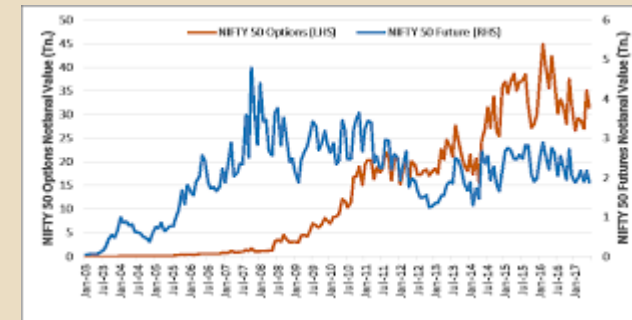
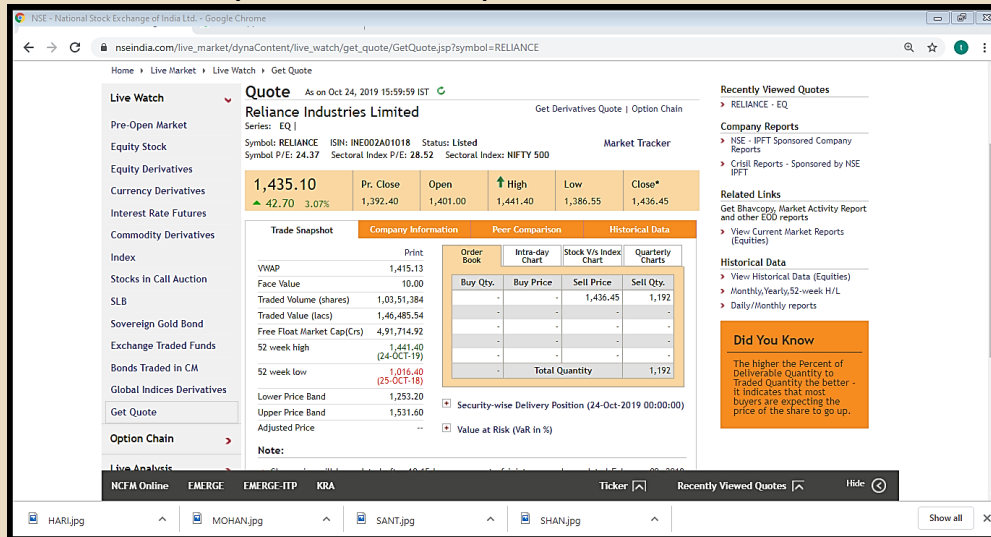
SHARE MARKET

It is a place where **shares** of public listed companies are traded. The primary **market** is where companies float **shares** to the general public in an initial public offering (IPO) to raise capital. A stock may be bought or sold only if it is listed on an exchange.



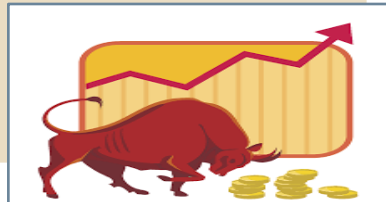
NSEINDIA

- The National Stock Exchange of India Limited (NSE) is the leading stock exchange of India, located in Mumbai
- National Stock Exchange has a total market capitalization of more than US\$2.27 trillion, making it the world's 11th-largest stock exchange as of April 2018.
- It has 25 years of experience in stock exchanging in live market analysis



ALERT MESSAGE TO USER

- It is no use that we grab the data or read in it.
- So that we implemented the data to online mail alert to user. Because all are not have enough time to check simultaneously.
- So we drop a code to read and send an info user as per their request.
- i.e., For example if the share value is 1000 when you buy.
 - If suddenly the price increase to 1100 in an hour, gain is 100.
 - If suddenly the price decrease to 950 in another hour loss is 50.
 - User cannot monitoring stock all day instead he put an alert telling that the price value is decreasing at some constant value 900. He got mail.





Thank you for giving us this opportunity

