Title: Data Scraping from flipkart using Blueprism

1. Introduction to Scraping

1.1 Overview

Web scraping is the process of collecting structured web data in an automated fashion. It's also called web data extraction. It can be defined as the process of using bots to extract content and data from a website. Unlike screen scraping, which only copies pixels displayed onscreen, web scraping extracts underlying HTML code and, with it, data stored in a database. The scraper can then replicate entire website content elsewhere. Some of the main use cases of web scraping include price monitoring, price intelligence, news monitoring, lead generation and market research among many others. It's extremely simple, in truth, and works by way of two parts: a web crawler and a web scraper. The web crawler is the horse, and the scraper is the chariot. The crawler leads the scraper, as if by hand, through the internet, where it extracts the data requested.

1. The crawler

A web crawler, which we generally call a "spider," is an artificial intelligence that browses the internet to index and searches for content by following links and exploring, like a person with too much time on their hands. In many projects, you first "crawl" the web or one specific website to discover URLs which then you pass on to your scraper.

2. The scraper

A web scraper is a specialized tool designed to accurately and quickly extract data from a web page. Web scrapers vary widely in design and complexity, depending on the project. An important part of every scraper is the data locators (or selectors) that are used to find the data that you want to extract from the HTML file - usually, XPath, CSS selectors, regex, or a combination of them is applied.

Automated web scrapers work in a rather simple but also complex way. After all, websites are built for humans to understand, not machines. First, the web scraper will be given one or more URLs to load before scraping. The scraper then loads the entire HTML code for the page in question. More advanced scrapers will render the entire website, including CSS and Javascript elements. Then the scraper will either extract all the data on the page or specific data selected by the user before the project is run. Ideally, the user will go through the process of selecting the specific data they want from the page. For example, you might want to scrape an Amazon product page for prices and models but are not necessarily interested in product reviews.

Lastly, the web scraper will output all the data that has been collected into a format that is more

useful to the user. Most web scrapers will output data to a CSV or Excel spreadsheet, while more advanced scrapers will support other formats such as JSON which can be used for an API.

1.2 Purpose

Some of the main use cases of web scraping include price monitoring, price intelligence, news monitoring, lead generation, and market research among many others. In general, web data extraction is used by people and businesses who want to make use of the vast amount of publicly available web data to make smarter decisions. If you've ever copy and pasted information from a website, you've performed the same function as any web scraper, only on a microscopic, manual scale. Unlike the mundane, mind-numbing process of manually extracting data, web scraping uses intelligent automation to retrieve hundreds, millions, or even billions of data points from the internet's seemingly endless frontier.

1.3 About RPA and Blueprism

Robotic Process Automation(RPA) is a methodology in which robots perform a set of tasks by following a process, without any human intervention. All of these technologies reduce the manual workforce, by giving them time and harnessing their talents in better ways. Robotic process automation does not include actual robots automating your tasks. But, it is the RPA Tools/Vendors providing you a set of libraries and runtime environments for automating business processes.

RPA Tools

There are a various number of tools in the RPA market. But, the top trending tools in today's market are UiPath, Blue Prism & Automation Anywhere.

What is RPA Blue Prism?

Blue Prism is an RPA Tool which holds the capability of virtual workforce powered by software robots. This helps the enterprises to automate the business operations in an agile and cost-effective manner. The tool is based on Java Programming Language and offers a visual designer with drag and drop functionalities. Formed in 2001, this tool differs from the other tools in the market, by using a Top-Down approach. Also, Blue Prism offers a visual designer with no recorders, scripts or any intervention.

Features of Blue Prism

The features of RPA Blue Prism are as follows:

- Secure & Accurate – There is no limit to the number of processes which need to be executed in this tool. Blue Prism delivers secure and accurate results to any number of processes that you wish to automate.
- Robust – It provides robust features like load balancing, data encryption, and end-to-end

auditing. So, every change is audited and related back to the user related to it.

- Scalable & Resilient – It allows scalability with central management. So, all the processes can be automated as per the need and can be monitored centrally.
- 24*7 Workforce – The tool is designed to work intelligently without a person physically monitoring every action occurring on the screen.
- Analytics – Provides extended features to configure Dashboards so that the session information can be redirected to the Monitoring Systems.
- Data Security and Data Abstraction – As the tool is designed to work autonomously, all the processing is performed and is stored in a data center. This provides a well-defined data abstraction of data and process security.
- Provide Cloud support – Provides working capacity support as per the business requirement. So, users have to just create workers on demand and manage them centrally.
- Execution Intelligence – Robots connect to systems and react dynamically to the responses in the data on multiple environments.

2 Proposed Solution

The project aims to build an application to extract iphone type and price from the flipkart by auto launching the website. This project performs the following:

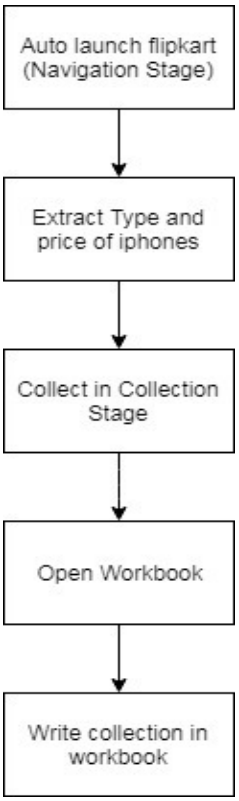
- ➤ Auto Launch the flipkart iphone page
- ➤ Extract the Data -Type and price
- ➤ Store the Data Extracted in excel named iphones.xlsx

Tool used:

- ➤ Blue Prism

3. Theoretical Analysis

3.1 Block Diagram

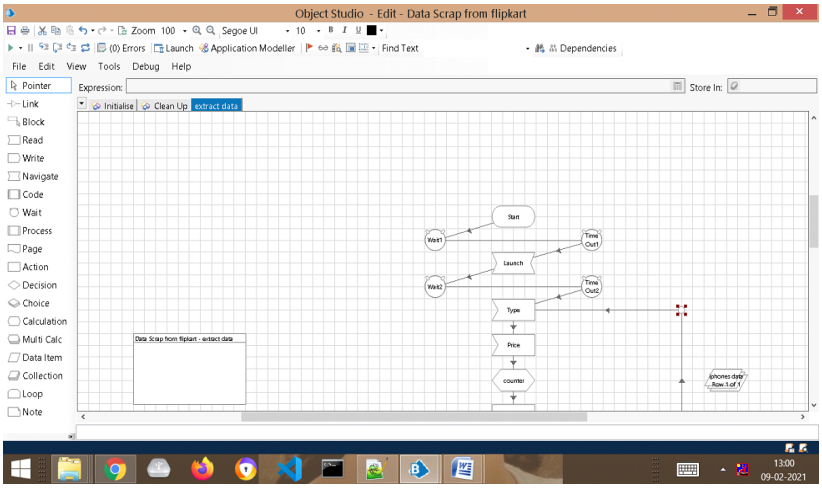


3.2 Hardware/Software Requirement

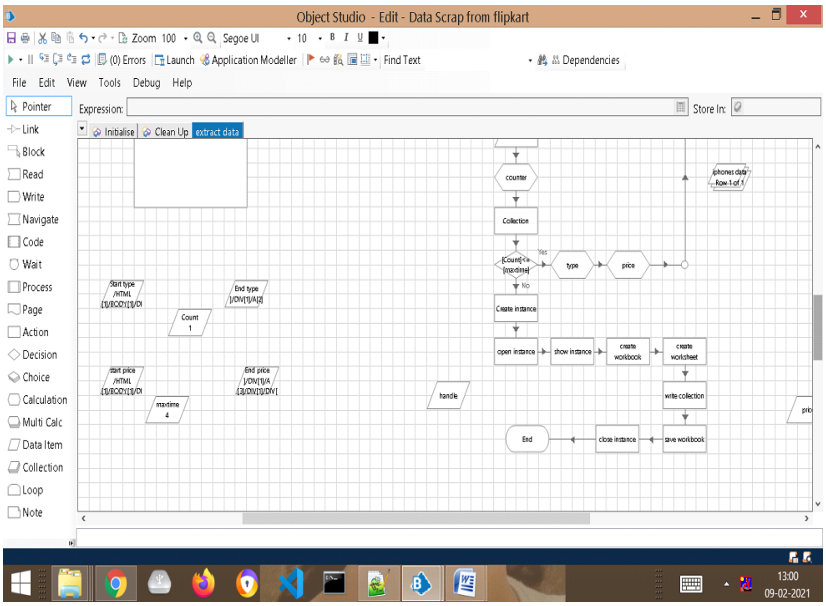
- OS: Windows 8 or above
- Tool : Blueprism

4. Screenshots

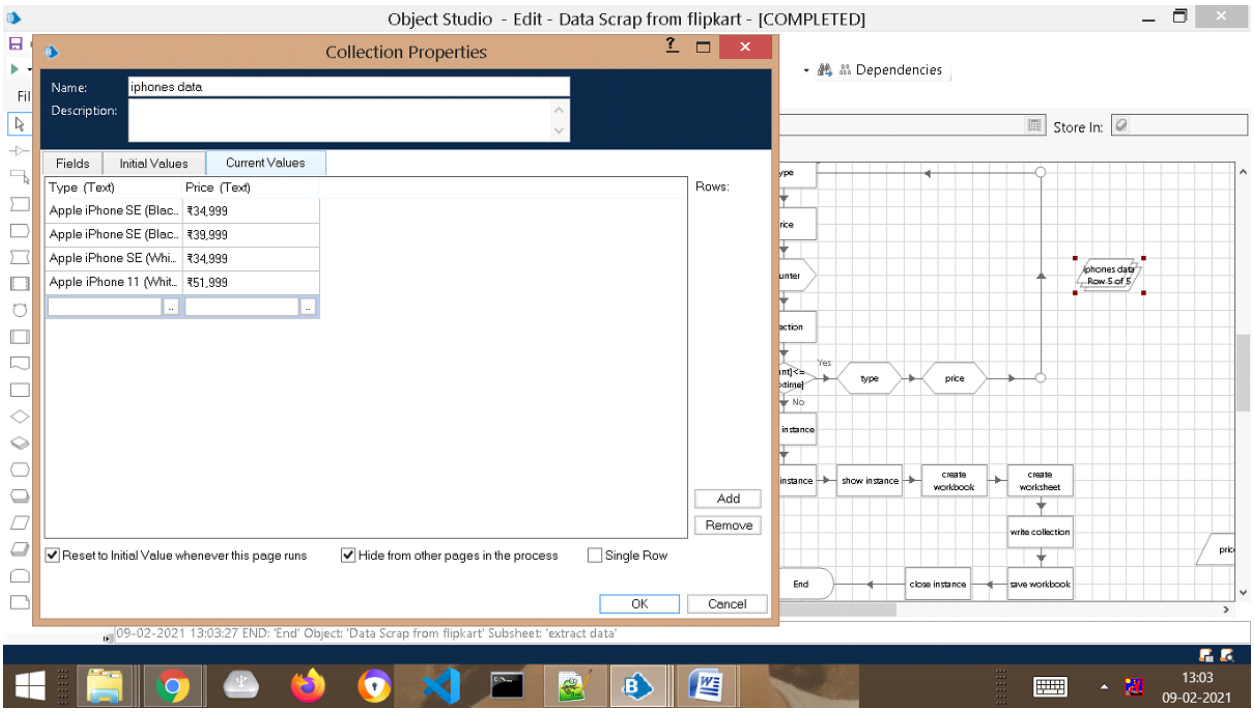
Business Object Flow:



Business Object Flow:



Data Extracted in collection stage:



5. Applications

Web Scraping can be used for

- Price intelligence
- Market research
- Alternative data for finance
- Real estate
- News & content monitoring
- Lead generation
- Brand monitoring
- Business automation
- MAP monitoring
- Real-Time Analytics
- Predictive Analysis
- Natural Language Processing
- Machine Learning Training Models

6. Conclusion

As the Internet has grown astronomically and businesses have become increasingly dependent on data, it is now a compulsion to have access to the latest data on every given subject. Data has become the basis of all decision-making processes whether it's a business or a non-profit organization. Therefore, web scraping has found its applications in every endeavour of note in contemporary times. It is also becoming increasingly clear that those who will make creative and advanced use of web scraping tool will race ahead of others and gain a competitive advantage.

In the project, Blueprism tool is used to extract data from flipkart website. Type of the iphone and price are extracted and written in a collection stage. Then the extracted collection is written in an excel workbook for further purpose.

7. Bibliography

www.google.com
www.portal.blueprism.com
www.imperva.com
www.towardsdatascience.com