Caption

PROJECT:
DATASCRAPING USING BLUEPRISM TOOL

Submitted by: Submitted to:
M.Sai Sravani   G. Hemanth kumar
CBIT HYD.

1.overview

Data scraping, also known as web scraping, is the process of importing information from a website into a spreadsheet or local file saved on your computer. It's one of the most efficient ways to get data from the web, and in some cases to channel that data to another website. Popular uses of data scraping include:

- Research for web content/business intelligence

- Pricing for travel booker sites/price comparison sites

- Finding sales leads/conducting market research by crawling public data sources (e.g. Yell and Twitter)

- Sending product data from an e-commerce site to another online vendor (e.g. Google Shopping)

1.purpose

Data-entry, although sounds outdated, is still a job in many companies. Such personnel can instead be converted to RPA developers and be allowed to take advantage of the technology. Also, companies still scraping data manually should definitely be using Web scraping with Robotic Process Automation, to increase efficiency and reduce errors and costs. RPA is not bound by any specific use cases and in reality, you could make RPA perform any task that you think is too boring and repetitive to be done by a human being. Web data extraction happens to be one of major areas where RPA has immense potential in the digital world.

robots utilize the user interface to capture data and manipulate applications just like humans do. They interpret, trigger responses, and communicate with other systems to perform a vast variety of repetitive tasks. Only substantially better: an RPA software robot never sleeps and makes zero mistakes.It is always best to choose tasks that are repetitive, rules-based, and do not require human judgment. The non-standard processes are difficult to automate, and human interaction is required to complete such processes. So, there are limited tasks that you can automate with RPA.

Type to enter text

2.1 Existing problem

Existing approaches or method to solve this problem.

It is always best to choose tasks that are repetitive, rules-based, and do not require human judgment. The non-standard processes are difficult to automate, and human interaction is required to complete such processes. So, there are limited tasks that you can automate with RPA.

After completion of entire process I just run the program so I got stopped in the reader (which I named as Samsung model)

I faced the problem is internal program is failed because the to match the items are same in the page. Again I got trouble in Samsung model (reader) which I should ext
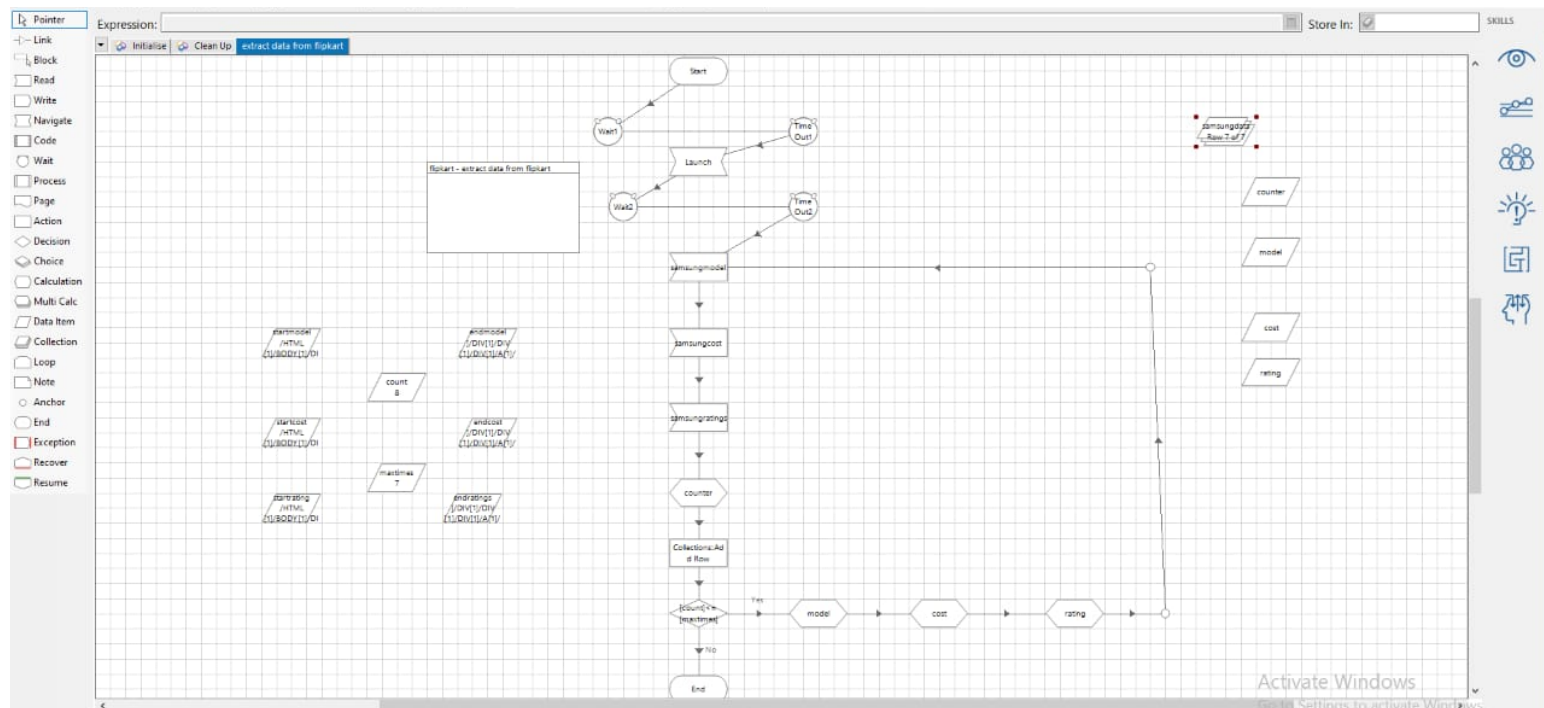
2.2 proposed solution

What is the method or solution by you?

The solution is we should not open the tabs of launched websites more than one because the matching items that we should identify will be more than one because the multiple websites which we have opened at the same time.

3. Theoretical analysis

3.1 block diagram

Caption

3.2 Hardware / Software designing

Software requirements of the project.

Blue Prism

Blue Prism is an RPA Tool that provides a virtual workforce to the organizations. It helps organizations to automate manual, repetitive, and rule-based business processes in an agile and cost-effective manner. It includes drag and drop support to automate the tasks.

Features

- It is platform-independent so that it can be used on any platform.

- It contains robust features like load balancing, data encryption, and end-to-end auditing. Thus, every change is audited.

- Blue Prism also supports automation of codes written in Mainframe, Java, Windows applications, and even web-based applications.

- It supports all major cloud platforms, like Microsoft Azure and Amazon AWS. So, users can manage most of the tasks centrally.
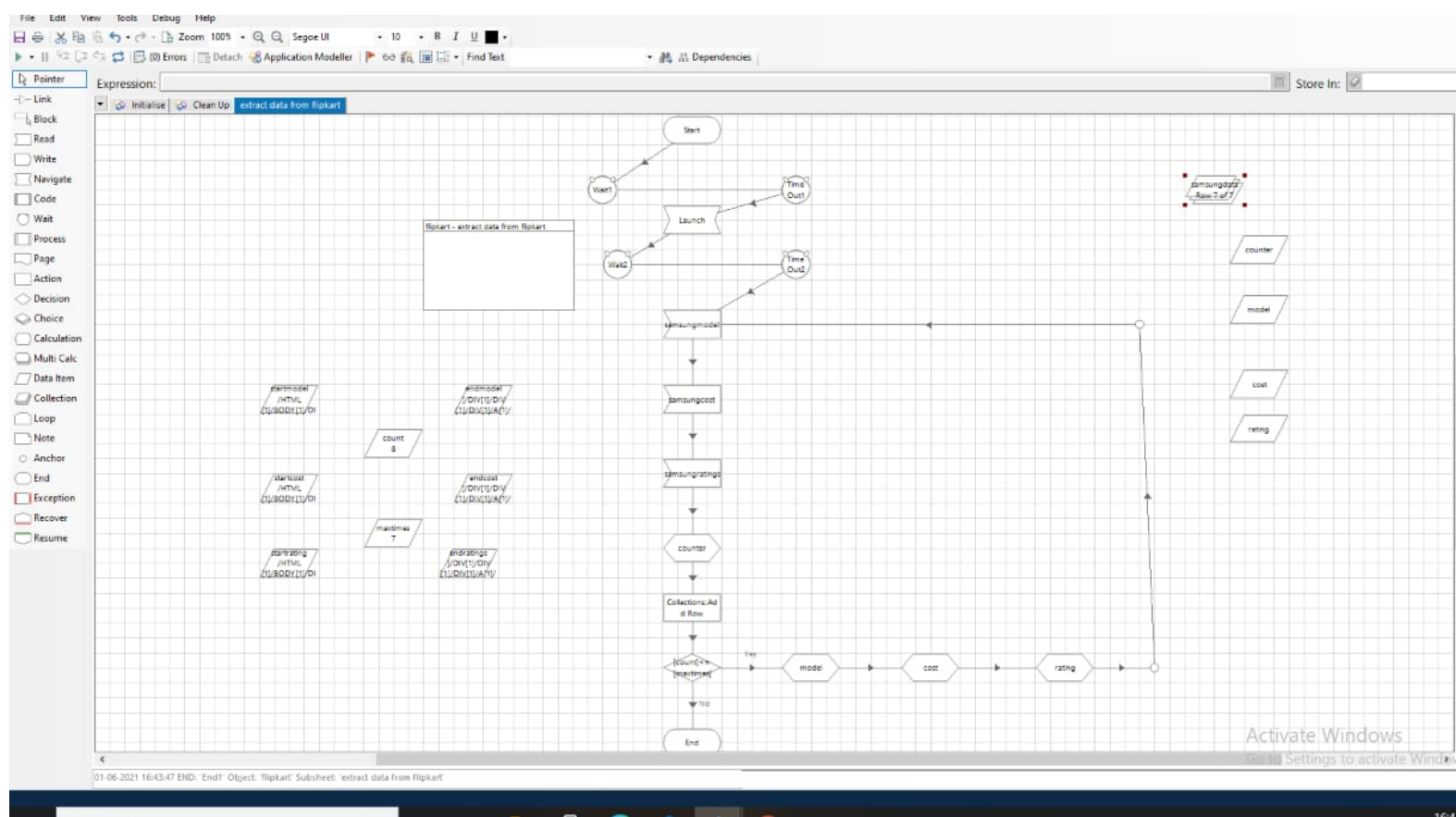
4.Experimental investigations

Analysis or the investigation made while working on the solution.

I analyze about blue prism tools and investigated the process of datascraping .

5.Flowchart

Diagram showing the control flow of the solution

Caption

6.RESULT

Caption

7.Advantages

- Eliminates the need for manual **data** entry.
- Reduces costly human errors.
- No need for custom processing and filtering algorithms.
- Easy drag-and-drop features eliminate the need to write scripts.
- ## Automation leading to lesser errors and costs

RPA isn't used for programming, rather its main use is to take over repetitive and boring tasks where humans are bound to make more errors. RPA helps make sure that lesser human errors occur, and also drives down costs. The previous Web Scraping with Robotic Process Automation was a classic example in which, a human had a high chance of making an error and would definitely require more time.

- ## Customised Scraping, not just Screen Grabbing

Contrary to what many believe, RPA is not Screen Grabbing. What it instead does, is that it creates a script file, after seeing your actions, then it runs the same script file to mimic your actions and get similar work done. In the end, its work is generating a script file that will take care of repetitive work, and not screen-grabbing.

- ## Do more than just crawl

You can do more than just basic Web Scraping with Robotic Process Automation. You could even teach it how to clean data by performing certain steps manually first and then letting it take over. RPA tools are also used for automating quality and assurance testing, sending the same emails to multiple people, and hundreds of other similar tasks that are repetitive and boring.

- ## No need to know to code

Every person needs automation, whether he is a business associate or a research associate or a full-blown coder. But

everyone is not comfortable coding, let alone automating tasks. Moreover automating one task might be very different from automating another. That is why RPAs are a boon to companies since they can be used by anybody irrespective of their coding knowledge.

- ### Faster to set up

RPAs are much easier to set up than teaching people how to code. Or even teaching any specific subject for that matter. Most automation techniques are already well established and applying them to a new set of web scraping requirements isn't very time consuming. What one basically has to do is, know how to make the system record a series of steps and how to ask it to do the same steps on other values.

- ### No need of a web scraping team

Web scraping with Robotic Process Automation is one of the growing uses of RPA. An entire web scraping team can be replaced by a single RPA expert who can train the scraping system to collect data from various types of webpages. And the best part is, it can be customised for any complex use cases of web scraping.

- ### Gathering Social Media data

Social media data collection has become more and more important as companies try to analyse their customers and understand their mindset. The problem is that most social media websites do not allow you to gather data automatically and doing so in a manual fashion would require too much time, too many personnel and would be prone to errors. Using an RPA based web scraping solution to perform this task would be much simpler and a favoured choice.

- ### Automating batch download tasks

Downloading images, videos, or even text documents in batches is something we have all done and been bored of doing. I even remember sitting and searching for songs of an artist on Google and downloading almost 50 of his songs one by one. Such batch download tasks required by companies can be given to RPA solutions, which can easily build a script that would download all your requirements.

Disadvantages

- **Difficult to analyze –** 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.
- **Data analysis –** The data that has been extracted will first need to be treated so that they can be easily understood. In certain cases, this might take a long time and a lot of energy to complete.
- **Time –** 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 scrapping language. This means that such services can take some days before they are up and running at full speed.

8.Applications

1.Product intelligence and demand analysis
2.Price intelligence
3.SEO Monitoring
4.Designing targeted ad campaigns
5.Campaign monitoring
6.MAP compliance
7.Lead generation
8.Content marketing
9.Market research
10.Brand monitoring

9.Conclusion

Data entry seems simple enough—go to a website, copy and paste the data into a spreadsheet, and then crunch the numbers. But manually gathering and entering data can add up over time. And unfortunately, humans are prone to errors, leaving you with potentially unreliable information. Luckily, there is a way to reclaim your time and eliminate errors—with robotic process automation from Automate by using Blueprism or UIpath or automation anywhere tools.

10.Future scope

 The solution is not user-friendly. It has a very high learning curve. People should be able to learn it easily so that they will get interested in using it.

While the solution is more secure, it's very hard to find people trained on that. I need different people, not only those who are trained on RPA tools. I cannot get people in Blue Prism. Without the resources, people just move to Automation Anywhere or UiPath which are more user-friendly. In comparison, UiPath is much easier to use and you can find people who are well-versed in it.

The product needs to put out more videos, similar to Automation Anywhere, which does that a lot. You can find a lot of videos online in relation to Automation Anywhere and UiPath, however, this is not the case with Blue Prism. There's just less information available.

Blue Prism needs to provide better training. They need to start something similar to Automation Anywhere University or UiPath Academy. If they had some courses at different levels (basic, advanced, and master), there would be more educated personnel available.

The solution needs to provide a trial license - whether it is on cloud or on-premises. They need to provide a standard environment to work with. We need to have practice in installing the data center and connecting it to the database.

We need to understand how we can migrate from one lesser version to the higher version and what is the load balance and how we can manage that. We need to understand better how our core system is managing that. Proper training would help with understanding.

11. BIBILOGRAPHY

Reference:

https://youtu.be/iVYtNdCYgFo

APPENDIX

A.Source code

https://drive.google.com/file/d/17zileKcpe7lZ6zW8LeEyOgRicP_jzhEc/view?usp=drivesdk