

AUTOMATED HOUSE FINDER

A PROJECT REPORT

Submitted by

Rishi Karthick P (220701225)

in partial fulfillment for the course

OAI1903 -INTRODUCTION TO ROBOTIC PROCESS AUTOMATION

for the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

**RAJALAKSHMI ENGINEERING COLLEGE RAJALAKSHMI
NAGAR**

THANDALAM CHENNAI – 602105

NOVEMBER 2024

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI -602105

BONAFIDE CERTIFICATE

Certified that this project report “**Automated House Finder**” is the bonafide work of “**RISHI KARTHICK P(220701225)**” who carried out the project work for the subject OAI1903 - Introduction to Robotic Process Automation under my supervision.

SIGNATURE

Dr. N. Durai Murugan, M.E., Ph.D.,
SUPERVISOR, Associate Professor,
Department of Computer Science and
Engineering, Rajalakshmi Engineering
College, Rajalakshmi Nagar,
Thandalam, Chennai – 602105.

Submitted to Project and Viva Voce Examination for the subject OAI1903 -
Introduction to Robotic Process Automation held on _____.

Internal Examiner

External Examiner

ABSTRACT

The project aims to automate the collection and delivery of real estate data for houses and apartments based on user-provided inputs such as state name and email address. Using UiPath, the workflow extracts property details like price, location, size, and other relevant specifications from real estate websites through web scraping techniques. This approach ensures the accurate and efficient gathering of data while minimizing the need for manual intervention.

Once the data is extracted, it is structured and stored in an Excel sheet for better readability and analysis. The process is designed to handle varying data formats from multiple real estate platforms, ensuring consistency and accuracy in the final dataset. Additionally, the project leverages UiPath's robust automation features to dynamically process input parameters and manage large volumes of data seamlessly.

After the data is processed, the Excel sheet is attached to an automated email and sent to the user-provided email address. This ensures that users receive real-time insights into the real estate market directly in their inbox. The project demonstrates the power of Robotic Process Automation (RPA) tools in enhancing productivity and streamlining repetitive tasks, particularly in data-intensive domains like real estate

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S. Meganathan, B.E, F.I.E.**, our Vice Chairman **Mr. Abhay Shankar Meganathan, B.E., M.S.**, and our respected Chairperson **Dr. (Mrs.) Thangam Meganathan, Ph.D.**, for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. Murugesan, M.E., Ph.D.**, our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. Kumar, M.E., Ph.D.**, Professor and Head of the Department of Computer Science for her guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides, **Mrs. G. M. Sasikala, M.E.**, Assistant Professor (SG), Department of Computer Science and Engineering, Rajalakshmi Engineering College for their valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinators, **Dr. N. Durai Murugan, M.E., Ph.D.**, Associate Professor, and **Mr. B. Bhuvaneswaran, M.E.**, Assistant Professor (SG), Department of Computer Science and Engineering for their useful tips during our review to build our project.

Rishi Karthick P (220701225)

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	iii
	LIST OF FIGURES	vi
	LIST OF ABBREVIATIONS	vii
1.		
	INTRODUCTION	
1.1	INTRODUCTION	1
1.2	OBJECTIVE	3
1.3	EXISTING SYSTEM	3
1.4	PROPOSED SYSTEM	4
2.	LITERATURE REVIEW	5
3.	SYSTEM DESIGN	9
3.1	SYSTEM FLOW DIAGRAM	9
3.2	ARCHITECTURE DIAGRAM	10
3.3	SEQUENCE DIAGRAM	11
4.	PROJECT DESCRIPTION	12
4.1	MODULES	
12		
4.1.1.	INPUT HANDLING AND	
12		
4.1.2.	CONTENT ANALYSIS	
12		
4.1.3.	RESULT MANAGEMENT	
13		
4.1.4.	COMPLETION AND REPORTING	13
5.	OUTPUT SCREENSHOTS	14
6.	CONCLUSION	18
	APPENDIX	19
	REFERENCES	25

LIST OF FIGURES

Figure No.	Figure Name	Page No.
3.1	System Flow Diagram	9
3.2	Architecture Diagram	10
3.3	Sequence Diagram	11
5.1	Input Dialog	14
5.2	Web Scraping	14
5.3	Data Storage in Excel	15
5.4	Email Automation	16

LIST OF ABBREVIATIONS

ABBREVIATION	ACRONYM
RPA	Robotic Process Automation
SMTP	Simple Mail Transfer Protocol
API	Application Programming Interface
CSV	Comma-Separated Values
OCR	Optical Character Recognition

CHAPTER 1 INTRODUCTION

1.1 INTRODUCTION

In today's fast-paced digital age, automation plays a critical role in optimizing time-consuming and repetitive tasks. This project leverages the capabilities of **UiPath**, a leading Robotic Process Automation (RPA) tool, to simplify and streamline the process of collecting real estate data. The manual process of searching for property information, compiling it into a structured format, and delivering it to users can be labor-intensive and error-prone. By automating this workflow, the project ensures accurate and efficient data extraction, storage, and dissemination.

The automation begins with user input, where the state name and email address are provided. Using web scraping techniques, the system accesses real estate websites to collect detailed property information such as location, price, size, and other relevant attributes. This data is then systematically organized into an Excel sheet to maintain consistency and usability. The inclusion of error-handling mechanisms ensures that the system operates robustly, even when encountering unexpected challenges, such as website changes or missing data.

To deliver the extracted information seamlessly, the project integrates an automated email system. The generated Excel sheet containing real estate data is sent to the user's provided email address, ensuring quick and convenient access. This project not only highlights the versatility of UiPath but also demonstrates how automation can be effectively applied to real-world problems, saving time and effort while enhancing accuracy and productivity.

1.2 OBJECTIVE

The primary objective of this project is to develop an automated system for real estate data collection and delivery that is efficient, accurate, and user-friendly. The specific objectives include:

1. **Automating Real Estate Data Retrieval:** Eliminate the manual effort involved in searching for property details by automating the process of web scraping from real estate websites based on user-provided inputs, such as state name.
2. **Structuring and Storing Data Effectively:** Organize the scraped property information, including price, location, size, and specifications, into a standardized Excel sheet format for better readability and usability.
3. **Enhancing Data Accessibility:** Automate the delivery of the generated Excel sheet via email to the user's specified email address, ensuring quick and seamless access to the required information.
4. **Implementing Robust Error Handling:** Incorporate mechanisms to handle unexpected challenges, such as changes in website structures or incomplete data, ensuring the system remains reliable and consistent.
5. **Leveraging UiPath's Capabilities:** Utilize UiPath's robust features to design and implement the automation workflow, demonstrating its effectiveness in solving real-world automation challenges.
6. **Reducing Manual Effort and Errors:** Minimize human intervention in repetitive tasks, thereby reducing errors and saving time for users while ensuring data accuracy.
7. **Improving Scalability and Adaptability:** Design the system to handle large datasets and adapt to different real estate platforms, making it suitable for various use cases and user requirements.

By achieving these objectives, this project aims to deliver a scalable and efficient solution for real estate data automation, highlighting the transformative potential of Robotic Process Automation (RPA) in data-driven tasks.

1.3 EXISTING SYSTEM

In the existing real estate data collection process, organizations typically rely on manual methods for gathering, storing, and processing information. This process often involves multiple time-consuming steps, including manually browsing real estate websites, extracting details of properties, and then copying the data into spreadsheets or other storage systems. Afterward, these details are often analyzed and reported using manual effort, such as creating reports or sending email updates.

This manual system has several limitations:

1. **Time-Consuming Process:** Manually browsing websites, searching for property listings, and recording the data requires significant effort, particularly when dealing with multiple properties across different locations.
2. **Error-Prone Workflow:** Manual entry of property details such as prices, location, and features often leads to errors, such as incorrect data entry, missed properties, or outdated information, undermining the accuracy and reliability of the real estate reports.
3. **Lack of Scalability:** As the volume of property listings increases, the manual process becomes increasingly inefficient. It becomes difficult to handle a large number of listings, making the system unsuitable for businesses that deal with large-scale data collection.
4. **Data Inconsistencies:** Manually copying data from different sources and inputting it into different systems can lead to inconsistencies, such as missing details or formatting errors, which affect the quality and completeness of the final data.
5. **Limited Reporting Flexibility:** The manual approach provides limited flexibility for generating reports or providing data in different formats, such as sending the details in an email or converting them to Excel for further analysis.
6. **Slow Response Time:** Since the process involves several steps, including browsing multiple listings and compiling the information manually, the response time can be slow. This

delays the time-sensitive decision-making process for real estate agencies or individuals relying on up-to-date property data.

Despite its inefficiencies, many organizations still use this manual approach because it's familiar and no affordable automated solutions have been implemented.

1.4 PROPOSED SYSTEM

The proposed **Houses Data Collection and Reporting System** automates the process of gathering, organizing, and sending property data. The system performs the following steps:

1. **Automated Data Collection:** The system automatically scrapes property listings from predefined real estate websites named Zillow based on user input (such as state or city). It gathers data such as property names, prices, locations, number of bedrooms, bathrooms, square footage, and more.
2. **Data Storage:** The scraped property data is automatically organized into an Excel file, with fields structured for easy viewing and analysis (e.g., property name, price, location, etc.). This eliminates the need for manual data entry and ensures accuracy.
3. **Report Generation:** Once the data is collected, the system generates a report in Excel format, with the option to filter or sort based on specific parameters (e.g., properties in a certain price range or location).
4. **Email Automation:** Once the report is generated, the system automatically sends the Excel report to a specified email address, ensuring that the relevant stakeholders receive the latest property data without manual intervention.
5. **Error Handling:** The system includes built-in error handling to detect issues such as missing data, changes in website layout, or incorrect data entries. This ensures that the data collection process remains smooth and consistent over time.

This system saves time, reduces errors, and improves efficiency, making it scalable for organizations of any size. By automating the process, it ensures real-time, accurate data for better decision-making.

CHAPTER 2 LITERATURE REVIEW

2.1 Survey on Robotic Process Automation (RPA) in Education:

Robotic Process Automation (RPA) has emerged as a key technology in automating repetitive tasks, particularly in industries like education and business. In the context of document generation, RPA has proven to be transformative, especially in automating tasks like ID card creation, certificate generation, and report processing. The primary advantage of RPA is its ability to streamline workflows, reduce manual efforts, and improve accuracy. Studies have shown that RPA tools, such as UiPath, are being utilized to automate processes like student record management, result generation, and the creation of report cards. This not only saves time but also enhances scalability by allowing educational institutions to handle large volumes of documents efficiently.

For instance, a study conducted by researchers in the educational sector showed how RPA significantly improved administrative efficiency by automating repetitive tasks that were traditionally done manually, such as result generation. However, challenges remain, particularly in the dynamic customization of documents and ensuring compatibility across different file formats. Despite these challenges, RPA tools have proven to be cost-effective solutions, particularly for organizations with limited resources, making automation accessible without large upfront investments.

2.2 Survey on Template-Based Document Automation:

Template-based document automation has been a prominent research area, particularly in the context of industries requiring large-scale document generation. Traditional methods often involve manual entry of data into pre-designed templates, leading to inefficiencies and inconsistencies. Template-based systems, when combined with automation frameworks, have the potential to significantly streamline document workflows by improving speed, accuracy, and consistency.

A key study in the field, published by the International Journal of Innovative Technology and Exploring Engineering (IJITEE), explored the use of RPA to dynamically populate templates using structured input data from tools like Microsoft Excel. This approach has been found to increase the accuracy of outputs by eliminating human error, particularly in tasks that involve high-volume document generation, such as ID cards, certificates, and reports. The findings of this research reinforce the potential of RPA in document automation, as it offers not only faster processing but also increased scalability and customization. By integrating automation into template-based systems, organizations can create tailored documents at scale, reducing manual intervention and improving overall efficiency.

2.3 Survey on Error Reduction in Document Workflows:

Ensuring accuracy in document workflows is a critical concern, as human errors in manual processes, such as incorrect data entry or formatting mistakes, can lead to significant inefficiencies. Research into RPA-driven solutions has highlighted the role of automation in minimizing errors by automating data extraction and template population processes.

One notable paper from the field of business process automation discussed how automation systems can validate input data and flag inconsistencies before document generation. This approach is particularly useful in ID card creation, where errors like misspelled names, incorrect photos, or mismatched data can result in serious administrative challenges. The study found that RPA systems not only catch errors in real-time but also ensure that the generated documents adhere to predefined formats, eliminating the need for manual validation. This feature is especially beneficial in reducing the time and resources spent on error correction, making the overall document generation process more reliable.

2.4 Summary of the Intersection of RPA and Document Automation:

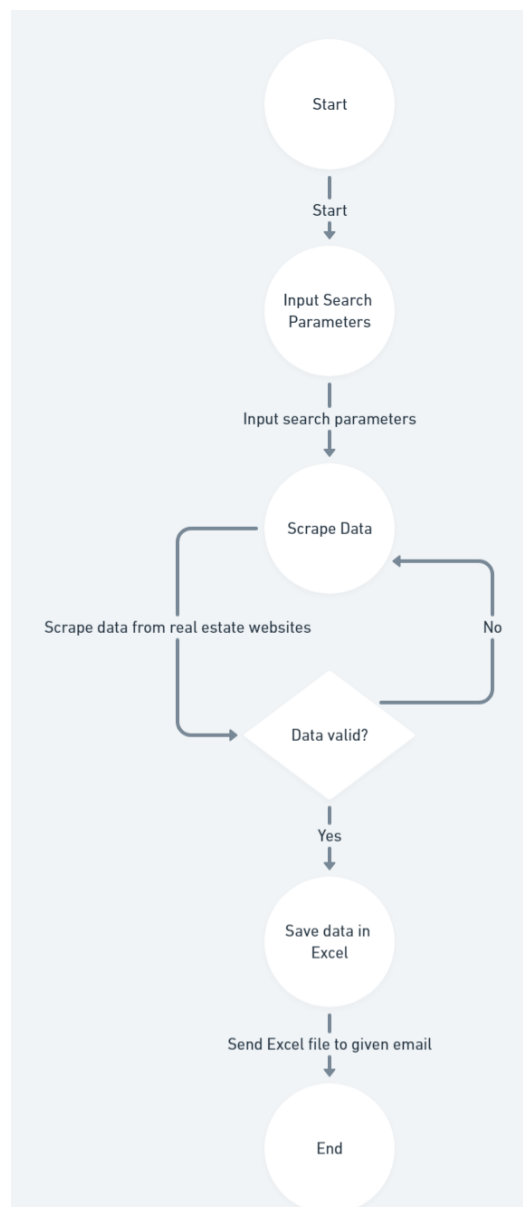
The integration of Robotic Process Automation with document generation tools, as explored in the literature, demonstrates its transformative potential. By leveraging platforms like UiPath alongside Microsoft Excel and Word, organizations can automate the entire document creation workflow. This includes data entry, template population, and document generation, ensuring accuracy, scalability, and customization.

Drawing from the studies above, it is clear that RPA addresses several critical challenges in document management, such as the inefficiency of manual processes, the potential for human error, and the complexity of handling large volumes of data. The proposed **ID Card Generator Project** exemplifies how RPA can bridge these gaps by offering a solution that not only automates the creation of ID cards but also incorporates error-checking mechanisms and dynamic template population. This combination positions the project as an effective and scalable solution for modern document management needs in education and beyond.

CHAPTER 3 SYSTEM DESIGN

3.1 SYSTEM FLOW DIAGRAM

A flowchart is a type of diagram that represents an algorithm, workflow or process. The flowchart shows the steps as boxes of various kinds, and their order by connecting the boxes with arrows. This diagrammatic representation illustrates a solution model to a given problem.



3.2 ARCHITECTURE DIAGRAM

An architecture diagram is a graphical representation of a set of concepts that are part of an architecture, including their principles, elements and components. The architecture diagram for this project is in Fig. 3.2.

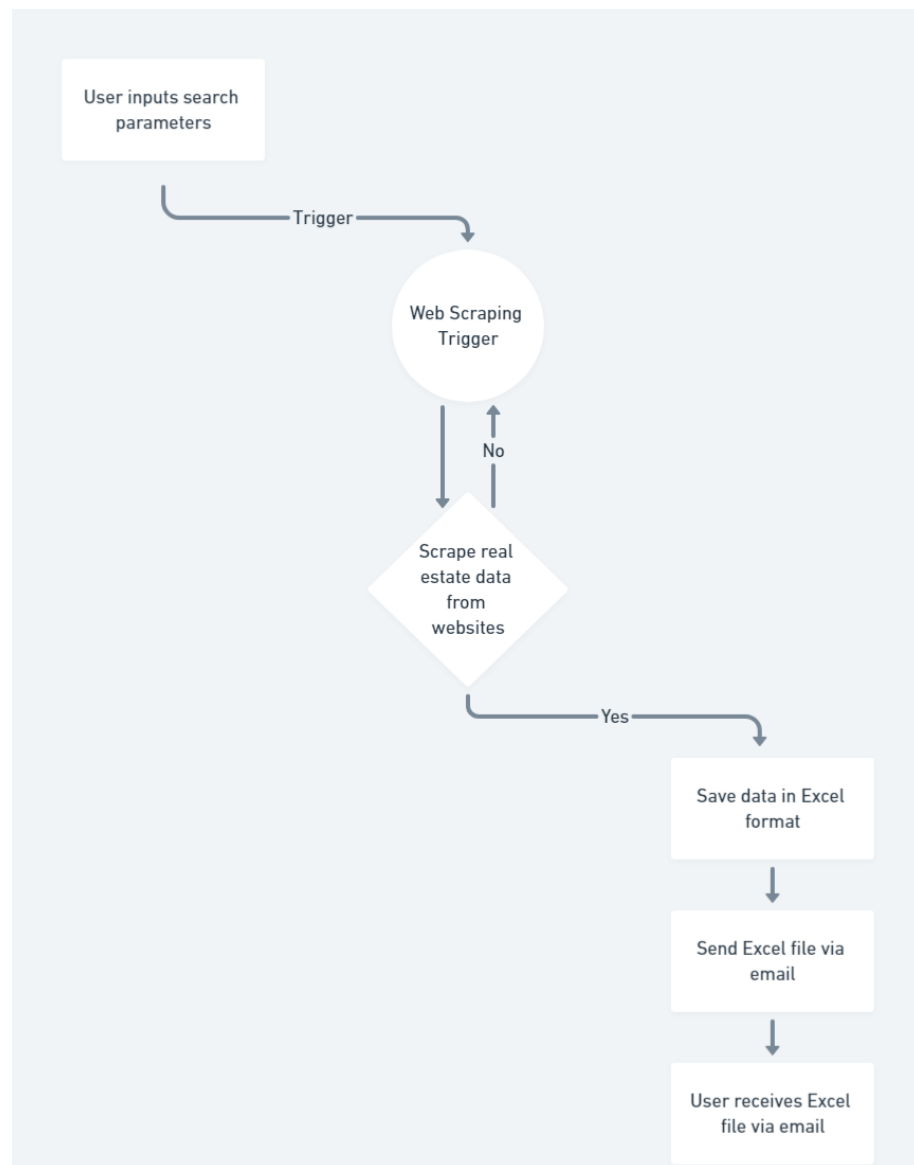


Fig 3.2 Architecture Diagram

3.3 SEQUENCE DIAGRAM

A sequence diagram is a type of interaction diagram because it describes and shows in what order a group of objects works together. The sequence diagram for this project is in Fig. 3.3.

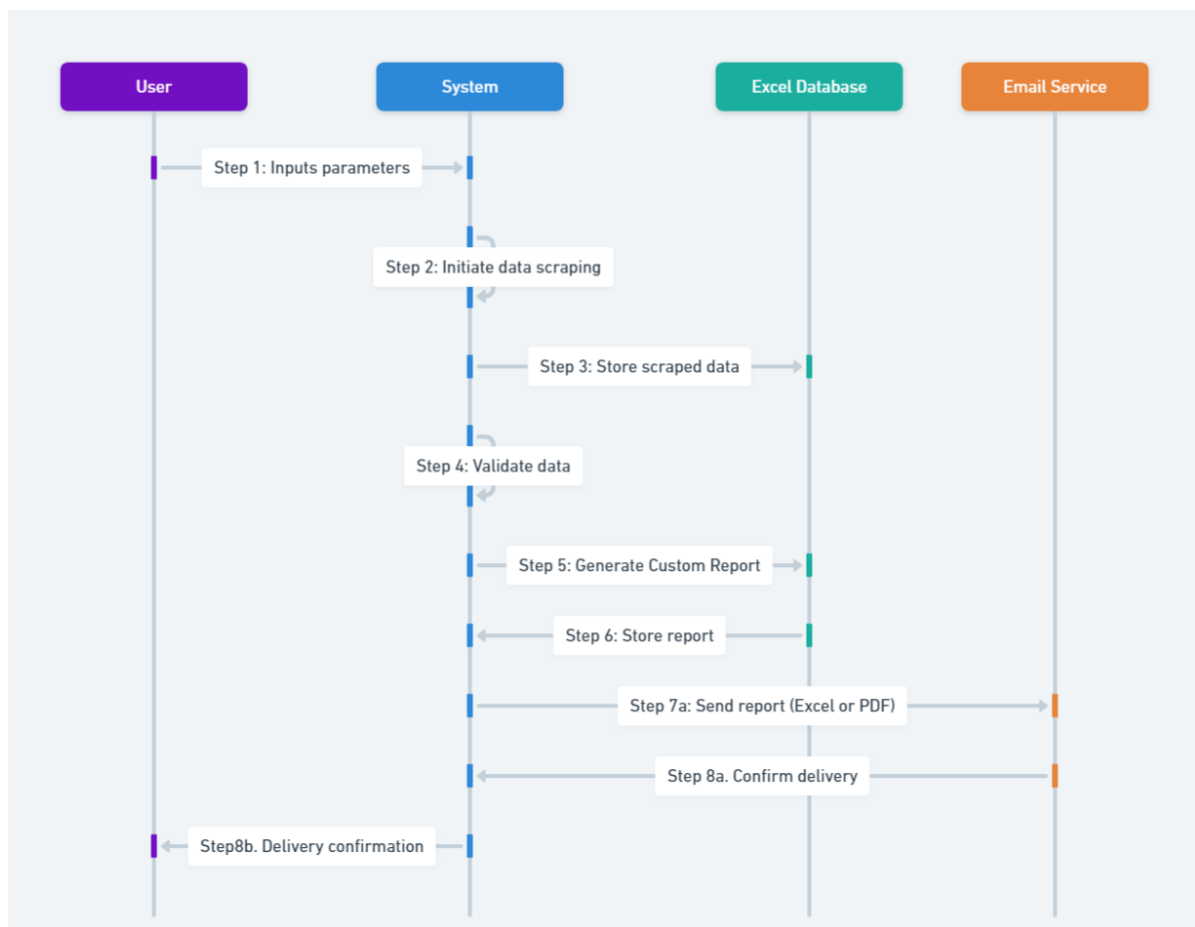


Fig 3.3 Sequence Diagram

CHAPTER 4 PROJECT DESCRIPTION

The **Houses Data Collection and Reporting System** uses web scraping and automation to simplify and accelerate the process of gathering, storing, and distributing real estate data. This solution integrates web scraping for data collection, Excel for data storage, and email automation for report delivery, providing an efficient and scalable system for real estate professionals. The system allows users to input specific criteria (e.g., location, price range) and automatically scrapes relevant property data from real estate websites. The data is then organized into a report and sent via email, all without manual intervention.

4.1. MODULES:

4.1.1. DATA COLLECTION MODULE:

The Data Collection module is responsible for gathering property data from real estate websites. Based on user-defined parameters (such as location, price range, and property type), the system scrapes listings from predefined sources. The module uses web scraping techniques to extract relevant data fields such as property name, price, location, and number of bedrooms. This module ensures accurate and up-to-date data by frequently refreshing the information from the sources.

4.1.2 DATA STORAGE AND VALIDATION MODULE:

Once the data is collected, it is stored in an organized Excel file or database. This module is responsible for verifying that the scraped data is complete, consistent, and accurate. It checks for missing or incorrectly formatted information and flags these entries for correction. The module performs basic validation, such as ensuring numerical values are valid and addresses are correctly formatted, making the data ready for further processing.

4.1.3 REPORT GENERATION MODULE:

The Report Generation module creates customized reports based on the stored data. The system allows users to filter and sort the collected property data according to various criteria (e.g., price range, number of bedrooms, etc.). Once the data is filtered, the module generates a detailed report in Excel or PDF format, which includes all relevant property information.

4.1.4 EMAIL AUTOMATION MODULE:

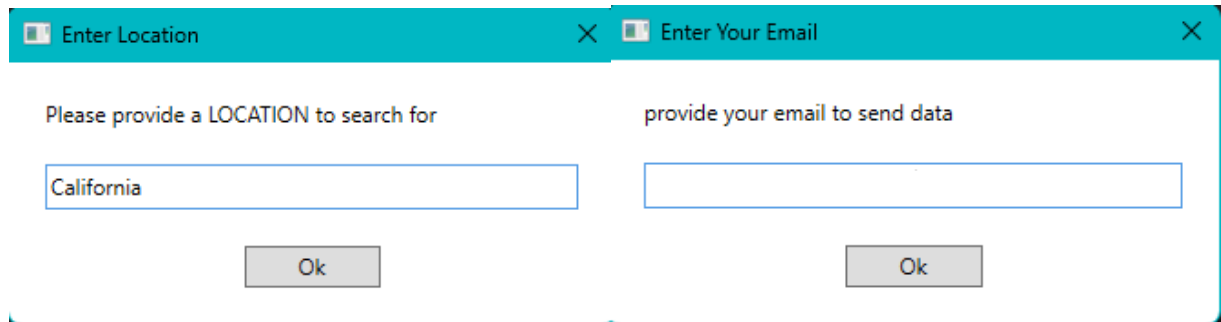
The Email Automation module is responsible for delivering the generated reports to the specified recipients. Once a report is ready, this module automatically sends the report via email to the user or stakeholders without manual intervention. Users can set up recurring emails to send reports at specific intervals (e.g., daily, weekly). The module allows for customization of email content and subject lines, ensuring that communication is professional and tailored.

4.1.5 ERROR HANDLING AND VALIDATION MODULE:

This module ensures that the entire system runs smoothly by detecting and managing any errors that may arise during the data collection, storage, or report generation process. It automatically checks for issues such as website access problems, missing data, or incorrect formatting, and generates alerts for the user to address. This proactive approach helps maintain the reliability and integrity of the data and workflow.

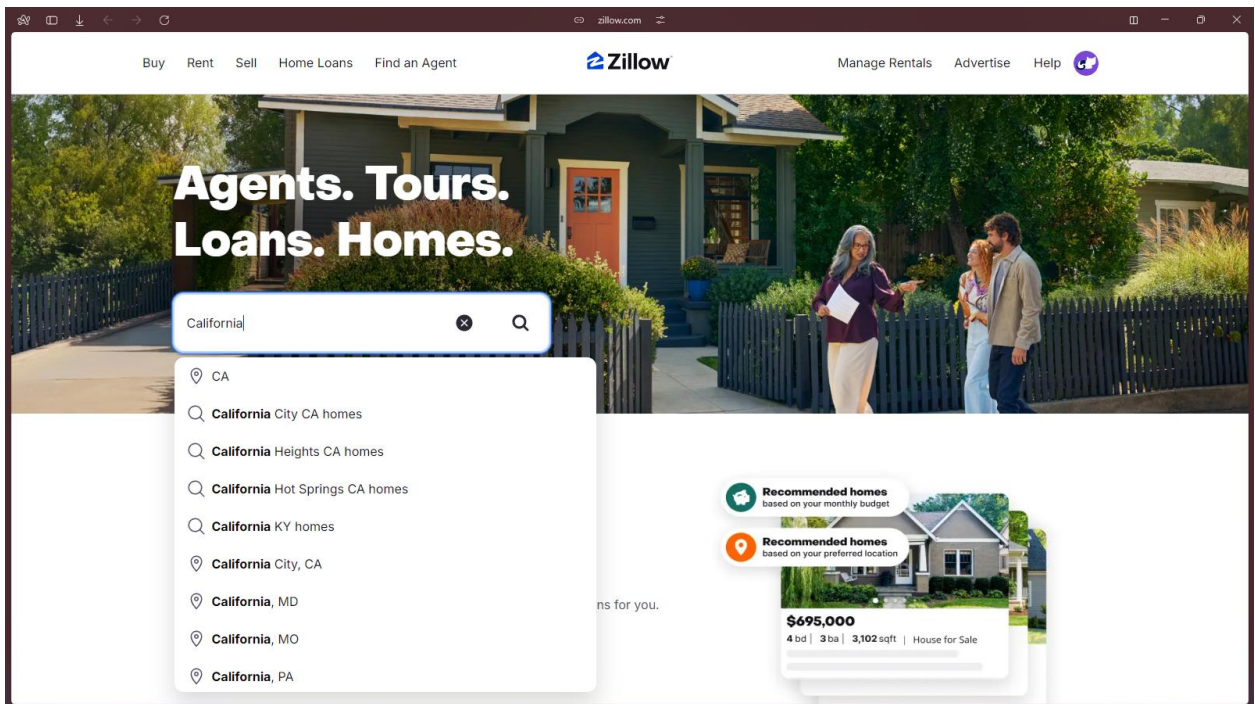
CHAPTER 5

OUTPUT SCREENSHOTS



The image shows two side-by-side input forms. The left form is titled 'Enter Location' and contains the text 'Please provide a LOCATION to search for'. Below this is a text input field with 'California' entered, and an 'Ok' button. The right form is titled 'Enter Your Email' and contains the text 'provide your email to send data'. Below this is an empty text input field and an 'Ok' button.

Fig 5.1 – Input ScreenShot in Fig 5.1



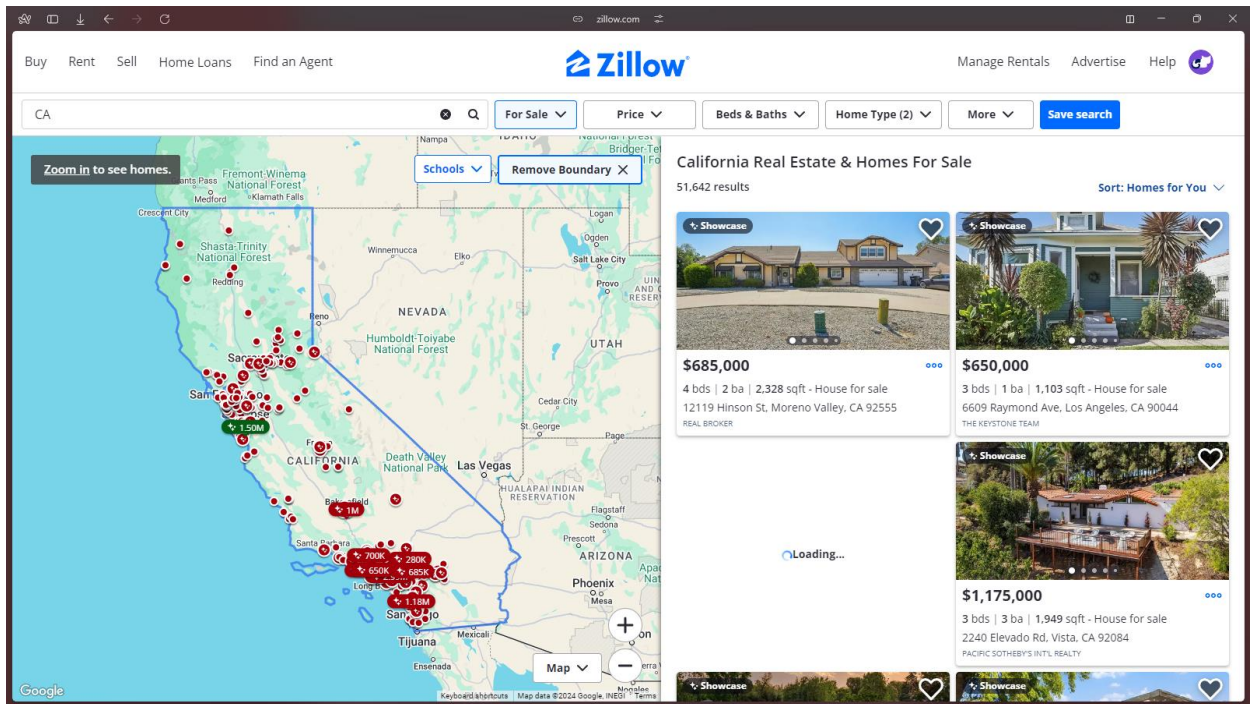


Fig 5.2 – Web Scrapping

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Price	Beds	Baths	Sqft	Price per sqft	Street Address	City	ZIP	Listing URL									
2	999900	4	4	4149	241.00	2221 Myrtle	Bakersfield	93301	https://www.zillow.com/homedetails/2221-Myrtle-St-Bakersfield-CA-93301/18882407_zpid/									
3	699999	3	2	1274	549.45	12780 Rajah	Sylmar	91342	https://www.zillow.com/homedetails/12780-Rajah-St-Sylmar-CA-91342/20102260_zpid/									
4	1465000	4	4	2524	580.43	3378 Via Loma	Fallbrook	92028	https://www.zillow.com/homedetails/3378-Via-Loma-Fallbrook-CA-92028/139758101_zpid/									
5	650000	3	1	1103	589.30	6609 Raym	Los Angeles	90044	https://www.zillow.com/homedetails/6609-Raymond-Ave-Los-Angeles-CA-90044/20934854_zpid/									
6	1175000	3	3	1949	602.87	2240 Elevado	Vista	92084	https://www.zillow.com/homedetails/2240-Elevado-Rd-Vista-CA-92084/16628778_zpid/									
7																		

Fig 5.3 – Excel File

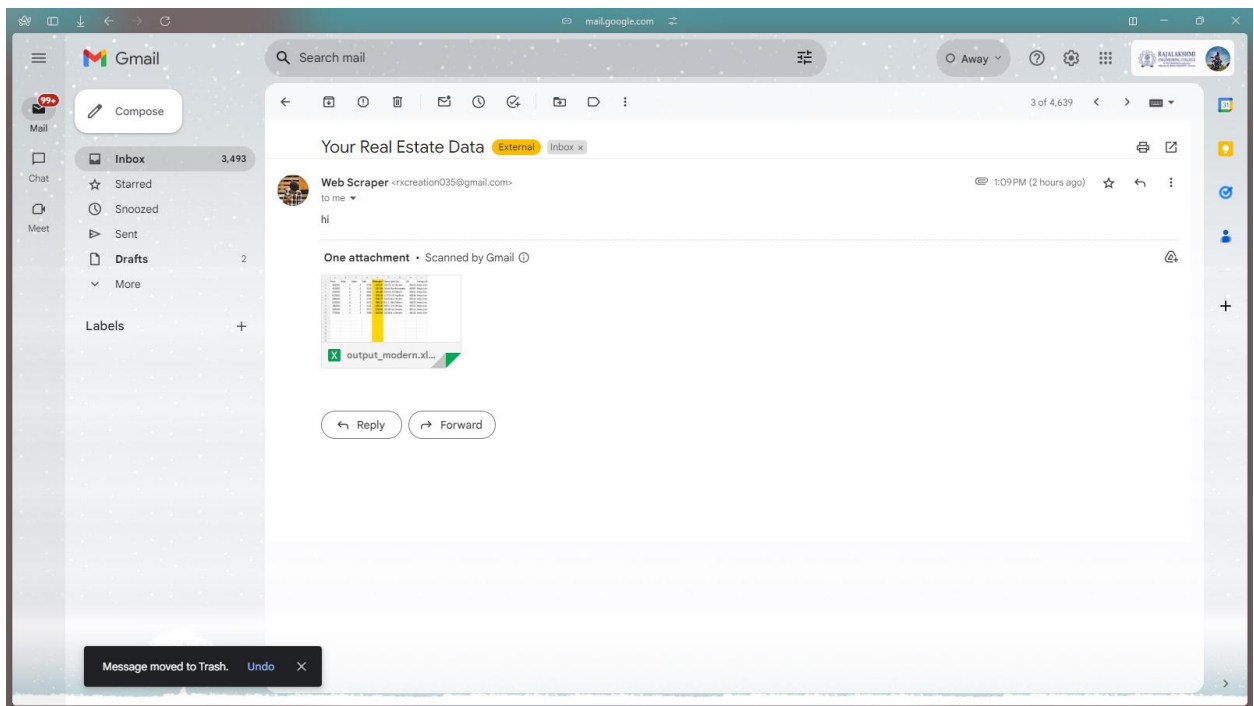


Fig 5.3 – Email Automation

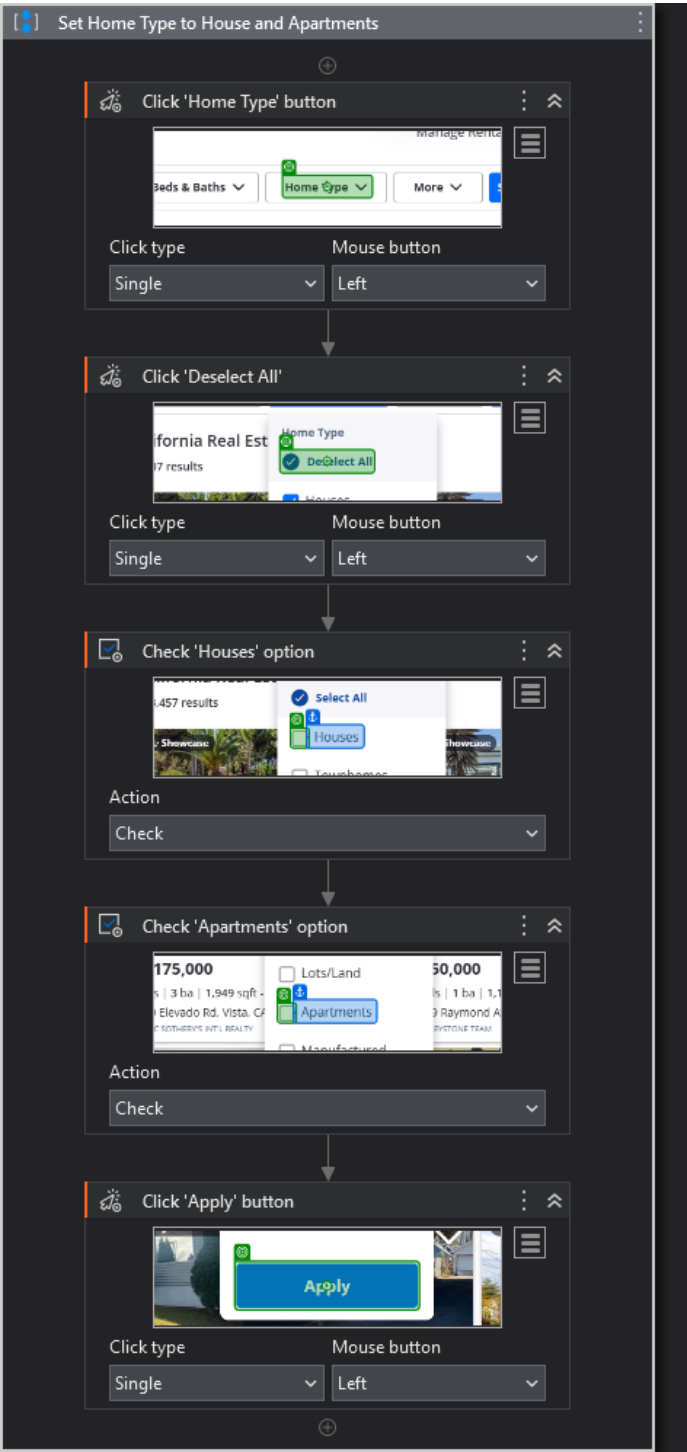
CHAPTER 6 CONCLUSION

The **Houses Data Collection and Reporting System** offers a powerful, automated solution to streamline the collection, organization, and distribution of real estate data. By integrating web scraping, Excel for data storage, and email automation, the system significantly reduces manual effort, enhances accuracy, and ensures timely delivery of essential property information.

The system's modular design allows for easy customization, enabling users to tailor the process to their specific needs, whether for small-scale operations or large real estate enterprises. With automated data collection, validation, report generation, and email notifications, the solution improves operational efficiency, enabling real estate professionals to make informed, data-driven decisions without the complexities of manual data handling.

Furthermore, the scalability of the system ensures it can handle growing data volumes and adapt to changing market needs. This project not only addresses the challenges of managing and analyzing real estate data but also empowers users with a reliable, cost-effective tool for enhancing productivity and decision-making in the fast-paced real estate industry.

APPENDIX PROCESS WORKFLOW



Insert new columns

Insert 'Street Address' column

Range *

Excel.Sheet(varLocation)

Has headers

Where *

After

Relative to column *

"Full Address"

Add header

"Strei

Format data as type

Set Format

Insert 'City' column

Range *

Excel.Sheet(varLocation)

Has headers

Where *

After

Relative to column *

"Street Address"

Add header

"City

Format data as type

Set Format

Insert 'ZIP' column

Range *

Excel.Sheet(varLocation)

Has headers

Where *

After

Relative to column *

"City"

Add header

"ZIP"

Format data as type

Set Format

Insert 'Price per Sqft' column

Range *

Excel.Sheet(varLocation)

Has headers

Where *

After

Relative to column *

"Sqft"

Add header

"Prici

Format data as type

Set Format

Insert new columns

+

Insert 'Street Address' column

Range *

{}

 Excel.Sheet(varLocation)

↵

+

☒ Has headers

Where *

After

Relative to column *

{}

 "Full Address"

↵

+

Add header

{}

 "Stre"

↵

+

Format data as type

Set Format

↓

+

Insert 'City' column

Range *

{}

 Excel.Sheet(varLocation)

↵

+

☒ Has headers

Where *

After

Relative to column *

{}

 "Street Address"

↵

+

Add header

{}

 "City"

↵

+

Format data as type

Set Format

↓

+

Insert 'ZIP' column

Range *

{}

 Excel.Sheet(varLocation)

↵

+

☒ Has headers

Where *

After

Relative to column *

{}

 "City"

↵

+

Add header

{}

 "ZIP"

↵

+

Format data as type

Set Format

↓

+

Insert 'Price per Sqft' column

Range *

{}

 Excel.Sheet(varLocation)

↵

+

☒ Has headers

Where *

After

Relative to column *

{}

 "Sqft"

↵

+

Add header

{}

 "Pric"

↵

+

Format data as type

Set Format

+

Populate new columns row by row

For each
CurrentRow

In range *

{ } Excel.Sheet(varLocation) L7 +

☒ Has headers ☒ Save after each row

Do

Check if the 'Full Address' can be split into 'Street Address', 'City' and 'ZIP'

Condition *

{ } System.Text.RegularExpressions.Regex.IsMatch L7 +

Then

Write 'Street Address' row by row

What to write *

{ } CurrentRow("Full Address").ToString.Split(", ")(0).Trim L7 +

Where to write *

{ } CurrentRow("Street Address") L7 +

☐ Auto increment row

Write 'City' row by row

What to write *

{ } CurrentRow("Full Address").ToString.Split(", ")(1).Trim L7 +

Where to write *

{ } CurrentRow("City") L7 +

☐ Auto increment row

Write 'ZIP' row by row

What to write *

{ } CurrentRow("Full Address").ToString.Split(", ")(2).Trim.Split(" ") L7 +

Where to write *

{ } CurrentRow("ZIP") L7 +

☐ Auto increment row

Else

Drop activity here or generate with Autopilot

Write 'Price per Sqft' row by row

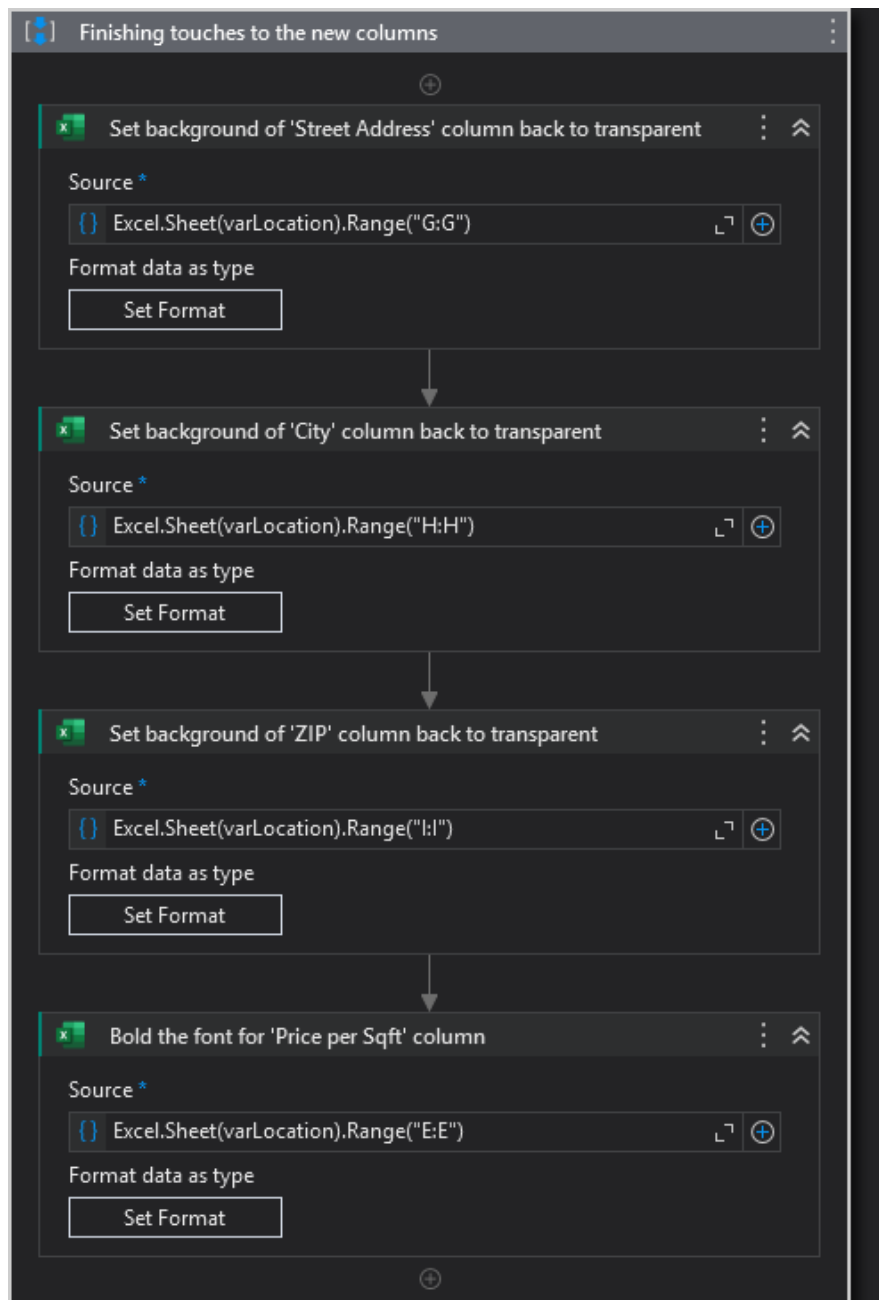
What to write *

{ } Cdbl(CurrentRow("Price"))/Cdbl(CurrentRow("Sqft")) L7 +

Where to write *

{ } CurrentRow("Price per Sqft") L7 +

☐ Auto increment row



Send SMTP Mail Message

To: email

Subject: "Your Real Estate Data"

Body: "hi"

[Attach Files](#)

REFERENCES

- [1] Kuppusamy, Palanivel & Joseph K, Suresh. (2020). [Robotic Process Automation to Smart Education](#). 3775.
- [2] Patil, Dr & Mane, Vinod & Patil, Dr. (2019). [Social Innovation in Education System by using Robotic Process Automation](#) (Rpa). International Journal of Innovative Technology and Exploring Engineering. 8. 3757-3760. 10.35940/ijitee.K2148.0981119.
- [3] Elkhatat, A.M., Elsaid, K. & Almeer, S. [Evaluating the efficacy of AI content detection tools in differentiating between human and AI- generated text.](#) *Int J Educ Integr* **19**, 17 (2023). <https://doi.org/10.1007/s40979-023-00140-5>
- [4] H. Alamleh, A. A. S. AlQahtani and A. ElSaid, "[Distinguishing Human-Written and ChatGPT-Generated Text Using Machine Learning](#)," 2023 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 2023, pp. 154-158, doi: 10.1109/SIEDS58326.2023.10137767.
- [5] Tomáš Foltýnek, Norman Meuschke, and Bela Gipp. 2019. [Academic Plagiarism Detection: A Systematic Literature Review](#). ACM Comput. Surv. 52, 6, Article 112 (November 2020), 42 pages. <https://doi.org/10.1145/3345317>
- [6] H. A. Chowdhury, D. K. Bhattacharyya, "[Plagiarism: Taxonomy, Tools and Detection Techniques](#)", 19th National Convention on Knowledge, Library and Information Networking, 2018.