November, 2022

A project report on

JOB PORTAL SEARCH AUTOMATION

Submitted in partial fulfillment for the J Component project of

CSE3119 – Robotic Process Automation

by

RANJANI T 20BRS1009

Signature of the Candidate

Signature of the Faculty

Sakthivel V

Dr.V.Sakthivel, AP(Sr.G)/SCOPE

BONAFIDE CERTIFICATE

Certified that the project report entitled **"JOB PORTAL SEARCH AUTOMATION"** is a bonafide work of **KARTHAVYA DEVANATHAN- 20BRS1012, RANJANI T-20BRS1009 and LIKITH SAI -20BRS1265** who carried out the Project work under my supervision and guidance for CSE2023 – Robotic Process Automation

Dr. V. Sakthivel

Assistant Professor (Senior)

School of Computer Science and Engineering (SCOPE)

VIT University, Chennai

Chennai – 600 127.

TABLE OF CONTENTS

ABSTRACT

From a newly graduated student or an experienced professional in the field, lookingfor a new job can be highly demanding and intimidating. Websites which confer fraudulent job positions or allow risks of personal identity theft must be strictly avoided. With the increasing number of job portals which lists different jobs during different times, it becomesa frustrating experience. It was becoming laborious to continually check and change expectations of the individual to receive best results and new roles which had been posted.

We already have a plethora of data on the search, from LinkedIn, indeed, Glassdoor etc., thathost many opportunities per day. Rushing into interviews with only a few offers leads to settling for lesser offers can induce great struggle in the future. But have you ever noticed that usually, they all contain different jobs? So, how about we scrap the data from these websites to build a collective job search portal? The main idea of the project is to scrap popular job portal websites and obtain information like the date of the job posting, salary details, job industry, company name, etc. Then, the problem statement writes as an automated, combined job search portal which enables newest information on job position meeting the demands of the user. And, to handle this huge amount of information from manydifferent job websites such as Naukri, Glass Door etc., for which we will be using the method of Data Scraping.

CHAPTER 1

INTRODUCTION

RPA is an exciting new technology that is being seen as the first step to using new and advancing technologies to automate enterprise processes. We will learn how RPA brings a level of advancement to automation that is much more evolved than the earlier automation process through its use of scripts, macros, and so on. RPA is advancing rapidly and is part of a bigger movement to low code and no-code tools. Compared to traditional coding these tools are easier to use and faster to deploy Organizations are now using these rapid development tools with emerging technologies such as process mining, artificial intelligence (AI), and analytics to enable end-to-end process automation. In terms of job portals, most of the job seekers face Irrelevant recruiter and candidate connection, Loss of information and update trackers and also Long-time consuming and lag process. So, in order to find the jobs faster and apply for it before the deadline, we are in need of an application where it makes our work easier by automating various job portals and provide us with the data easily so that we can apply for the job in a single click. So, in this project we will be automating job portals like Naukri, glass door and so on.

WORKFLOW ARCHITECTURE
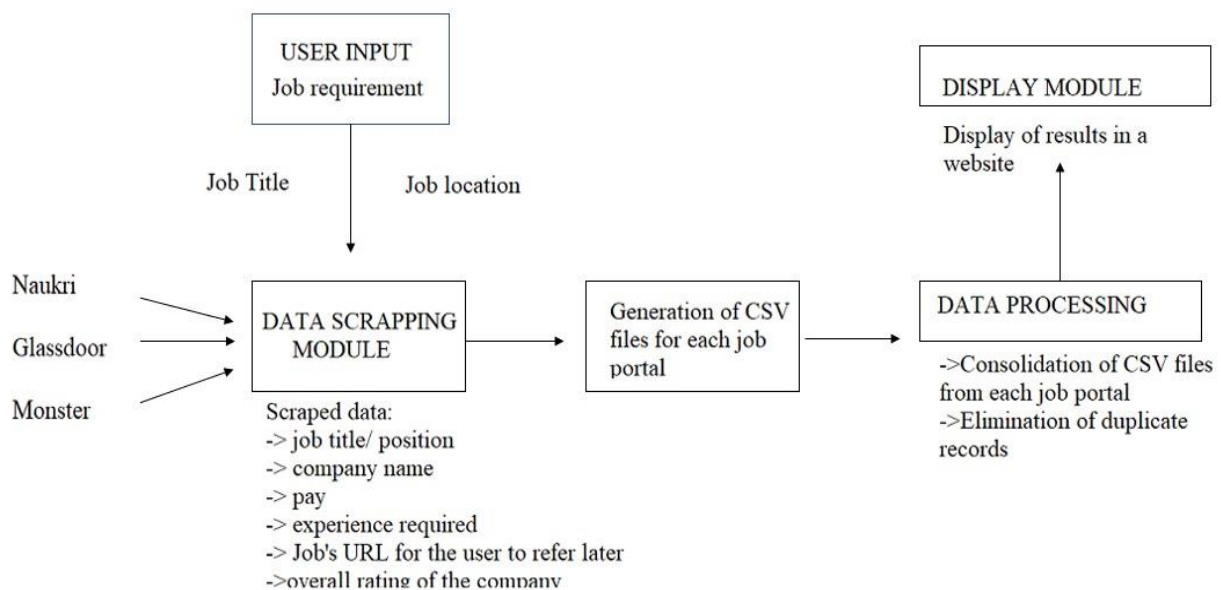
CHAPTER 2

SOFTWARE

UiPath is a robotic process automation tool for large-scale end-to-end automation. For an accelerated business change, it provides solutions for businesses to automate routineoffice activities. It uses a variety of methods to transform tedious tasks into automated processes. UiPath can extract anything you can see in a web browser. This includes statistics, finance and stock info, real-estate data, product catalogues, search-engine results, job listings, social networks feeds, customer opinions, and competitive pricing.

Within a company, you can find even an even larger variety of data formats that UiPath can handle: reports, dashboards, customers, employees, finance, and medical data that you need to transform and migrate. It can automatically log in to a web site, extract data spanning multiple web pages, filter and transform it into the format of your choice, before integratingit into another application or web service.

CHAPTER 3

## MODULES

3.1 DATA SCRAPING MODULE:

In this module, we will collect data from the user based on their requirement. We collect such type of information from the user. This forms the user module. After collecting the data from various websites, we must send the data received to the data scrapping module to proceed further with the user's requirement. Extracting information from the website in Data Table format from separate workflow for NAUKRI, GLASSDOOR and MONSTER. The data from these websites are stored to a file and read from a MAIN workflow where all the data are compared, processed by removing duplicate rows and sent to the website.

Data scraping enables you to extract structured data from your browser, applicationor document to a database, .csv file or even Excel spreadsheet. Structured data is a specific kind of information that is highly organized and is presented in a predictable pattern.

For example, all Google search results have the same structure: a link at the top, a string of the URL and a description of the web page. This structure enables Studio to easily extract the information, as it always knows where to find it. In our project, we are trying to scrape data from popular job search portals such as Naukri, indeed, LinkedIn. We are going to input the requirement from the user. We will analyze and scrape the required data and at last we'll provide a combined result of various jobs availablein these popular websites.

Our team has worked on creating the job portal site from the pre-processed information obtained from the following websites:

NAUKRI:
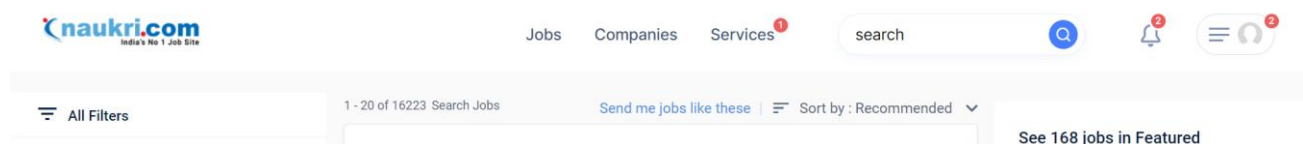
GLASSDOOR:
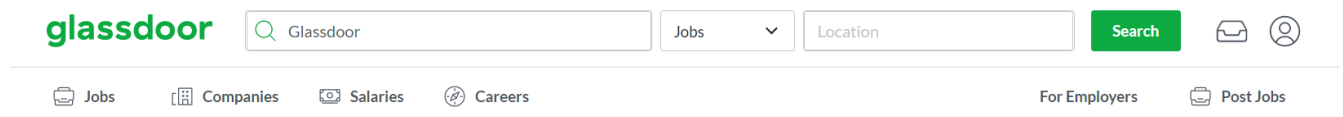
MONSTER:
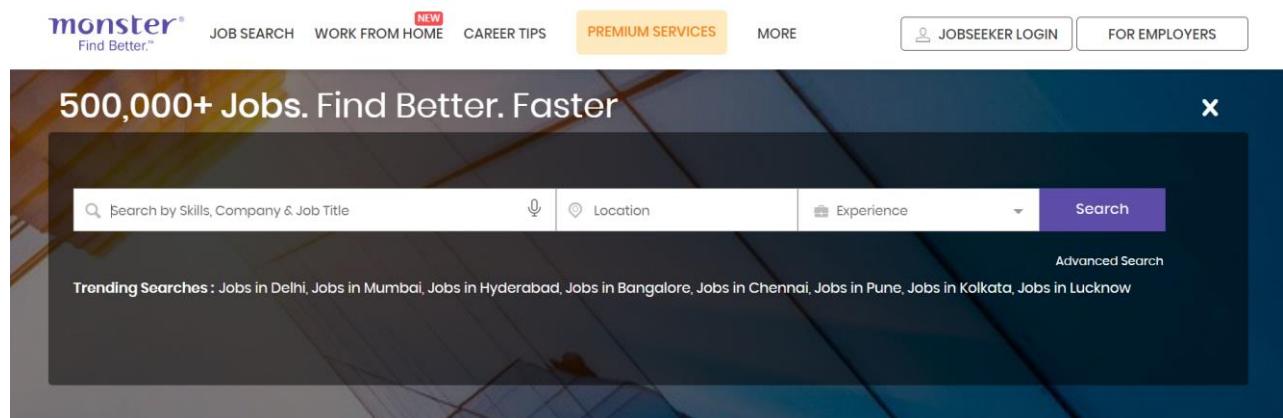
Data Scraping Workflow - Naukri:

The Data Scraping Extract Wizard allows the selection of each data element and another element of the same type in the list type in order to enable multi-extraction without needing to indicate all the repeated elements.

We have indicated and extracted the following columns of information using the Wizard as follows:

Similarly, we can extract other information such as: Job Experience required, description, Company Name, Ratings, Salary provided.

Data Scraping Activities,
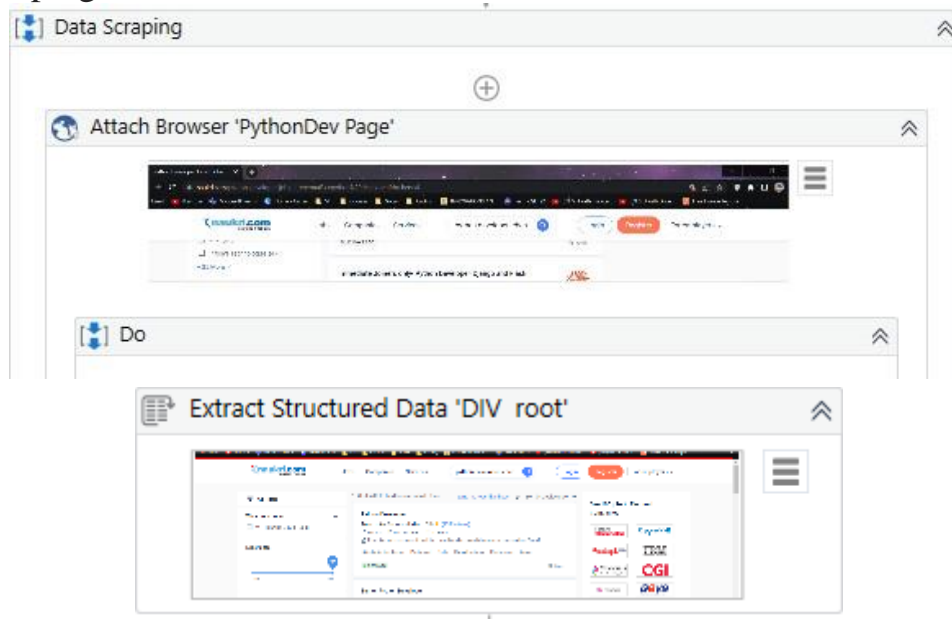
Similarly, the data scraping workflows for the other websites are made and the data is saved to individual files which can be read in the main workflow.

3.2 DISPLAY MODULE:

So, we have gathered the requirement from the user, we have scraped the necessary data from various job portals and we have separate csv files for each portal which has all thenecessary information. Now, we are going to develop a full stack website with a proper frontend design to display the consolidated information which is convenient for the user to check.The website displays the job vacancies from various portals at a single place.

We take input from a targeted user, based on their requirement we are scrapping data from popular sites such as Naukri and Indeed separately. After the scrapping, we are generating an excel file with the top 50-100 records of job information extract and all the details about various jobsare stored as separate excel files for Naukri jobs and Indeed jobs. We will be displaying the consolidated file without any duplicate columns in our website

Our Job portal Automation Site, this site has a home page when you click your jobs button it will re-direct you to the list of jobs available respective to the input given. This site also has a about page which displays the main idea of the website. Contact page which displays the details of the developers and finally the feedback page which asks for the user to login and give feedback
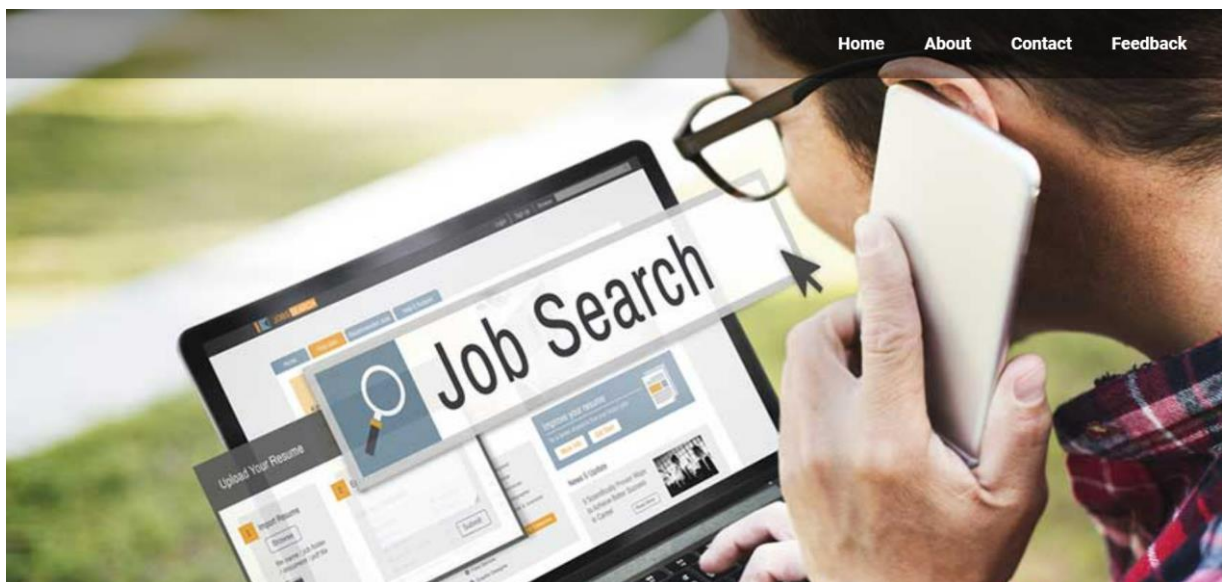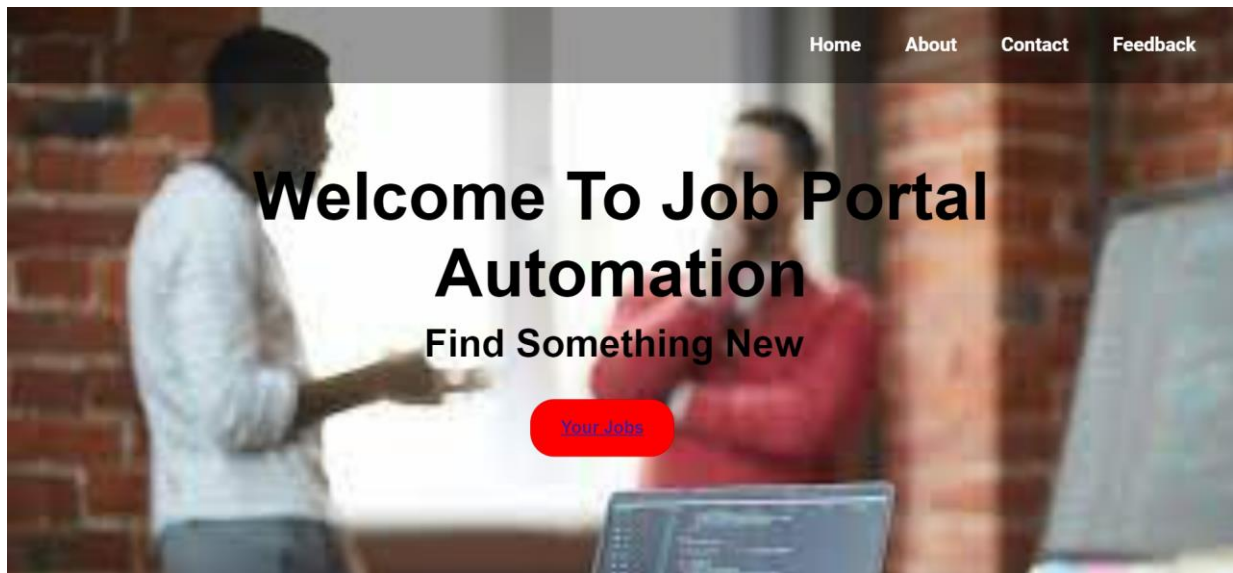
Who We Are ?

From a newly graduated student or an experienced professional in the field, looking for a new job can be highly demanding and intimidating. Websites which confer fraudulent job positions or allow risks of personal identity theft must be strictly avoided. With the increasing number of job portals which lists different jobs during different times, it becomes a frustrating experience. It was becoming laborious to continually check and change expectations of the individual to receive best results and new roles which had been posted.

We already have a plethora of data on the search, from LinkedIn, indeed, Glassdoor etc., that host many opportunities per day. Rushing into interviews with only a few offers leads to settling for lesser offers can induce great struggle in the future. But have you ever noticed that usually, they all contain different jobs? So, how about we scrap the data from these websites to build a collective job search portal? The main idea of the project is to scrap popular job portal websites and obtain information like the date of the job posting, salary details, job industry, company name, etc. Then, the problem statement writes as an automated, combined job search portal which enables newest information on job position meeting the demands of the user. And, to handle this huge amount of information from many different job websites such as LinkedIn, Naukri, indeed etc, we'll be using the method of Data Scraping.

Contact information

This website is designed and developed by

Karthavya D

Ranjani T

Likith Sai

✉ Karthavya.t2020@vitstudent.ac.in   ✉ ranjani.t2020@vitstudent.ac.in   ✉ likithsai2020@vitstudent.ac.in
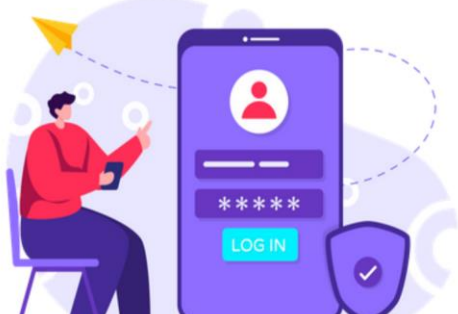
📇 9587364678   📇 9587364678   📇 9587364678

Register to Join!

✉ Email

🔒 Password

REGISTER

CHAPTER 4

WORKFLOW

The MAIN Workflow:

Collecting requirements from the user in the MAIN WORKFLOW:

Invoking the individual workflows to the MAIN file and passing the requirement arguments:

Reading each file from the main workflow:

Read CSV

Read from file
"Naukri.csv"

Output to
naukri

☑ Has headers

Read CSV

Read from file
"Glassdoor.csv"

Output to
glassdoor

☑ Has headers

Read CSV

Read from file
"Glassdoor.csv"

Output to
glassdoor

☑ Has headers
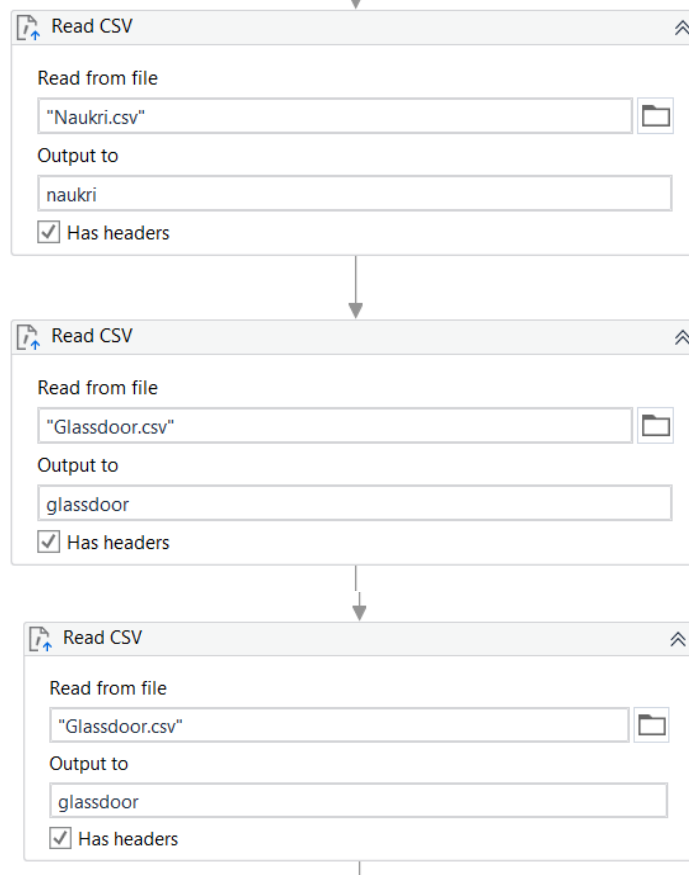
Removing duplicate rows in data tables and declaring variables for pre -processing:

Remove Duplicate Rows

DataTable
naukri   ...

Remove Duplicate Rows

DataTable
glassdoor   ...

Remove Duplicate Rows

DataTable
mons   ...

Clear Data Table

DataTable
gd2   ...

Clear Data Table
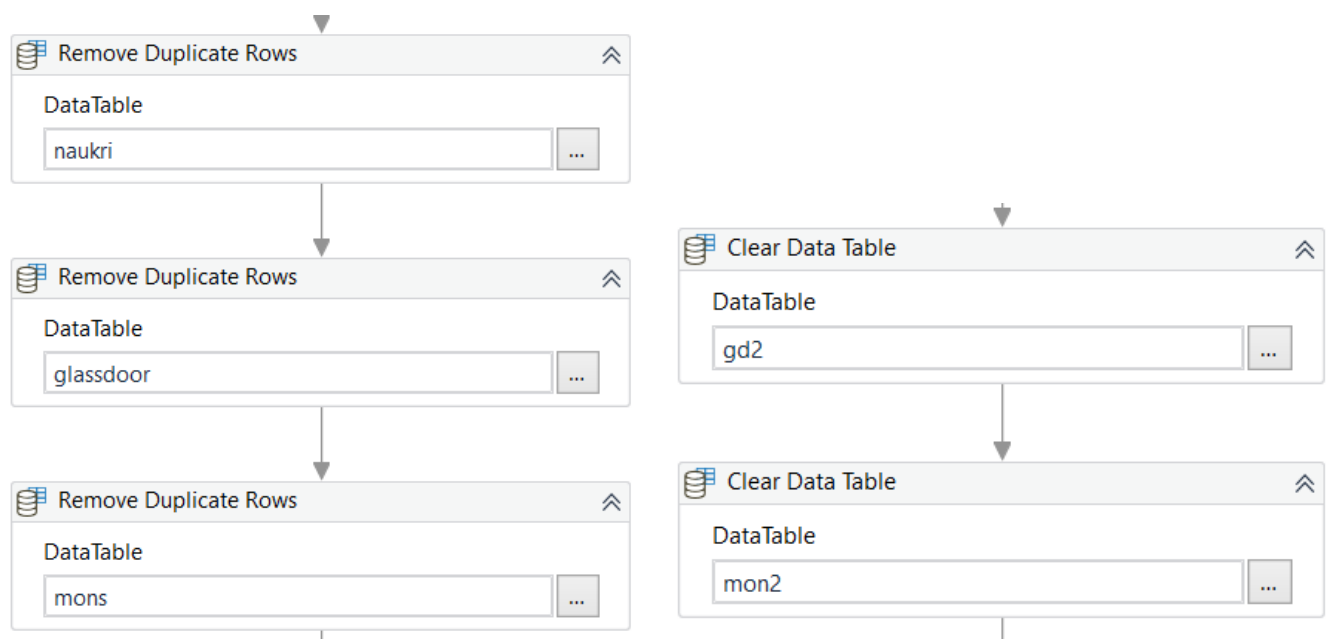
DataTable
mon2   ...

Removing repeated job positions from data tables to print only unique and relevant information in the Website for easy access of the users:

Comparing Glassdoor and Naukri:

For Each Row in Data Table
ForEach: CurrentRow
In: glassdoor
Body
Lookup Data Table
Column: The variable containing the column you want to sea
ColumnIndex: The index of the column you want to search for. Onl
ColumnName: "Job Position Name"
DataTable: naukri
LookupValue: CurrentRow("Position").ToString

If
Condition: r<>-1
Then | Else
Drop activity here
Sequence
Add Data Row
ArrayRow: CurrentRow.ItemArray
DataRow: The DataRow object to be added to the DataTable.
DataTable: gd2

Comparing Monster and Naukri:

For Each Row in Data Table
ForEach: CurrentRow
In: mons
Body
Lookup Data Table
Column: The variable containing the column you want to sea
ColumnIndex: The index of the column you want to search for. Onl
ColumnName: "Job Position Name"
DataTable: naukri
LookupValue: CurrentRow("Position").ToString

If
Condition: r<>-1
Then | Else
Drop activity here
Sequence
Add Data Row
ArrayRow: CurrentRow.ItemArray
DataRow: The DataRow object to be added to the DataTable.
DataTable: mon2

Comparing Glassdoor and Monster:

For Each Row in Data Table
ForEach: CurrentRow
In: mon2
Body
Lookup Data Table
Column: The variable containing the column you want to sea
ColumnIndex: The index of the column you want to search for. Onl
ColumnName: "Position"
DataTable: glassdoor
LookupValue: CurrentRow("Position").ToString

If
Condition: r<>-1
Then | Else
Drop activity here
Sequence
Add Data Row
ArrayRow: CurrentRow.ItemArray
DataRow: The DataRow object to be added to the DataTable.
DataTable: mon3

Saving the extracted and processed information to the final file to send the data to the website:
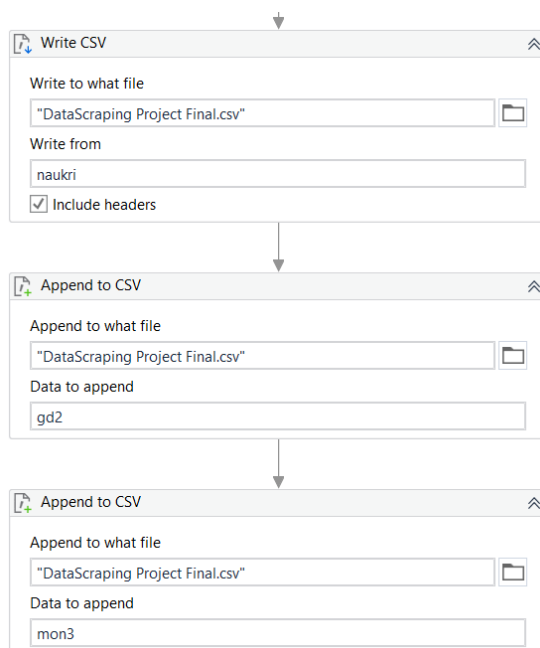
CHAPTER 5

RESULTS AND DISCUSSION

The flow of the process can be mapped through the below screenshots of the output obtained from the MAIN Workflow to the Website:
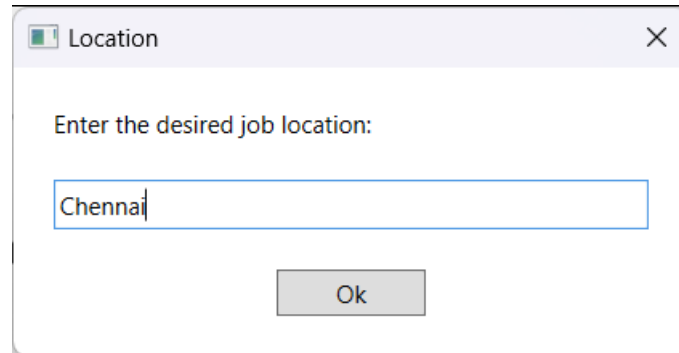
Requirements Input:

Position:

Location:

Scraping the information for the above requirement:

Extracted individual information from all websites with a total of **300 data rows:**

Naukri:

94	Associate	Shell	Not disclo:	3-5 Yrs	4.4	https
95	Senior Dat	Kaleidofin	Not disclo:	4-9 Yrs	3.7	https
96	Data analy	Adecco Inc	8,50,000 -	5-10 Yrs		https
97	Data Analy	L A Consul	Not disclo:	2-6 Yrs		https
98	Clinical Da	Huquo Cor	Not disclo:	5-10 Yrs		https
99	Assistant N	Huquo Cor	Not disclo:	2-7 Yrs		https
100	Data Analy	Dotsolved	7,00,000 -	5-9 Yrs	3.8	https
101	Data Analy	Arcadia	Not disclo:	1-6 Yrs		https

Glassdoor:

99	Senior Cor	AstraZeneca	
100	Senior Infr	BNY Mello	₹330K - ₹800K (Glassdoor est.)
101	Business A	Calibraint	₹455K - ₹1M (Glassdoor est.)
102			

Monster:

88	Team Lead	NielsenIQ	3-4 Years
89	Data Proce	NielsenIQ	Data Processing Analyst - Core SO
90	Business P	Ramboll India Private	3-5 Years
91	Hyperion A	Zillion Technologies (I	5-14 Years
92			

The final csv file after pre-processing consists of only **231 data rows** after removing and filtering out repeated work postings from multiple websites:

The data as displayed in the Website:

Your Automated job list

<div>Load Data</div>

Job Position Name	Company Name	Pay	Experience	Rating	URL
Sr.Executive Assistant - Business Analyst	DATA PATTERNS (I) PVT LTD	Not disclosed	7-9 Yrs	3.7	https://www.naukri.com/job-listings-sr-executive-assistant-business-analyst-data-patterns-i-pvt-ltd-chennai-7-to-9-years-120222500107
Clinical Data Analyst	Navitas life science	Not disclosed	2-3 Yrs	3.9	https://www.naukri.com/job-listings-clinical-data-analyst-navitas-life-science-chennai-bangalore-bengaluru-2-to-3-years-240822501456
data analyst	Disha Consultants	Not disclosed	2-4 Yrs	4.6	https://www.naukri.com/job-listings-data-analyst-disha-consultant-pune-chennai-jaipur-bangalore-bengaluru-2-to-4-years-

CHAPTER 6

CONCLUSION

The job automation portal basically scrapped the data from three websites: Naukri, Glass Door and Monster. For example: The user wanted to search for Data science analyst, in the location Chennai. For the above required condition, the data will be scraped from all the three websites and will get stored in a separate csv file. All the files will be pre-processed such that repetitive jobs scrapped from all the three sites is removed. We have also processed in such a way that the best 100 results will be displayed without duplicated jobs available in each site. Finally, the result obtained is stored a single csv file. The downloaded final csv file is loaded in our website and it's displayed for the user. Our website will also have some additional features like registration/login and feedback.

References:

1. Inspiration for the Project from UiPath Forum – https://forum.uipath.com/t/how-do-you-scrape-the-jobs-listings-from-indeed-com-and-apply-to-ones-that-mention-your-degree-skills/286486

2. Data Scraping Reference – https://docs.uipath.com/studio/docs/about-data-scraping

3. Invoke Workflow – https://docs.uipath.com/activities/docs/invoke-workflow-file

4. Reusing Automation Workflows – https://www.uipath.com/learning/video-tutorials/reusing-ui-automation-invoke-workflow-and-templates

5. Remove Duplicate Rows UiPath Helpdesk – https://forum.uipath.com/t/remove-duplicate-rows-from-datatable/160509

6. CSV File Activities – https://docs.uipath.com/activities/docs/read-csv-file

7. Website Development – HTML - https://www.w3schools.com/html/

8. Website Development – CSS https://www.w3schools.com/css/

9. Website Development – JavaScript https://www.w3schools.com/js/

10. https://www.geeksforgeeks.org/begin-web-development-with-a-head-start/

11. Website Development Application applied – https://code.visualstudio.com/

12. Server used for hosting the developed website locally – https://www.wampserver.com/en/