AUTOMATE DATA COLLECTION OF THE EXAMINATION RESULTS IN AFFILIATED COLLEGESUSING WEB SCRAPING TECHNIQUES

Mr. Raman RK, Asst. Professor, Dept. of Computer Science and Engineering
Mr. K. Srinivas Rao, Asst. Professor, Dept. of Computer Science and Cyber Security
Loyola Academy, Alwal, Hyderabad
[1]raman.rk@loyolaacademy.edu.in [2]srinivasrao.kunchala@loyolaacademy.edu.in

Abstract - Since the evolution of the Internet, the technology has grown leaps and bounds to process the data. Once upon a time gathering data was a huge task but now the scenario has changed and it is challenging to process the huge data. In order to resolve this challenge, numerous data analysis techniques are being undertaken. One of the main data analysis approaches is data scraping or web scraping. Using web scraping, the unstructured data from the website is transformed into a proper structured data. The traditional techniques involve copying and pasting the results of each student in Excel sheet one by one which is quite time consuming. Using Robotic Process Automation(RPA) we can try to resolve the issue. This paper mainly focuses to provide a solution for the challenge that is imposed when coming to data gathering and analyzing the examination results of students in affiliated colleges.

Keywords: *RPA, Excel, Automation, DataAnalysis, Web Scraping*

I. INTRODUCTION

The data plays a major role in any kind of research, either it can be academic or business, research etc. The users are prone to huge data around them and they might want to analyze the data from various websites. With the advent of the Internet, most of the industries, institutions etc. have started implementing their websites. All these websites contain various information that belong to diverse categories. Each data is presented in different formats and the users will have to work a lot on gathering data at one place as it is spanned across multiple pages under various sections. Even most of the websites are very secure and do not allow the users to copy the data. In those situations it will be very difficult to gather the data. If the website does not have any restrictions like that then the user will have to copy and paste the data manually every time which is quite a tedious job. In order to resolve this drawback, a technique called

Web Scraping is used. Using Web Scraping, we can extract or scrap the data from multiple web pages into a spreadsheet or a database. When coming to the field of education still certain aspects are done manually, especially examination results of the students where the data and marks of every student is copied and pasted in excel sheet. The main motive behind this study is to offer a solution for the Affiliated colleges to automate the data collection of their examination results. Section II describes the basic overview of web scraping. Section III describes the problems in existing systems performed with Web Scraping. Section IV represents the proposed system.

II. OVERVIEW OF WEB SCRAPING

As the data grew more and more every day, there was a huge requirement for a technology to extract data from multiple sites and put them in one place. To achieve this feat, Web scraping, also known as web extraction, is a technique to extract data from the Internet (WWW) and save it to a spreadsheet or database. (Draxl, 2018). The main feature of Web Scraping is to extract the data which is required based on the current scenario and it can be even customized based on the users. Web scraping can be used in different scenarios (i.e.) to monitor or compare the products, prices, weather monitoring, business development (Zhao, n.d.). Web Scraping is normally done using Robotic Process Automation (RPA) by the creation of bots for any operation. There are many tools that can be used in performing Web Scraping such as Beautiful Soup, Scrapy, Selenium etc. The software used in this study is Power Automate Desktop which provides many options for the data collection from different sources. The data is extracted from the website in HTML format using the

web scraper and can be rendered in various formats such as database, spreadsheet etc. as shown in Figure 2.1(Irfan Ali Kandhro, 2022).

Figure 2.1 Working of Web Scraping

III. PROBLEMS IN EXISTINGSYSTEM

As we have already discussed, most of the Affiliated Colleges across India do not have the proper efficient mechanism to collect the results of each student when the results are released by the University. Normally, for the affiliated colleges during the release of the semester results they rely on a few third party websites such as *Schools9, Manabaddi etc.* apart from the regular *Osmania University* website in the state of *Telangana*. During the time of results the faculty are facing a huge difficulty as they need to enter the hall ticket number of each student every time to extract the data and then manually copy the data such as hallticket no, name and grades of each subject along with the SGPA score from the websites mentioned above. Once they copied it, they needed to paste it in the excelsheet again manually which is quite timeconsuming and inefficient. (K. Himaja et al.,2017)

IV. PROPOSED SYSTEM

In order to break this monotony mentioned in the previous section, the examination results from the websites can be extracted automatically by using the Robotic Process Automation (RPA) technique. In RPA we have many different softwares to automate but in this study we use *Microsoft Power Automate*. Using the Power Automate, new flow can be created for streamlining the repetitive tasks done during the result analysis. An Excel Spreadsheet is provided as an input containing the Hall Ticket numbers and the name of the student with necessary subjects as the column headings. Since the data present in the website is not according to the format which is required by the Institution, the bot is designed to extract the data of each student and kept in a new spreadsheet.(Shreya V. Dhoke et al., 2022) Later, using that spreadsheet as the input file the bot processes the records for each subject by checking the name of the subject. If the condition is satisfied then the bot will write the grades of the subjects automatically for each student since it is being given under the loop. Once the bot completes processing all the records from the input file the data will be extracted and placed it according to the format which is required as shown in the flowchart figure 4.1

Figure 4.1 Flowchart of Bot Execution inthe Proposed System

The following are the screenshots of the proposed system which was developed using Power Automate

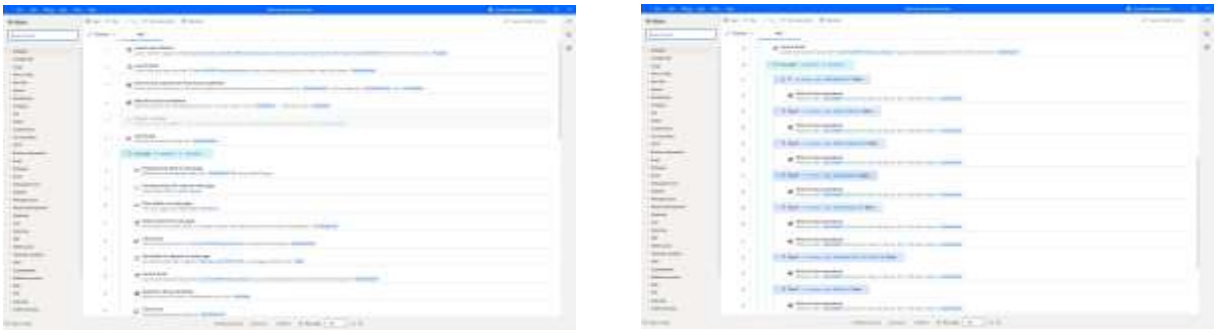

Figure 4.2 Bot Development Flow of theProposed System

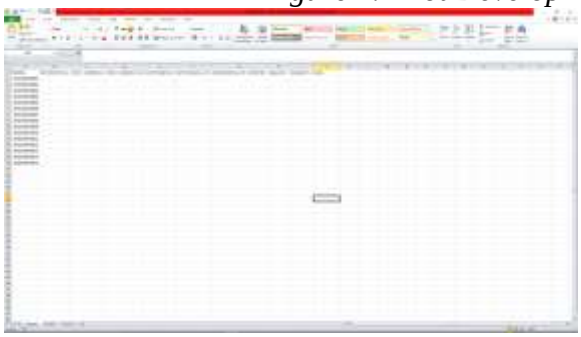

Figure 4.3 Input as Excel Spreadsheet

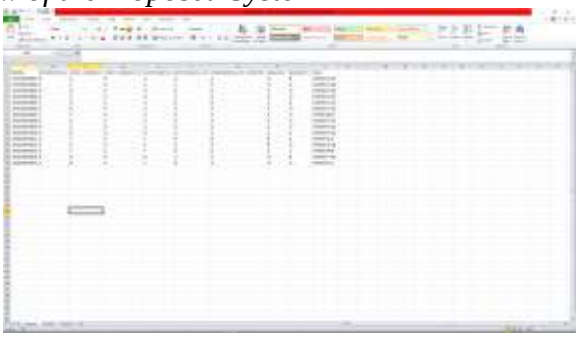

Figure 4.4 Output as Excel Spreadsheet

V. FUTURE RECOMMENDATION

This study is implemented using only excel as the input file but this bot can be improved by taking the input from the database using Excel OLEDB. The main advantage of this improvement is that the users can extract the results based on their requirements by writing customized queries which can extract data from the database. The other advantage is the efficiency of the system will also be improved.

VI. CONCLUSION

By analyzing the drawbacks and difficulties faced during the data collection of results from the third part websites. This study has proposed a solution to automate the tasks to reduce the manual intervention to extract the results into an Excel Spreadsheet. This study even suggests that the Affiliated Colleges across India implement this strategy of extracting the results through Automation which saves lots of time and there will be nodata mismatch.

REFERENCES

[1] Draxl, V. (2018, January 1). *Bachelor paper web scraping data extraction from websites*. Academia.edu. RetrievedNovember 6, 2022, from https://www.academia.edu/35901535/BACHELOR_PAPER_Web_Scraping_Data_Extraction_from_websites

[2] Zhao, B. O. (n.d.). *Web scraping*.SpringerLink. https://link.springer.com/referenceworkentry/

[3] Qingli Niu, & Irfan Ali Kandhro. (2022, May 8). *Web Scraping Tool For Newspapers And Images Data Using Jsonify*. Open Access Journal - Journal of Applied Science and Engineering. https://jase.tku.edu.tw/articles/jase-202304 -26-4-0002.pdf

[4] K. Himaja, lalam Sreevidya, & N.Srinivasan. (2017, July 13). *RESULTANALYSIS AUTOMATION*. ARPNJournals. https://www.arpnjournals.org/jeas/research_papers/rp_2017/ jeas_0717_6170.pdf

[5] Shreya V. Dhoke, Anupama D. Sakhare, & Satish J. Sharma. (2022, June). *EFFICIENT SCRAPING OF DATA FROM WEBSITES USING SELENIUM*. Journal JETIR follow UGC CARE List, UGC-CARE journal, ugc care, ugc care approved list, ugc approved journal, ugc care Journal, UGC-CARE, UGC Approved UGC CARE, UGC-CARE, UGC Approved List of Research Journal, ugc approved journal, research journal. https://www.jetir.org/papers/JETIRFM06063.pdf

[6] Manjushree B, & Sharvani G. (2020, June). Survey on Web scraping technology. Home. https://www.wthtjsjs.cn/gallery/1-whjj-june-5412.pdf