

A Mini-Project Report on :

# **DATA SCRAPING USING BLUE PRISM**

Submitted By

|                         |                     |
|-------------------------|---------------------|
| <b>L. Lokesh Goud</b>   | <b>(19R11A04N0)</b> |
| <b>K. Naithikasree</b>  | <b>(19R11A0473)</b> |
| <b>G.Kusuma Clayton</b> | <b>(19R11A0465)</b> |

Under the Guidance of  
**Mr. Hemant Kumar Gahlot**



**Department of Electronics and Communication Engineering**

**GEETHANJALI COLLEGE OF  
ENGINEERING AND TECHNOLOGY**  
AUTONOMOUS

(Accredited by NBA, Approved by AICTE, Affiliated to JNTUH)  
Cheeryal Village, Keesara (M), Medchal Dist, Hyderabad - 501301)

| S.No | CONTENTS   | PAGE.No |
|------|--|---------|
| 1.   | INTRODUCTION<br>1.1 Overview<br>1.2 Purpose                                    | 1-2     |
| 2.   | LITERATURE SURVEY<br>2.1 Existing Problem<br>2.2 Proposed Solution             | 2       |
| 3.   | THEORITICAL ANALYSIS<br>3.1 Block Diagram<br>3.2 Hardware / Software Designing | 3-4     |
| 4.   | EXPERIMENTAL INVESTIGATIONS  | 5       |
| 5.   | FLOWCHART  | 6       |
| 6.   | RESULT   | 7-8     |
| 7.   | ADVANTAGES AND DISADVANTAGES   | 9       |
| 8.   | APPLICATIONS   | 10      |
| 9.   | CONCLUSION   | 10      |
| 10.  | FUTURE SCOPE   | 11      |
| 11.  | BIBILOGRAPHY   | 12      |

# 1. INTRODUCTION

## 1.1 Overview

Data Scraping is the process of importing information from a website into a spreadsheet or local file. It is also known as *Web Scraping*. Thus, the key element that distinguishes data scraping from regular parsing is that the output being scraped is intended for display to an end-user, rather than as an input to another program. Automation can possibly yield extraordinary advantages for organizations and associations. Notwithstanding, this sort of information extraction is for all intents and purposes inconceivable when you need a lot of information from numerous websites for a business use case. This is when automated web scraping comes into the image. Digital Workers can finish the extracting data from any stage ,viably and precisely. Executing the automated tools for data scraping will make a grating between the current and recently detailed practices. This investigation inspects the extraction of data from Flipkart website as far as both individual just as organizational level. Workers may see this effect in an unexpected manner in comparison to the association would.

## 1.2 PURPOSE

By this point, you can likely think about a few unique manners by which web scrubbers can be utilized. We've put probably the most well-known ones beneath (in addition to a couple of novel ones).

- Scraping stock costs into an application Programming interface
- Scraping data from Yellow Pages to create leads
- Scraping data from a store finder to make a rundown of business areas
- Scraping item data from locales like Amazon or eBay for contender examination
- Scraping sports details for wagering or dream associations
- Scraping webpage data before a site relocation
- Scraping item subtleties for correlation shopping
- Scraping monetary data for statistical surveying and bits of knowledge

The rundown of things you can do with web scraping is practically perpetual. All things considered, it is about how you can manage the data you've gathered and how important you can make it. By utilizing this strategy we can have the option to make our work simple and quick with high exactness rate which is additionally extremely accommodating for some organizations and the freelancers.

## 2. LITERATURE SURVEY

### 2.1 Existing Problem

Sites for the most part depict in their terms of utilization and in their robots.txt document on the off chance that they permit scrubbers or not. We can't see the data of the all items all at once, it is be exceptionally hard to open every single item to see the subtleties of the item. Indeed, even through subsequent to scratching the data through some application we can see all the undesirable information associated and connected to it. Some of the time the laptop/PC which we are utilizing, gets stuck out because of a great deal of the extraction data and furthermore the undesirable data gets over-burden with the real substance which we need.

### 2.2 Proposed Solutions

While organizations might have the option to do limited scope data extraction, challenges emerge when they shift to enormous scope extraction. Here in the Blue Prism what we have done is, we just made data items and additions the matter what we need to extract, so that in the yield it is just noticeable barring every one of the undesirable substances. It is additionally an exceptionally fast interaction and furthermore we can restrict the check of items as per our need. This solitary takes or gives the required and wanted yield other than undesirable stuff. we have additionally kept the alarm stage to get the information. It shows notifications likewise to the Facebook notifications you get inside your program when you remain signed in.

## 3. THEORITICAL ANALYSIS

### 3.1 Block Diagram

Here in the beneath we could see the diagrammatic outline of the data scraping flow :



### 3.2 Hardware / Software Designing

#### 3.2.1 HARDWARE REQUIREMENTS :

The hardware prerequisites are the necessities of a hardware gadget. Most hardware just has working framework prerequisites or similarity. For instance, a printer might be viable with Windows XP yet not viable with more up to date forms of Windows like Windows 10, Linux, or the Apple macOS. In the event that a hardware gadget isn't viable with your PC, it is dependent upon the producer to deliver drivers.

The hardware requirements are the requirements of a hardware device. Most hardware only has operating system requirements or compatibility. Shockingly, numerous producers just delivery refreshed drivers to fix issues with more seasoned drivers and regularly don't deliver drivers for more up to date working frameworks or option working frameworks.

Coming up next is the Hardware needed to finish this task:

- Internet connection to download and activate
- Administration access to install and run Blue Prism
- Minimum 10GB free disk space
- Windows 8.1 or 10 (64-bit version only)OR CLOUD: Get started free, \*Cloud account required.
- Minimum System Requirements to run Office Excel 2013, your computer needs to meet the following minimum hardware requirements:
  - a. 500 Megahertz (MHz)
  - b. 256 Megahertz (MB) RAM
  - c. 1.5 Gigabytes (GB) available space
  - d. 1024x768 or Higher resolution monitor

### 3.2.2 SOFTWARE REQUIREMENTS :

Requirement is a condition or capacity moved by the software or system segment to tackle a genuine issue. The issues can be to mechanize a piece of a system, to address deficiencies of a current system, to control a gadget, etc. A condition or capacity that should be met or moved by a system or system segment to fulfill an agreement, standard, detail, or other officially forced reports. Requirements portray how a system should act, show up or perform. Requirements vary starting with one client then onto the next and starting with one business measure then onto the next.

- Operating System : Windows XP /Vista or any main stream OS
- Installation and Setup Guide for Blue Prism
- Installation and Setup Guide for MS Excel
- Blue Prism Version: 6.10.1
- Blue Prism License File
- Blue Prism Installation Software 64 bit
- MS Excel
- Windows 10

## 4. EXPERIMENTAL INVESTIGATIONS

Data scraping is the cycle of naturally mining data or gathering data from the Internet. It is a field with dynamic improvements offering a shared objective to the semantic web vision, a yearning drive that actually requires forward leaps in text preparing, semantic agreement, man-made brainpower and human-PC collaborations. Current web scraping arrangements range from the specially appointed, requiring human exertion, to completely computerized frameworks that can change over whole sites into organized data, with limits. In December 1993, the primary crawler-based web search tool, JumpStation, was dispatched. In examination, JumpStation brought another jump, being the primary WWW web crawler that depended on a web robot.

Web Scraping or Data Crawling or Data Gathering has been into the presence however long the actual Web. The reason behind business web scraping has consistently been to acquire simple business benefits like contender's item costs, taking leads, commandeering showcasing efforts, diverting APIs, and the through and through burglary of substance and data.

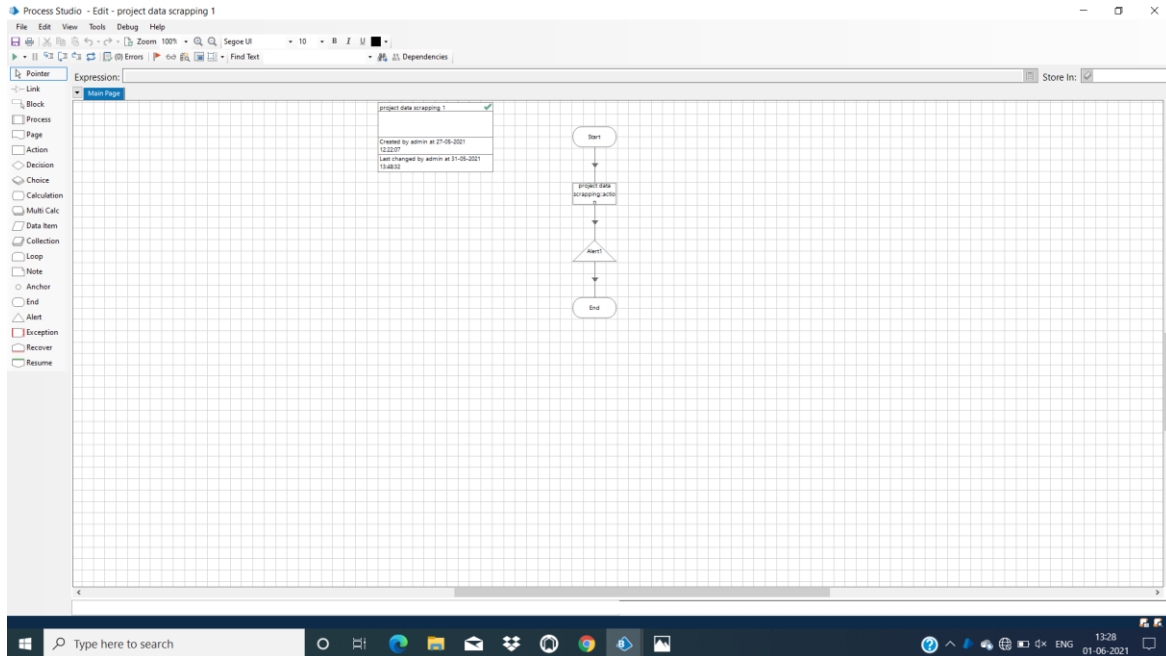
Web scraping is the technique which assists with taking or concentrate the substance from a website with the plan of utilizing it for purposes outside the immediate control of the webpage proprietor. The main utilization of web scraping was to connect with testing structures.

Web scraping is similar to web ordering, the interaction by which web crawlers record web content. Web indexers ("great bots") observe the guidelines; web scrubbers, then again, essentially take whatever substance they've been modified to bring – costs, advancements, offers, or data that would somehow or another just be accessible to paid endorsers or approved colleagues.

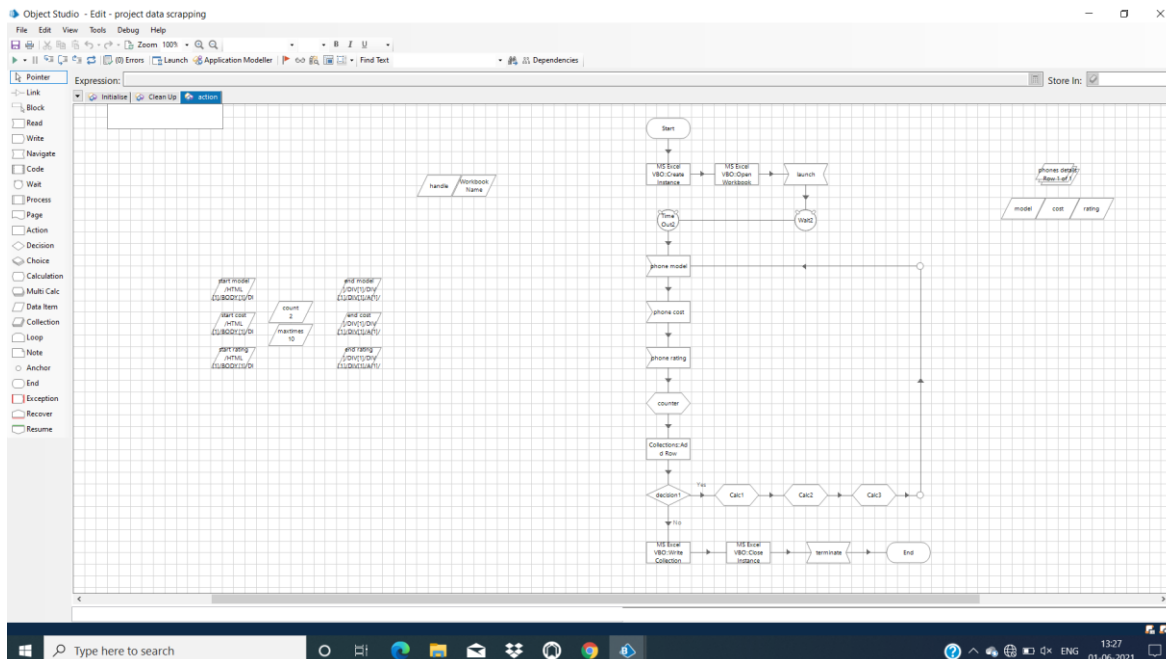
# 5. FLOWCHART

The below picture are the flows in Process and Object studios

Process studio's Flow:



Object studio's Flow :

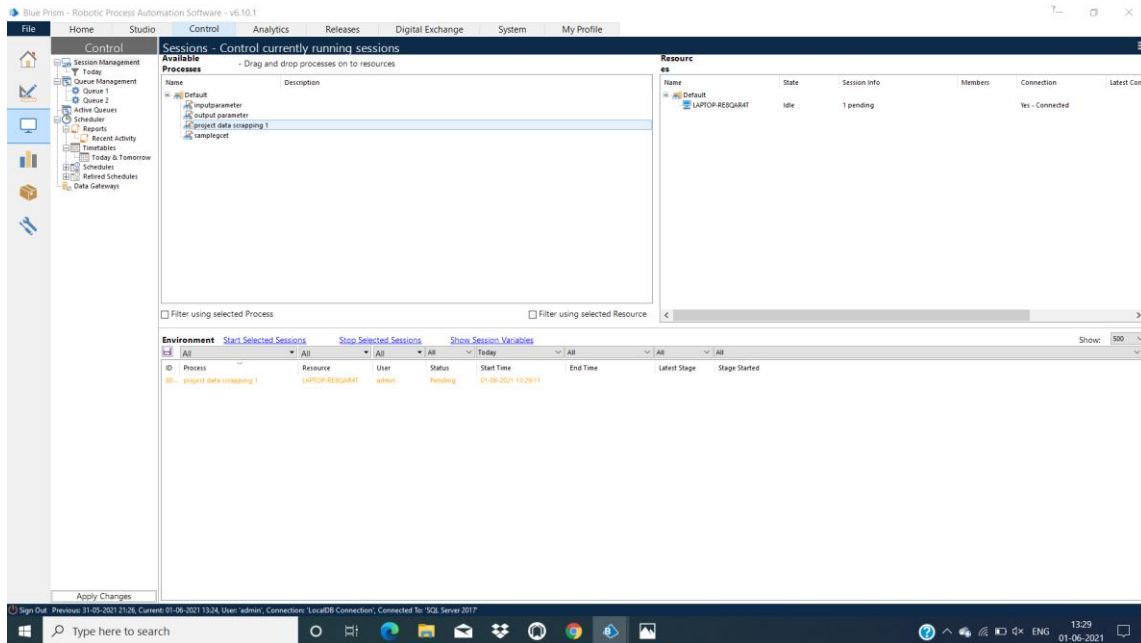




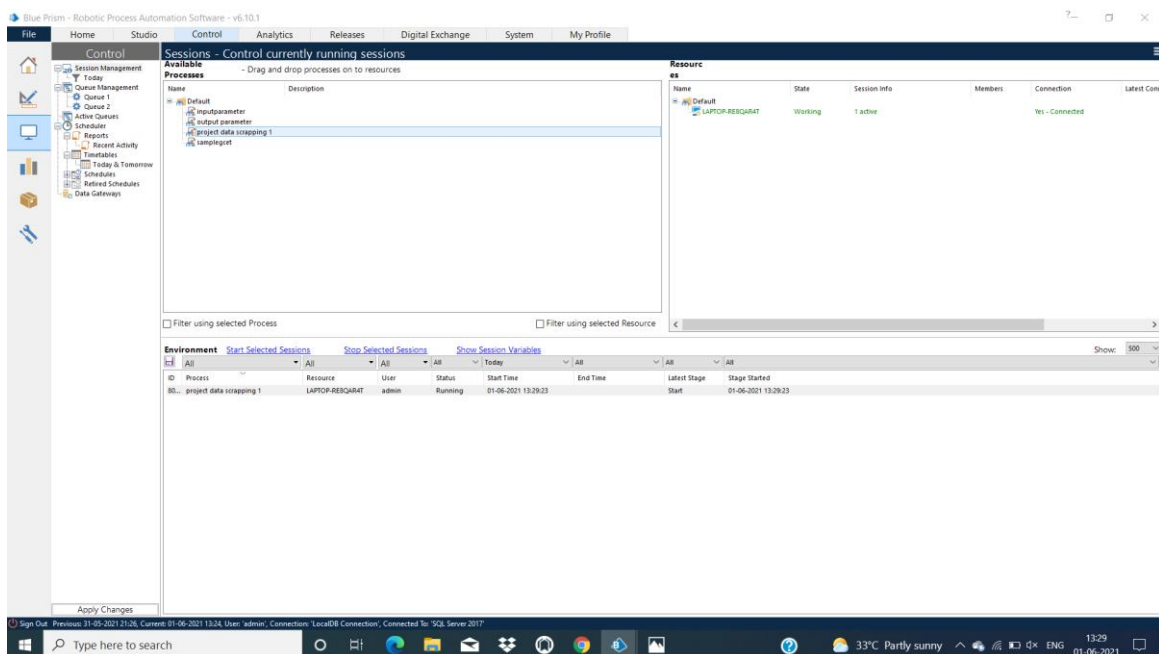
## 6. RESULT

Below are the output of the flows :

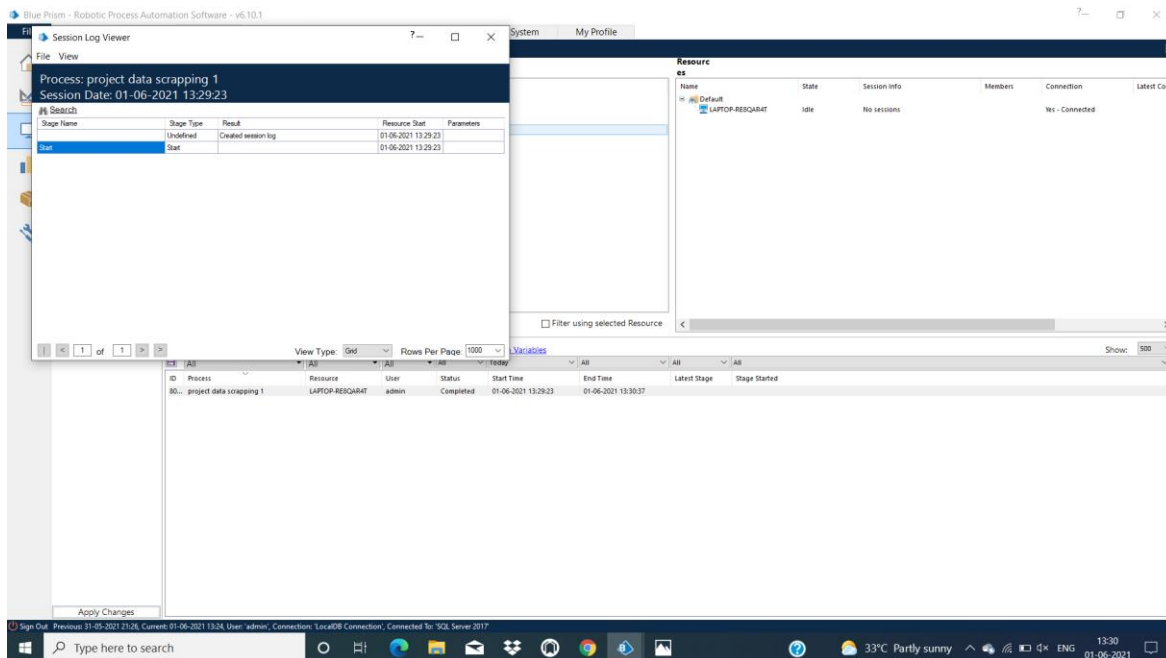
View of the Control Room when the flow is in pending (which is about to run) :



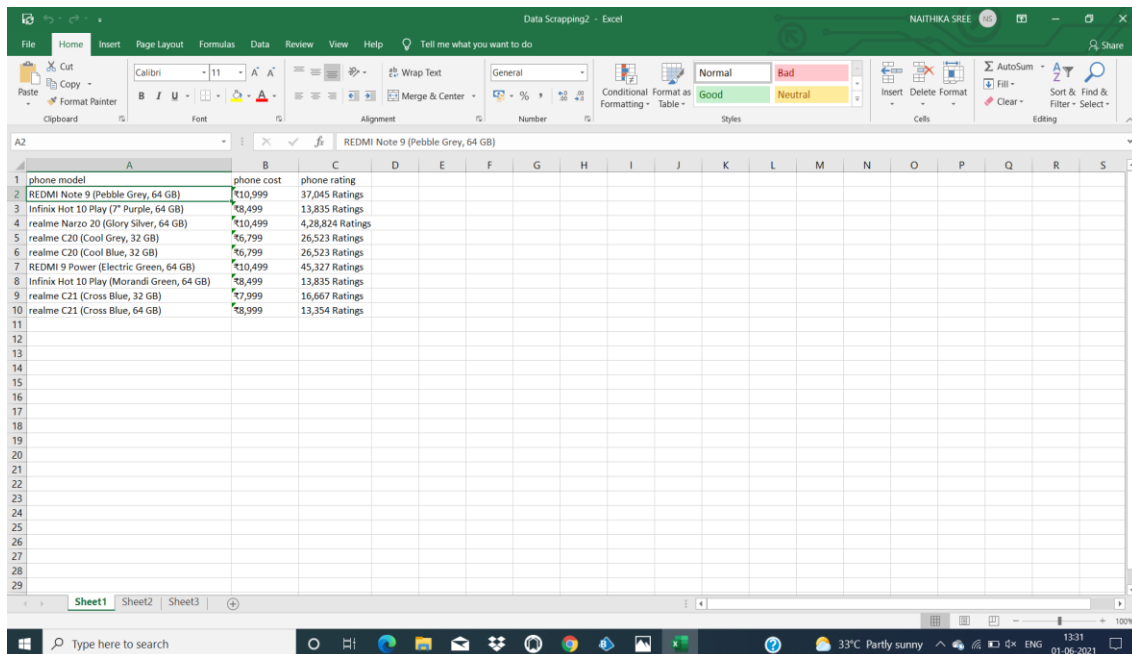
View of the Control room when the flow is running :



Output displaced in the Control room:



Here is the picture of the output in Excel. The information has been extracted from the Flipkart website:



# 7. ADVANTAGES AND DISADVANTAGES

## ADVANTAGES :

- Scrape data from two different websites and compare their similarities. Especially useful for eCommerce stores where you can compare pricing and other data points.
- Create a consistent influx of online reviews as left by your customers across a wide range of review platforms.
- Analyze your competition by scraping SEO and Social Media data on a daily basis.
- Generate new leads by acquiring customer information and their contact data using nothing but crawlers.
- Create large sheets of product details and otherwise from eCommerce platforms like Etsy, Amazon, eBay, and others.
- Strengthen your marketing strategy by gathering large data sets of user feedback on any given subject.

## DISADVANTAGES :

- For anybody who is not an expert, the scraping processes are confusing to understand. Although this is not a major problem, but some errors could be fixed faster if it was easier to understand for more software developers.
- The data that has been extracted will first need to be treated so that they can be easily understood.
- It is common for new data extraction applications to take some time in the beginning as the software often has a learning curve. Sometimes web scraping services take time to become familiar with the core application and need to adjust to the scraping language.
- Most web scraping services are slower than API calls and another problem is the websites that do not allow screen scraping. In such cases web scraping services are rendered useless.

## 8. APPLICATIONS

The Applications of the Data Scraping are :

- Scrape Property details and agent contact details from real estate websites
- Scrape contact details of businesses as well as individuals from yellow pages websites.
- Scrape people profiles from social networks like Facebook, LinkedIn etc. for tracking online reputation.
- Scrape hospital/clinic websites to build a catalog of health physicians including their contact details
- Scrape reviews of products and places/hotels/restaurants.
- Crawl forums and communities to extract data from posts and authors.
- Scrape articles from various article/PR websites to feature in own website.
- Scrape data from various Government websites, most of which do not provide an easy way to download the data which they display.

## 9. CONCLUSION

As the Web has developed cosmically and organizations have gotten progressively reliant upon data, it is currently an impulse to approach the most recent data on each given subject. Data has become the premise of all dynamic cycles whether it's a business or a non-benefit association. Consequently, data scraping has discovered its applications in each attempt of note in contemporary occasions. It is likewise turning out to be progressively certain that the individuals who will make inventive and progressed utilization of web scraping instrument will race in front of others and gain an upper hand. The primary objective of this venture is utilizing blue prism to assemble data from the web and show it in Excel spreadsheet. We had the option to achieve this objective by utilizing data from Flipkart site page customized to our necessities. So influence data scraping and lift your possibilities in your picked space of attempt!

## 10. FUTURE SCOPE

With the expansion of increasingly more data in the realm of the web, the significance of web scraping is expanding. Numerous organizations are currently offering redone web scraping instruments to their customers in which they assemble data from everywhere the universe of the web and orchestrate them into helpful and effectively reasonable data. It decreases the valuable labor to physically visit every website and gather the data. Web Scrapers are planned and code for each and singular website and crawlers do expansive scraping. The Future of web scraping is to be sure brilliant and it will turn out to be increasingly more fundamental for each business with the progression of time.

Web scraping administrations are considered as perhaps the most polished exercises done by a large portion of the IT organizations and Ecommerce Stores that work across the globe. A typical inquiry that is regularly posed is the reason an organization, business or eCommerce store needs to separate the data from the web. The basic answer is that the Internet is the biggest wellspring of data on the planet and contains data in each field of life. The article clarifies the total capability of web scraping in driving business development. One of the greatest and fruitful suppliers of web scraping administrations is Information Transformation Services (ITS) working across the Globe. ITS has customers everywhere on the world and arrangements in each field and specialty. We have particular programming and methods that assist organizations with getting fast outcomes from the web. ITS offers 99% proficient and precise data. Our Services speed up and decrease the time and assets of the organization for get-together data from the Internet.

# 11. BIBLIOGRAPHY

Below are the References listed:

1. *"Search Engine History.com". Search Engine History. Retrieved November 26, 2019.*
2. ^ Song, Ruihua; Microsoft Research (Sep 14, 2007). *"Joint Optimization of Wrapper Generation and Template Detection" (PDF). The 13th International Conference on Knowledge Discovery and Data Mining: 894. doi:10.1145/1281192.1281287. ISBN 9781595936097. S2CID 833565. Archived from the original (PDF) on October 11, 2016.*
3. ^ Semantic annotation based web scraping
4. ^ Roush, Wade (2012-07-25). *"Diffbot Is Using Computer Vision to Reinvent the Semantic Web". www.xconomy.com. Retrieved 2013-03-15.*
5. ^ *"FAQ about linking – Are website terms of use binding contracts?". www.chillingeffects.org. 2007-08-20. Archived from the original on 2002-03-08. Retrieved 2007-08-20.*

[https://en.wikipedia.org/wiki/Web\\_scraping](https://en.wikipedia.org/wiki/Web_scraping)

<https://it-s.com/the-future-of-web-scraping-services/>

<https://www.webharvy.com/articles/web-scraper-use-cases.html>

<https://librarycarpentry.org/lc-webscraping/05-conclusion/index.html>

<https://www.promptcloud.com/blog/emerging-applications-of-web-scraping-technology/>

<https://www.xbyte.io/the-history-of-web-scraping.php>

<https://towardsdatascience.com/https-medium-com-hiren787-patel-web-scraping-applications-a6f370d316f4>