

Web Scraping

Author: Anusha J

Co-Author: DR. Samitha Khaiyum

AUTHOR M.C.A. Post Graduate Students D.S.C.E

CO-AUTHOR M.C.A. Associate Professor & HOD D.S.C.E

Abstract:

The scenario of web client and information trade is fast changing. As ordinary citizens join the web what's more, begin to utilize it, heaps of new procedures are elevated to support up the system. Simultaneously, to improve PCs and system office new advances were brought which results into naturally diminishing in cost of equipment and site's connected expenses. Because of all these changes, enormous number of clients are joined and utilize the web offices. Day by day utilization of web comes in to an enormous information is accessible on web. Business, academicians, specialists all are share their ads, data on web with the goal that they can be associated with individuals fastly and without any problem. Because of trade, offer and store information on web, another issue is emerge that how to deal with such information over-burden and how the client will get or get to the best data in least endeavors. To fathom this issues, scientist spot out new method called Web Scraping. Web scraping is basic strategy which is utilized to produce organized information based on accessible unstructured information on the web. Scraping produced organized information at that point put away in focal database and dissect in spreadsheets. Customary reorder, Text grapping and regular expression matching, HTTP programming, HTML parsing, DOM parsing, Web scraping programming, Vertical conglomeration stages, Semantic comment perceiving and Computer vision site page analyzers are a portion of the basic strategies utilized for information scraping. Already most client utilizes the normal duplicate nuisance procedure for social affair and breaking down information on the web, however it is a repetitive strategy where part of information replicated by the client and store on PC documents. When contrasted with this method web scraping programming is least demanding scraping strategy. Presently a days, there are loads of programming are accessible in the market for web scraping. Our paper is centered around the diagram on the data extraction procedure for example web scraping, various strategies of web scraping and a portion of the ongoing devices utilized for a web scraping.

Keywords- Web mining, information extraction, web scraping.

I. Introduction

In the field of advertising, logical or academic research information assumes a significant job. Analyst, advertise analyser or academicians accumulate information from various sites for their better improvement. Copping of information on the site to client neighbourhood stockpiling is prohibited by the vast majority of the site authority. So that the client need to physically adapting the information from site to nearby PC document stockpiling. Be that as it may, such an undertaking is debilitating what's more, tedious. Because of such confinement web scraping strategies are presents. By utilizing web scraping methods client can remove data accessible on various site into a single database or spreadsheets. So information can be without any problem imagine and break down for additional utilization. Web scraping method is a sub-control of web mining innovation. The primary goal of this paper is to audit on

the distinctive web scraping procedures and programming which can be utilized to remove required data from sites.

Web Mining remains at the intersection of Information Recovery, Information Extraction and Data Mining. Both Data Retrieval and Information Extraction have significant jobs to follow and mine important data out of unstructured information, before it is appropriate to be prepared by information mining applications. Investigating better these strategies is incredibly important to adapt to the measure of accessible information in the Information Overload Era. Likewise, with the Web being progressively situated towards the significance of semantics and reconciliation of data, these zones of study become very imperative to address the new future patterns of the Web

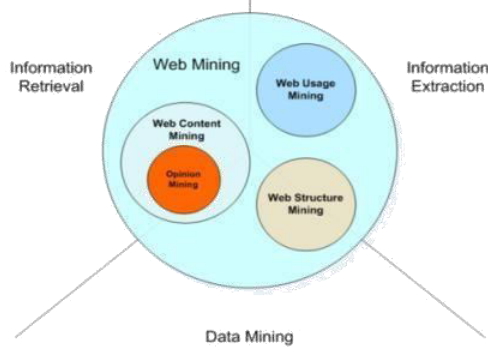


Figure 1: web mining [3]

Data Retrieval (IR) is a field of study worried about the recovery of records from an assortment of different archives (applicable and non-important), typically dependent on watchword look. With the Internet extension, Information Retrieval got a unique center, as web search tools turned into a predominant approach to discover data on the Web. These days, on account of their significance, web crawlers are the most agent name in Information Retrieval.

Data Extraction (IE) is a sub-portion of artificial intelligence. IE basically center to extricate significant information out of unstructured information. An extraction data framework is for the most part pointed on elements or articles distinguishing proof (individuals, places, organizations, and so forth) and extraction rules. Recordings, pictures, sound and content are a portion of the basic case of unstructured information. The primary data extraction frameworks were for the most part engaged on content, and still these days this is the most investigated sort of information by the examination network and business systems. The objective of IE is to distinguish valuable parts out of crude information (unstructured information) and concentrate them to at long last form more important data through semantic characterization. The outcome might be reasonable for other data handling assignments, for example, IR and Data Mining. Distinctively there is a distinction between the objectives of IR and IE, however in the genuine world they ought to be viewed as close reciprocal exercises to improve their exactness and precision [3].

Highlight that the term extraction characterize that the data which client need is expressly. The requirement for this clarification is to separate between statistical information mining strategies which reason data out of accessible information. Anyway the data recovery and data extraction can manufacture information from a given information set, it doesn't imply that they can be utilized on the other hand. As a matter of fact, data extraction is an essential pre-preparing venture to structure information before a measurable information mining calculation can manufacture information (now despite everything covered up) from it. To separate a helpful and significant information from a recovered data various procedures are utilized. Among those methods, in this

paper we are center around the web scraping procedures and examine on the a portion of the devices which are accessible for web scraping [3].

II. OVERVIEW OF WEB SCRAPING

Web Scraping is significant strategy utilized for extricating unstructured information from the sites and changing that information into organized. Web Scraping is additionally recognized as web information extraction, web information scraping, web gathering or screen scraping. Web scraping is a type of information mining. The essential and significant point of the web scraping process is to mine data from an alternate and unstructured sites and change it into a conceivable structure like spreadsheets, database or a comma-isolated qualities (CSV) record. Information like thing estimating, stock valuing, various reports, advertise evaluating and item subtleties, can be assembled through web scraping. Separating focused on data from sites commitments to take powerful choices in business process



Figure 2. Basic Architecture of Web Scraping

As we know, scraping is a method used to edit data from site pages dependent on content schedules. Web pages are records written in Hypertext Markup Language (HTML), and all the more as of late XHTML which depends on XML. Web archives are spoken to by a tree organized called the Archive Object Model, or just the DOM tree and the objective of HTML is to indicate the configuration of content showed by Web programs

From the activity perspective, a web scraping look like manual reorder task. The distinction here is that this activity is done in a sorted out and programmed way, by a virtual PC operator. At the point when an operator is following each connection of a site page, it is really playing out a similar activity that a individual would ordinarily do while interfacing with a web site. This operator can follow joins (by giving HTTP GET demands) and submit structures (through HTTP POST), perusing through a wide range of website pages. While a PC would perform manual undertakings at the speed of a PC guidance, a human would need to think, snatch the mouse, point to the connection lastly click on it. Presently the benefit appears to be clear when a client needs to tap on a few connections before getting to the real wanted page. Be that as it may, of course, the advantage of giving solicitations at a content speed, likewise brings an issue. On the off chance that one uses web scraping without a strategy for constraining solicitations, the mentioned server may find that somebody is attempting a Denial-of-Service assault, because of the extraordinary measure of solicitations activated in a short time [3].

Next stage is, the parser follows client indicated ways inside the report to recover the ideal data dependent on the information retrieved in past advance. These ways are indicated by CSS selectors or XPATHs. They utilize both relative and absolute paths (in view of the DOM tree) to guide the parser toward a explicit component inside a web record. Ordinarily web scraping activities utilizes ordinary articulations to limited or trim the found data, so as to recover information with a user specified granular size [3].

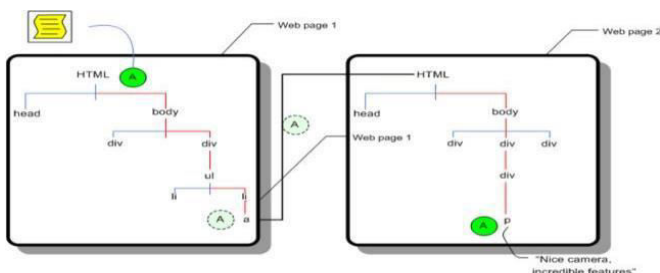


Figure 3: Last step of web scraping- Gathering information by scraping agent on the basis of DOM tree and XPATH [3].

III. WEB SCRAPER USES

Web Scrapers are additionally being utilized by Online Marketers to pull information secretly from the contender's sites, for example, high targeted keywords, valuable links, emails & traffic sources.. A portion of the zone where web scraper methods are for the most part utilizes are:

- Change detection on website
- Product Price comparison
- Weather broadcasting and data monitoring
- Research analytics
- Analyze data in graphics
- Web Indexing & rank checking
- Advertisement analysis
- Market Analysis

IV. DIFFERENT WEB SCRAPING TECHNIQUES

A. Classical copy and paste:

The human's manual assessment and reorder strategy is the best and the serviceable web scraping innovation. However, this is a tending to actualize or cause mistakes, and tedious strategy when client need to break down and store heaps of datasets [1].

B. Hypertext Transfer Protocol (HTTP) Programming:

By utilizing this procedure client can be separate data from static and dynamic site pages. Information can be recovered by presenting HTTP demands on the remote web server utilizing attachment programming.

C. Hyper Text Markup Language (HTML) Parsing:

Semi structured information question dialects, as XQuery and the Hyper Content Query Language (HTQL), can be utilized to parse HTML pages and to recover and change page content [1].

D. Document Object Model (DOM) Parsing:

By inserting a undeniable internet browser, for example, the Internet Explorer or the Mozilla program control, projects can recover the dynamic content created by customer side contents. These program controls additionally parse website pages into a DOM tree, in view of which projects can recover portions of the pages.

E. Web Scraping Software:

Now a days numerous product apparatuses are accessible, that can be utilized to tweak web-scraping arrangements. This product may endeavour to consequently perceive the information structure of a page or give a chronicle interface that evacuates the need to physically compose web scraping code, or some scripting capacities that can be utilized to remove and change substance, and database interfaces that can store the scraped information in nearby databases.

F. Computer vision web-page analyzers:

There are endeavors utilizing AI and PC vision that endeavor to recognize and remove data from site pages by deciphering pages outwardly as an individual would [1].

V. WEB SCRAPING TOOLS

Web Scraping Software are the mechanized program that are utilized to make the manual duplicate copy and paste work naturally .it additionally gather huge measure of data from sites like catalog destinations, land locales, grouped sites and occupation sheets and put away on neighborhood server. Assume you need to scrap land property subtleties of India then you have to name few folks to reorder subtleties from sites to exceed expectations by visiting every property operator pages. Along these lines it will take days what's more, even a long time to prepare your property information to utilize. So web scraping can computerize the manual work automatically by visiting each page and concentrate information from pages and parsing the html pages. There are number of Web Scraping Software that accessible in showcase that can assist you with scraping

information from any site you need. Following are the rundown of some scraping instruments [1].

Web scraping programming fill in as, to bot or web crawler get to the web information legitimately utilizing the Hypertext Transfer Protocol, or through an internet browser and concentrate the exact information from that site page. This removed information is then store into a focal nearby database or spreadsheet for some time in the future or examination.

Some of the tools are:

1. Mozenda:

It is a Business Intelligence Software. It is permits client to extricate information from reports as a similar way client can separate information from site page. Additionally client can be consolidate information from multiplesources into a solitary informational index. Mozenda as of now bolster reports scraping for a few well known arrangements like World, Excel, PDF and so forth key highlights of Mozenda are: industry Information Feeds, high-volume week after week information takes care of, venture building also, venture upkeep. Mozenda will consequently identify names and related qualities and assemble powerful informational indexes with insignificant design. Mozenda's Data Mining Software is stuffed loaded with valuable applications particularly for salesmen. Client can do things, for example, "lead age, guaging, gaining data for setting up financial plans, contender estimating investigation. This product is an extraordinary ally for advertising plan and deals plan making. Utilizing Refine Capture tex apparatus, Mozenda is sufficiently keen to channel the content you need remains clean or get the particular content or split them into pieces.

2. Visual Web Ripper:

Visual Web Ripper is one of the most development web scraping programming, made by Sequentum bunch in 2006. It offers more usefulness which will permits client to scrap information from any sites like Business Directories, Simple Web Pages, Grouped Sites, Forums and online business webpage like eBay, amazon, magento locales. In the wake of completing information scraping task information can be sent out to organized CSV, Excel, or XML group.

3. Web Content Extractor:

Another Simple and client situated device for information scraping is Web Content Extractor (WCE) which is created by New pro-soft. It has great wizard that manage client to arrangement scraper. Client can scrap information from site with scarcely any snaps and WCE is self-wise for placing information into various arrangements like Excel, content, HTML groups, Microsoft Access database, Organized Query Language(SQL) Script File, MySQL Script Document, Extensible Markup Language (XML) record, HTTP submit structure and Open Database Connectivity (ODBC) Data source.

4. Import.io:

Import.io is an online apparatus for extricating information from site without composing code. In the event that client need a quick outcome, at that point he/she will go after this instrument with the goal that he can change over site in brief timeframe. For extricating information, client enter URL and application naturally extricate information which client need to require, if client doesn't intrigued by the programmed extraction, the point and click interface help to choose information fields on site. As the information extraction is finished, the separated dataset is store on Import.io cloud server and ferther downloaded in CSV, Excel, JSON design.

5. Scrappy:

An open source and collective system for separating the information you need from sites. Crude is intended to scrap web content from destinations that are made out of numerous pages of comparable semantic structure. The framework is executed as a Firefox program augmentation, and works in three primary stages to scrap web information. Initial, a client explores to a page that he would like to scrap and creates a format for the substance that he might want from that page. Next, the client chooses a lot of connections that point to pages coordinating the substance layout characterized by the client. At last, the client chooses a yield information design and Crude slithers the connections determined by the client and scrapes content comparing to the client's template[9].Scrappy composed in Python and runs on Linux, Windows, and Mac.[9]

VI. Languages For Web Scraping:

The most ideal approach to choose the programming language for web scraping is that causes you scrap the information you need. There are numerous dialects that vie for the top spot. It's impractical for anyone to guarantee that someone or other language is the best language for web scraping. You could without much of a stretch settle on an inappropriate decision. You can wind up investing time and vitality into something that may not yield your ideal outcomes. the best programming language for web scraping is the one you definitely know. On the off chance that you know about a specific language and have involvement with programming, that would be a good thought to manufacture assets that help web scraping in that language. Since you are as of now acquainted with the language, you are probably going to come to speed a lot quicker while figuring out how to scrap with it. You can think about this as an underlying advance.

There are numerous outsider libraries accessible on the internet which encourages you out in web scraping. With the assistance of Google search, you can get your own web scraping library for the language you know.

Languages for web scraping are:

1. Python: Python is a standout amongst other programming dialects in web scraping. It's a finished item which can deal with all procedure identified with information extraction easily.

- Let us disclose to you why Python is a favored language to use for web scraping. Scrapy and BeautifulSoup are the broadly utilized structures dependent on Python that makes scraping simple.
- Python library — BeautifulSoup is intended for quick and profoundly productive information extraction.
- Scrapy has some incredible highlights like help for XPath, upgraded execution inferable from the twisted library and an assortment of troubleshooting instruments.
- Pythonic expressions are route, looking and changing a parse tree.
- The BeautifulSoup structure can likewise change over approaching archives to Unicode and active records to UTF-8.
- Lovely Soup takes a shot at famous Python parsers like lxml and html5lib, which permit you to attempt diverse parsing procedures [10].

2. Node.JS: Most definitely, Node.js is an especially incredible programming language that utilizes dynamic coding rehearses. Despite the fact that it bolsters circulated slithering, the steadiness of interchanges is generally frail and isn't suggested for enormous scope ventures.

Node.js use JavaScript occasions hover to make non-blocking I/O applications that can without a doubt advantage various concurrent occasions.

- So as to abuse the component of Node.JS, individuals utilize numerous occasions of same scraping as each Node.JS procedure takes one center on the CPU.
- Node.JS is the fitting and suggested for spilling, API, attachment based usage [10].

3. Ruby:

Ruby is one of the best open source programming dialects. It's one of the favored dialects over others as it is easy to comprehend and increasingly gainful nearly others. The linguistic structure it contains is basic and advantageous for composing. Ruby keeps up the offset of useful programming with the guide of basic programming. Ruby sets aside some effort to compose. Ruby on Rails is one of the most favored web systems that empower one to compose less code and forestall any kind of redundancy.

- NokoGiri, HTTParty and Pry can empower you to set up your web scraper with no issue.

- NokoGiri is a Rubygem that offers HTML, XML, SAX and Reader parsers with XPath and CSS selector support.
- HTTParty is the diamond that sends a HTTP solicitation to the pages that you need to remove information from. What it will achieve is that it will outfit all the HTML of the page as a string [10].

4. PHP:

PHP is maybe the least significant programming language for web scraping to assemble a crawler program. The significant downside of this language is weka support for multi-stringing and async. The errand booking and lining issues could be connected while utilizing PHP language for the web scraping. This is the reason PHP isn't suggested for web scraping. Utilizing URL library, you can extricate illustrations, recordings, and photos from different sites. Twist can move records utilizing a broad rundown of conventions including HTTP and FTP. This will assist you with making a web bug to download nearly anything from the web naturally [10].

5. C & C++:

C and C++ offer the best execution and yield, however the creating of web scraping arrangement on these languages costs you somewhat higher. Thus, it's not prudent to utilize and make a crawler utilizing C or C++ programming language.

- It's basic and straightforward.
- It's anything but difficult to parallelize your scraper utilizing C++.
- Utilizing lib curl to get URLs and afterward compose your own HTML parsing library that addresses your issues given your objective stage.
- C++ isn't the main decision for any web-related activities since it should likewise be possible utilizing any powerful language.
- Since it's very costly, it would be the last choice to use in web scraping [10].

VII. Why Web Scraping?

Web Crawling:

Web crawling is analogous to a spider crawling however the spot of crawling here is the web! It essentially visits a site and read site pages to assemble passages for web crawler list. The apparatuses that are utilized for web slithering are known as web crawlers or creepy crawlies. A progression of website pages are examined and connections to the pages on them are then followed for much more connections therefore it does a profound quest for removing of data. Acclaimed web search tools, for example, Google, Yahoo and Bing do web slithering and utilize this data for ordering site pages. Models are Scrappy and Apache nut.

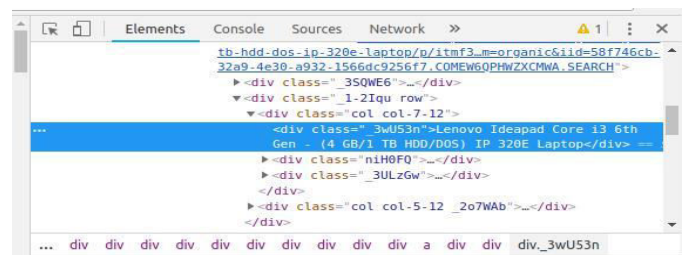
S.NO.	WEB SCRAPING	WEB CRAWLING
1.	Web crawling is what search engines do: going through the web, looking for any information you need, clicks on every link, etc. The main point of that is to collect as much information as you can. Every search engine is in some way a crawler.	When it comes to web scraping, it's a bit different. While scraping you not only looking for the information you need, you can also download it and save it for further use. It's more about the target process and web scraping is done for many purposes - market research, SEO, sales intelligence, etc.
2.	It is utilized for downloading information.	It is utilized for ordering of Web pages
3.	It need not visit all the pages of site for information.	It is utilized for ordering of Web pages
4.	Application zones incorporate Retail Marketing, Equity search and Machine learning.	Used in web indexes to give list items to the client.
5.	This needs crawl agent and a parser for parsing the response.	This only needs only crawl agent.

With the help of web scraping we can extract data from any websites, no matter how large is the data. Moreover, websites may have data that you cannot copy and paste manually or by any other technique. Sometimes you want to scrape email addresses from Facebook pages but there's no 100% guarantee that you'll get all the email addresses so the best method is to make use of a good web scraping technique. As above mentioned software is also best for web scraping but some software may charge a few \$\$ as license fee at that time we can make use of another web scraping technique to scrape the data by using programming languages.

A small example to scrape the data from web using web scraping:

- Step 1:** enter the target URL. Wait for the web page gets completely loaded within the built-in browser.
- Step 2:** Inspecting the Page

The data is usually nested in tags. So, we inspect the page to see, under which tag the data we want to scrape is nested. To inspect the page, just right click on the element and click on "Inspect".



- Step 3:** Find the data you want to extract
Let's extract the Price, Name, and Rating which is nested in the "div" tag respectively.
- Step 4:** Write the code, using python language

From selenium import webdriver
from BeautifulSoup import BeautifulSoup.

import pandas as pd

driver=webdriver.Chrome("/usr/lib/chromium-browser/chromedriver")

products=[]

prices=[]

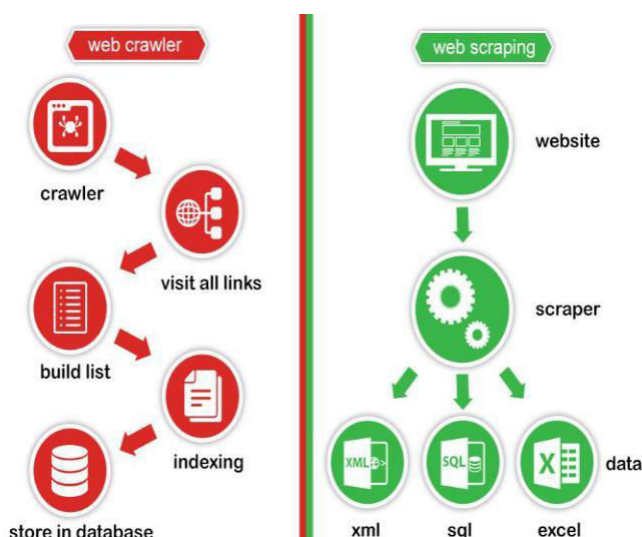
ratings=[]

driver.get("<ahref=https://www.flipkart.com/laptops/>https://www.flipkart.com/laptops/~buyback-guarantee-on-laptops-/pr?sid=6bo%2Cb5g&uniq")

content = driver.page_source

soup = BeautifulSoup(content)

for a in soup.findAll('a',href=True, attrs={'class':'_31qSD5'}):



```
name=a.find('div', attrs={'class':'_3wU53n'})

price=a.find('div', attrs={'class':'_1vC4OE _2rQ-NK'})

rating=a.find('div', attrs={'class':'hGSR34 _2beYZw'})

products.append(name.text)

prices.append(price.text)

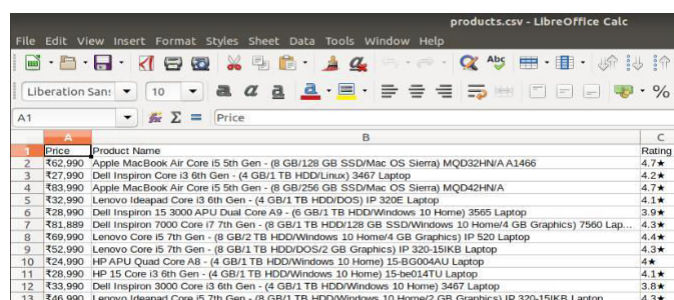
ratings.append(rating.text)

python web-s.py

df=pd.DataFrame({'Product
Name':products,'Price':prices,'Rating':ratings})

df.to_csv('products.csv', index=False, encoding='utf-8')
```

After running the above code, A file name “products.csv” is created and this file contains the extracted data.



	A	B	C
1	Price	Product Name	Rating
2	\$62,990	Apple MacBook Air Core i5 5th Gen - (8 GB/128 GB SSD/Mac OS Sierra) MQD32HN/A A1466	4.7★
3	\$27,990	Dell Inspiron Core i3 6th Gen - (4 GB/1 TB HDD/Linux) 3467 Laptop	4.2★
4	\$63,990	Apple MacBook Air Core i5 5th Gen - (8 GB/256 GB SSD/Mac OS Sierra) MQD42HN/A	4.7★
5	\$32,990	Lenovo Ideapad Core i3 6th Gen - (4 GB/1 TB HDD/DOS) IP 320E Laptop	4.1★
6	\$28,990	Dell Inspiron 15 3000 APU Dual Core A9 - (6 GB/1 TB HDD/Windows 10 Home) 3565 Laptop	3.9★
7	\$81,889	Dell Inspiron 7000 Core i7 7th Gen - (8 GB/1 TB HDD/128 GB SSD/Windows 10 Home/4 GB Graphics) 7560 Lap...	4.3★
8	\$69,990	Lenovo Core i5 7th Gen - (8 GB/2 TB HDD/Windows 10 Home/4 GB Graphics) IP 520 Laptop	4.4★
9	\$52,990	Lenovo Core i5 7th Gen - (8 GB/1 TB HDD/DOS/2 GB Graphics) IP 320-15IKB Laptop	4.3★
10	\$24,990	HP APU Quad Core A8 - (4 GB/1 TB HDD/Windows 10 Home) 15-BG004AU Laptop	4★
11	\$28,990	HP 15 Core i3 6th Gen - (4 GB/1 TB HDD/Windows 10 Home) 15-be014TU Laptop	4.1★
12	\$33,990	Dell Inspiron 3000 Core i3 6th Gen - (4 GB/1 TB HDD/Windows 10 Home) 3467 Laptop	3.8★
13	\$46,990	Lenovo Ideapad Core i5 7th Gen - (8 GB/1 TB HDD/Windows 10 Home/2 GB Graphics) IP 320-15IKB Laptop	4.3★

Data Set after scraping online scraping tool.

VIII. Conclusion

To get programmed data from site, web scraping is the best and productive method. Among all other techniques notice in this paper because other techniques is only going through the web, looking for any information we need, click on the link etc., but when it comes to web scraping it's not only looking for the information but we can also download the data and save it for further usage and also web scraping is fast, reliable which are utilized to extricate and store information, web scraping is progressively dependable, quick and automatic information recovery framework. By utilizing web scraping phrasing client can effectively remove unstructured information on single or various sites into an organized information consequently. The primary point of this strategy is to get data from web and total into a new dataset. In this paper we are talk about the essential of web mining. We likewise center around the methods utilized for web scraping. The last area of the paper advance the review of various programming for web scraping it's high time to decide and pick the right and best programming language for web scraping. However, it's really important to keep all the benefits and drawbacks in mind while choosing the right programming language.

REFERENCES.

- [1] S.C.M. de S Sirisuriya 2015, A Comparative Study on Web Scraping .Proceedings of 8th International Research Conference, KDU. <http://ir.kdu.ac.lk/handle/345/1051>
- [2] List of Web Harvester, Data Scraper, Web Scraping Software and Tools, n.d. WebData Scraping. <http://webdata-scraping.com/webscraping-software/>
- [3] Felipe Jordão Almeida Prado Mattosinho, Master Thesis, Mining Product Opinions and Reviews on the Web, TU Dresden. https://www.rn.inf.tudresden.de/uploads/StudentischeArbeiten/Masterarbeit_Mattosinho_Felipe.pdf
- [4] Eloisa Vargiu, Mirko Urru1, 2013, Exploiting web scraping in a collaborative filtering- based approach to web advertising, Artificial Intelligence Research, 2013, Vol. 2, No. 1, <https://dx.doi.org/10.5430/air.v2n1p44>.
- [5] Schrenk, M. Webbots, spiders, and screen scrapers: a guide to developing Internet agents with PHP/CURL. No Starch Press, 2007.
- [6] OsmarCastrillo-Fernández, 2015, Web Scraping: Applications and Tools, European Public Sector Information Platform, Topic Report No. 2015/10.
- [7] Deepak Kumar Mahto, Lisha Singh, A Dive into Web Scraper World, 2016 International Conference on Computing for Sustainable Global Development (INDIACom), 978-9-3805-4421-2/16/\$31.00 c , 2016 IEEE. <https://ieeexplore.ieee.org/document/7724353/>
- [8] MiloslavBeiiio, Jakub Misek, Filip Zavoral, AgentMat: Framework for Data Scraping and Semantization, 9781- 4244-2865-6/09/\$25.00 ©2009 IEEE. <https://scholar.google.cz/citations?user=EDwbpaoAAA&hl=en>.
- [9] Amir Ghazvinian, Sean Holbert, Nikil Viswanathan, Scrappy: Simple Web Scraping, Department of Biomedical Informatics, Department of Computer Science, Stanford University. <https://seanolbert.wordpress.com/2011/07/15/scrappy-simple-web-scraping/>.
- [10] X-Byte Enterprises Crawling, medium.com: Top languages for web scraping. <https://medium.com/>