NICHE HOBBY ENGAGEMENT BOT

A PROJECT REPORT

Submitted by

RUPALA N (220701233)

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 – 602 105

NOVEMBER 2024

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI - 602105

BONAFIDE CERTIFICATE

Certified that this project report **"The Niche Hobby Engagement Bot"** is the bonafide work of **"RUPALA N (220701233)"** who carried out the project work for the subject OAI1903 - Introduction to Robotic Process Automation under my supervision.

.

SIGNATURE

Dr. N. Durai Murugan
Assistant 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 _____.

ABSTRACT

The proposed project leverages UiPath's automation capabilities to create a personalized content delivery system tailored to user hobbies, specifically focusing on Reddit as the primary data source. This innovative solution automates the extraction of hobby-related information from Reddit and delivers it to users via email. By utilizing UiPath activities such as Data Scraping, For Each Row in Data Table, and Send SMTP Mail Message, the workflow processes user data efficiently, ensuring a seamless experience. The project begins by reading user information from an Excel file, including their hobbies and email addresses. For each user, the workflow dynamically searches Reddit for relevant discussions and extracts key insights related to their hobbies. The extracted data is formatted into a structured and personalized email, making it both informative and engaging. While the current implementation focuses solely on Reddit, the solution is designed with scalability in mind. Future iterations could incorporate additional platforms like Amazon, YouTube, or Twitter, making it a versatile tool for hobby-based content aggregation. The project emphasizes automation's ability to deliver tailored content, demonstrating its potential as a productivity tool for users with diverse interests. By simplifying the process of finding and curating information, the solution saves users time and effort, providing them with meaningful insights directly in their inbox. This project also underscores the potential for extending UiPath workflows to real-world applications such as personalized marketing, news aggregation, and customer engagement. With its current focus on Reddit and its future adaptability, this solution serves as a prototype for building advanced, scalable automation tools to meet individual user needs.

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour 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 endeavouring 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. Revathy, M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Design for her guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides, **Mrs. Roxanna Samuel, M.E.,** Assistant Professor (SG), **Ms. Farjana, M.E.,** Assistant Professor (SG), **Ms. Vinothini, 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.

- Rupala N (220701233)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATION	ACCRONYM
RPA	Robotic Process Automation
AI	Artificial Intelligence
API	Application Programming Interface
SMTP	Simple Mail Transfer Protocol

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

In the age of digital transformation, automation is playing a pivotal role in reshaping how individuals and businesses interact with information. One area where automation can be particularly impactful is in personalized content aggregation. People often spend hours searching for relevant information online about their hobbies, interests, or passions. This project aims to address this challenge by leveraging UiPath's capabilities to develop a solution that automates the discovery and delivery of hobby-specific content, focusing on Reddit as the initial data source.

The core concept of this project revolves around extracting hobby-related information from Reddit discussions and delivering it directly to users through personalized emails. Reddit, being a hub of niche communities, provides an excellent platform for gathering diverse and rich content. Users can specify their hobbies, such as photography, cooking, or fitness, through a structured input mechanism like an Excel sheet. The workflow then dynamically searches Reddit for discussions, posts, or threads relevant to each user's hobbies, extracts the required information, and compiles it into a cohesive email format.

The implementation of this project relies on UiPath's robust set of tools and activities. Key components include Data Scraping to extract information from Reddit pages, Excel Automation to process user data, String Manipulation for formatting extracted content, and Email Automation to deliver personalized messages. The use of these UiPath activities ensures that the workflow is not only efficient but also modular and easy to expand in future iterations.

Although the current project focuses solely on Reddit, the design is intentionally scalable. Future versions could integrate other platforms, such as Amazon for product recommendations or YouTube for video content. By building a solid foundation with Reddit, the project demonstrates its potential to evolve into a comprehensive multi-platform content aggregation tool.

This project highlights several fundamental RPA concepts, such as dynamic data extraction, control flow, error handling, and modular workflow design. It also emphasizes the importance of user-centric automation solutions, showcasing how UiPath can bridge the gap between manual content discovery and automated personalization.

In essence, this solution not only simplifies the process of finding relevant information but also enhances the overall user experience. By delivering tailored content directly to users' inboxes, it saves time and effort, making it an invaluable tool for individuals passionate about their hobbies. This project stands as a testament to the power of RPA in addressing real-world problems through efficient and innovative solutions.

1.2 OBJECTIVE

The primary objective of this project is to develop an automated solution using UiPath that simplifies personalized hobby related content aggregation for users. By leveraging Reddit as the primary data source, the project aims to extract relevant hobby-specific information and deliver it directly to users via email. The workflow is designed to dynamically process user inputs, such as hobbies, and scrape data from corresponding Reddit threads, ensuring personalized and meaningful results. This project emphasizes the practical application of RPA concepts like data scraping, control flow, and email automation, while laying the foundation for future scalability to integrate multiple platforms.

1.3 EXISTING SYSTEM

The current approach to engaging with hobby-specific content on platforms like Reddit is manual and time-consuming. Users must individually browse forums, search for relevant posts, and filter through vast amounts of data to find meaningful content. There is no automated mechanism to consolidate or personalize this information for users, which can lead to inefficiencies and missed updates. Additionally, keeping track of trending discussions across multiple hobbies requires significant effort, making it challenging for users to stay informed and engaged with their interests.

1.4 PROPOSED SYSTEM

The proposed system introduces an automated solution using UiPath to streamline content aggregation and delivery for hobby enthusiasts. By leveraging Reddit as the primary platform, the bot dynamically processes user inputs, such as hobbies, from an Excel sheet and scrapes relevant data from corresponding Reddit threads. This includes trending posts, discussions, or updates tailored to each user's interests. The scraped data is formatted into a concise, user-friendly summary and sent via email.

The workflow utilizes key RPA concepts like data scraping, string manipulation, and email automation, ensuring an efficient and personalized experience. This system not only reduces the manual effort required to stay updated but also sets the stage for future expansion to additional platforms, enhancing its utility and scalability. Additionally, the proposed system lays the foundation for future enhancements, such as integrating more platforms like Amazon or dedicated forums, incorporating sentiment analysis to prioritize impactful content, and expanding notification channels beyond email. This ensures a robust, adaptable solution for hobby enthusiasts seeking timely and relevant updates.

CHAPTER 2

LITERATURE REVIEW

2.1 Robotic Process Automation (RPA) in Data Handling:

RPA has emerged as a transformative technology for automating repetitive tasks, including data collection, extraction, and processing. Studies by Willcocks et al. (2015) highlight RPA's ability to emulate human interactions with digital systems, thereby increasing efficiency and reducing errors in data-driven workflows. UiPath, a leading RPA platform, has been instrumental in enabling businesses to implement scalable automation solutions, particularly for web scraping and email communication.

Web Scraping for Information Retrieval:

Web scraping is a widely used technique to extract structured information from unstructured web data. Research by Aghajarian et al. (2018) emphasizes its utility in aggregating content from forums, e-commerce platforms, and social media sites. This project specifically focuses on Reddit as a source, leveraging its community-driven discussions for personalized information. Reddit’s open API and structured page layouts facilitate efficient data extraction, making it an ideal platform for targeted scraping activities.

2.2 Personalization in User-Centric Systems:

Personalized systems have proven effective in enhancing user engagement by tailoring content based on individual preferences. Research by Resnick and Varian (1997) on recommender systems underscores the importance of understanding user preferences to provide relevant information.

The integration of hobbies as a personalization parameter in this project exemplifies the growing trend of user-centric automation systems.

With the rise of big data and advanced analytics, personalization has become increasingly sophisticated, leveraging machine learning algorithms to analyze user behavior and preferences at scale. The project uses hobby-based input to create a highly targeted experience for users, making the information more relevant and engaging.

Such personalization not only improves user satisfaction but also demonstrates how automation can deliver meaningful insights tailored to individual needs..

2.3 Challenges in Automation and Ethical Considerations:

Despite its benefits, RPA and web scraping raise challenges related to data privacy, scalability, and ethical considerations. Studies by Mitchell (2019) highlight the importance of adhering to platform-specific terms of service and avoiding excessive server load when implementing scraping activities. This project adheres to ethical scraping practices by focusing on publicly accessible data without violating user privacy.

Additionally, implementing rate-limiting mechanisms and respecting website crawling guidelines is crucial to ensure compliance and avoid disruptions. The project includes measures to maintain ethical standards, such as scraping data at controlled intervals and providing transparent processes for users to understand the data sources. These practices ensure the project aligns with current debates around responsible data use and automation ethics.

:

2.4 Future Implications:

The project paves the way for future enhancements, such as integrating multiple platforms and utilizing advanced Natural Language Processing (NLP) techniques for content summarization. By addressing the identified challenges and leveraging advancements in RPA and personalization technologies, this project demonstrates the potential of automation in delivering value to end-users.

Looking ahead, the system could be extended to include multi-lingual support and sentiment analysis for a broader scope of user engagement. These advancements would enable the system to cater to global audiences and extract insights with deeper contextual relevance. Furthermore, integrating visualization tools to summarize scraped data could improve user accessibility and foster greater understanding of the retrieved information. This highlights the immense potential for scaling the system beyond its current scope.

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. The system flow diagram for this project is in Fig. 3.1.

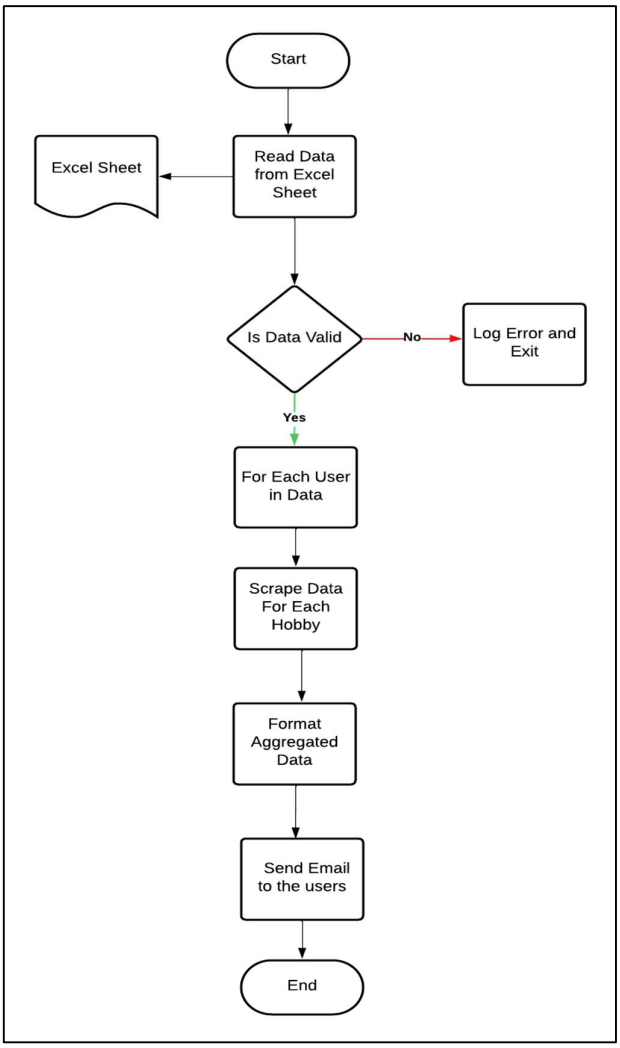


Fig 3.1 System Flow Diagram

3.1 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.

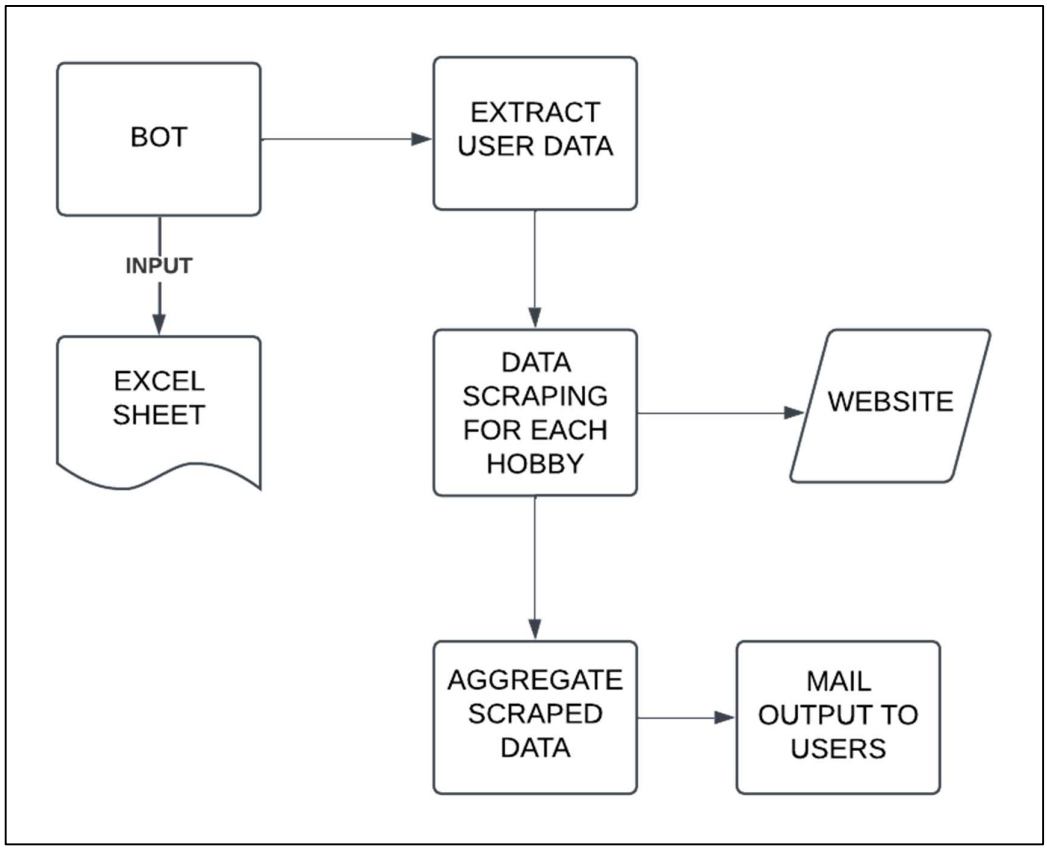


Fig 3.2 Architecture Diagram

3.2 SEQUENCE DIAGRAM

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

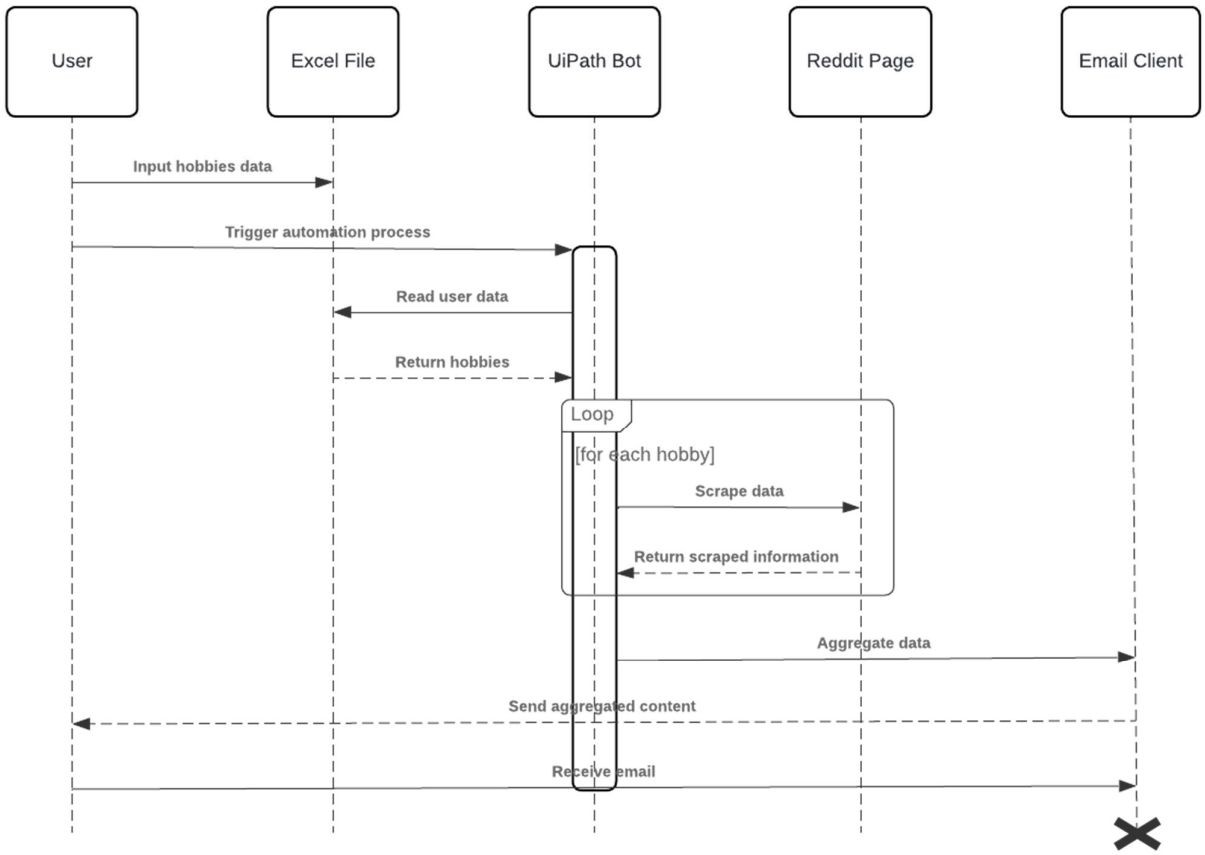


Fig 3.3 Sequence Diagram

CHAPTER 4

PROJECT DESCRIPTION

"The Smart Assignment Integrity Verification Bot" is a sophisticated Robotic Process Automation (RPA) project designed to address the challenges of AI-generated content and plagiarism in student assignments. Developed using UiPath, this intelligent bot streamlines the assignment assessment process, providing educators with an efficient tool to maintain academic integrity.

4.1. MODULES:

4.1.1. INPUT HANDLING AND INITIALIZATION:

4.1.1.1. Data Input:

- Read user data from an Excel file, including hobbies, platforms, and associated links.
- Process and categorize user information for individualized operations.

4.1.1.2. Hobby Filtering:

- Split hobbies into individual items for processing.
- Validate input data to ensure consistent and complete operation.

4.1.2 DATA EXTRACTION AND AGGREGATION:

4.1.2.1 Reddit Scraping:

- Open Reddit pages corresponding to each user's hobby link.

- Scrape relevant data such as trending posts or discussions for the selected hobby.

4.1.2.2 Data Aggregation:

- Combine all extracted data into a structured format.
- Append hobby-specific insights to a consolidated message for each user.

4.1.3 RESULT MANAGEMENT:

4.1.3.1 Data Formatting:

- Format extracted data into a user-friendly and readable format.

4.1.3.2 Email Preparation:

- Prepare email content with extracted data and ensure hobby-specific insights are presented clearly and concisely.

4.1.4 COMPLETION AND REPORTING:

4.1.4.1 Email Dispatch:

- Automatically send emails containing personalized insights to each user. Use a reliable email client for smooth communication.

CHAPTER 5

OUTPUT SCREENSHOTS

Fig 5.1 – Input

The bot gets the excel file from the project directory as shown in Fig 5.1.

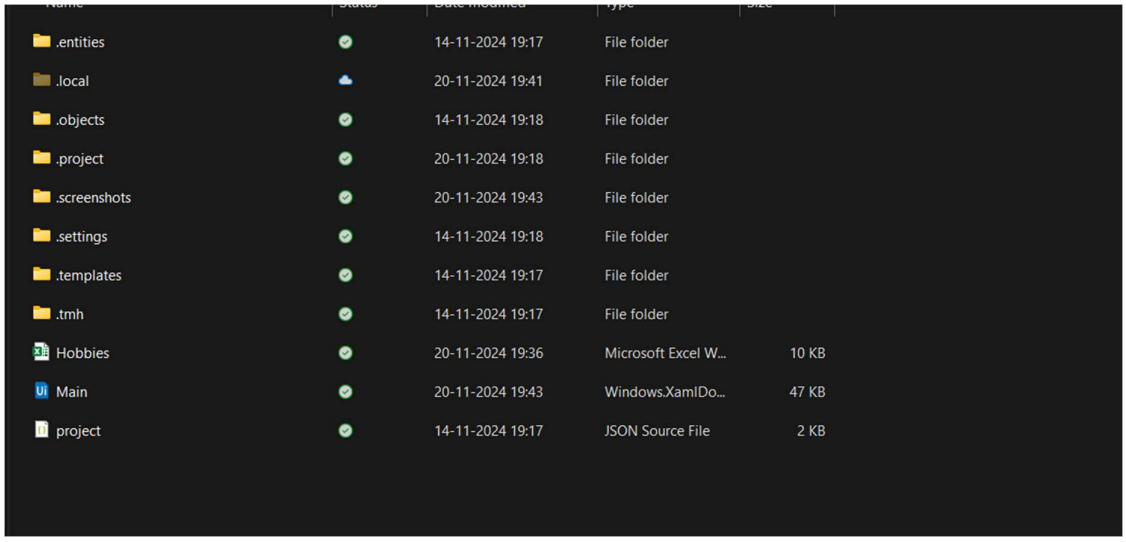
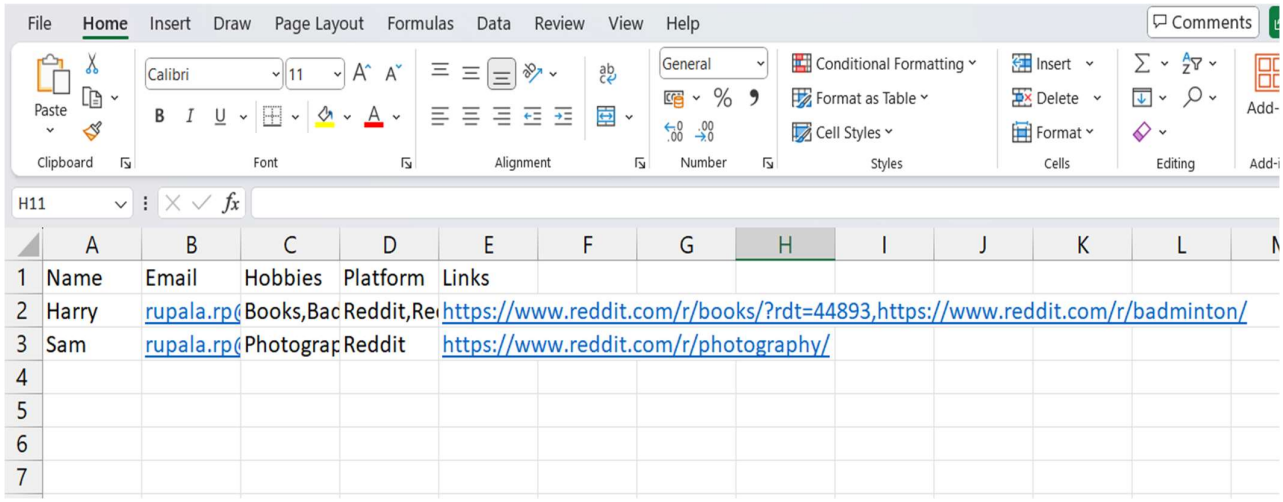


Fig 5.2 – Excel File Creation

An excel file report is created in the main directory for providing the details as shown in Fig 5.2.



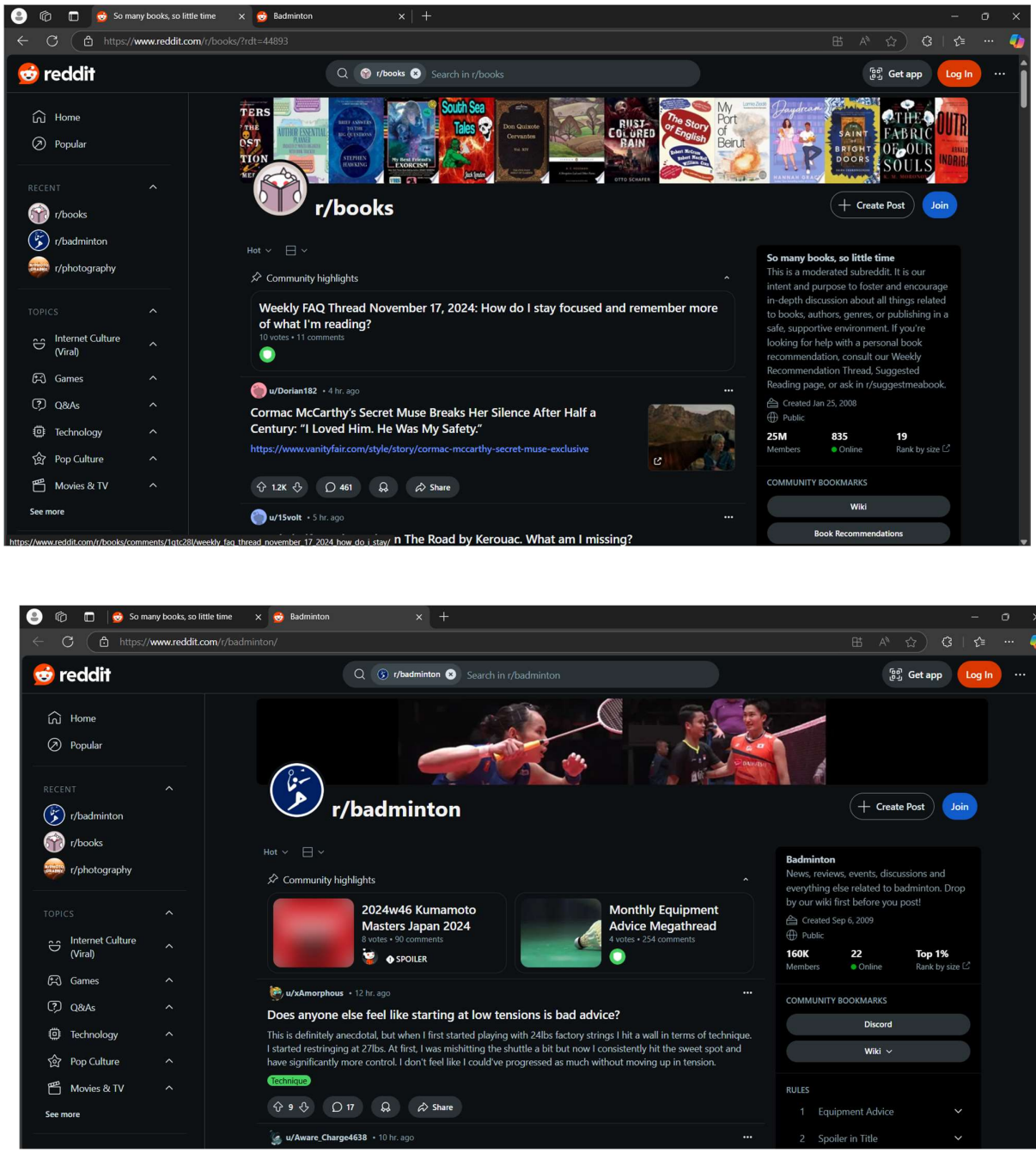


Fig 5.3 – Web Scraping

Fig 5.4 – Web Scraping

Hello,

Here is the information we gathered for you based on your preferences:

Hobby: Books
We found some engaging discussions on Reddit:

- *"Top 5 Fiction Books to Read This Winter"*
  - "Discover the most captivating fiction books to cozy up with this season."
  - *Top Comment:* "Loved the second recommendation! It's a must-read."
  - Read More
- *"How to Start a Book Club with Friends"*
  - "Tips and tricks for starting a fun and engaging book club."
  - *Top Comment:* "This guide made organizing so easy!"
  - Read More

Hobby: Badminton
We explored Reddit pages for badminton enthusiasts:

- *"The Best Badminton Rackets of 2024"*
  - "An in-depth review of this year's top-performing rackets."
  - *Top Comment:* "The lightweight option is great for beginners!"
  - Read More
- *"How to Improve Your Backhand Smash"*
  - "Proven techniques and drills to perfect your smash."
  - *Top Comment:* "This tip improved my game significantly!"
  - Read More

Thank you for using our service!

Fig 5.5 – Sample Email

The results are then forwarded to the mail of the users collected from the excel sheet as it is shown in Fig 5.5.

CHAPTER 6

CONCLUSION

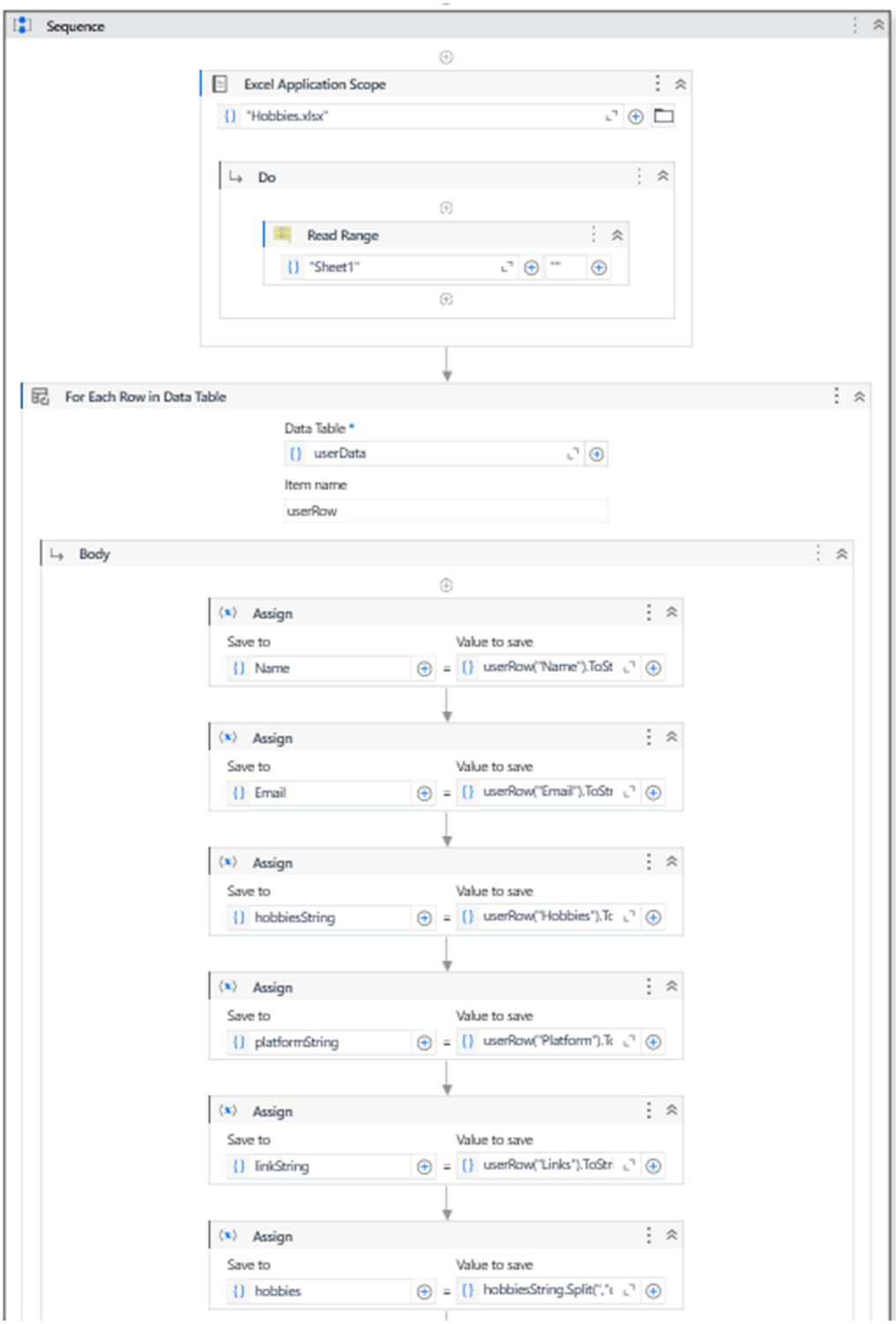
The Automated Hobby-Based Reddit Data Aggregator exemplifies the transformative potential of Robotic Process Automation (RPA) in simplifying complex workflows and providing personalized user experiences. By leveraging UiPath's robust capabilities, this project streamlines the process of hobby-specific data collection and aggregation, addressing the growing need for customized insights in a digital age.

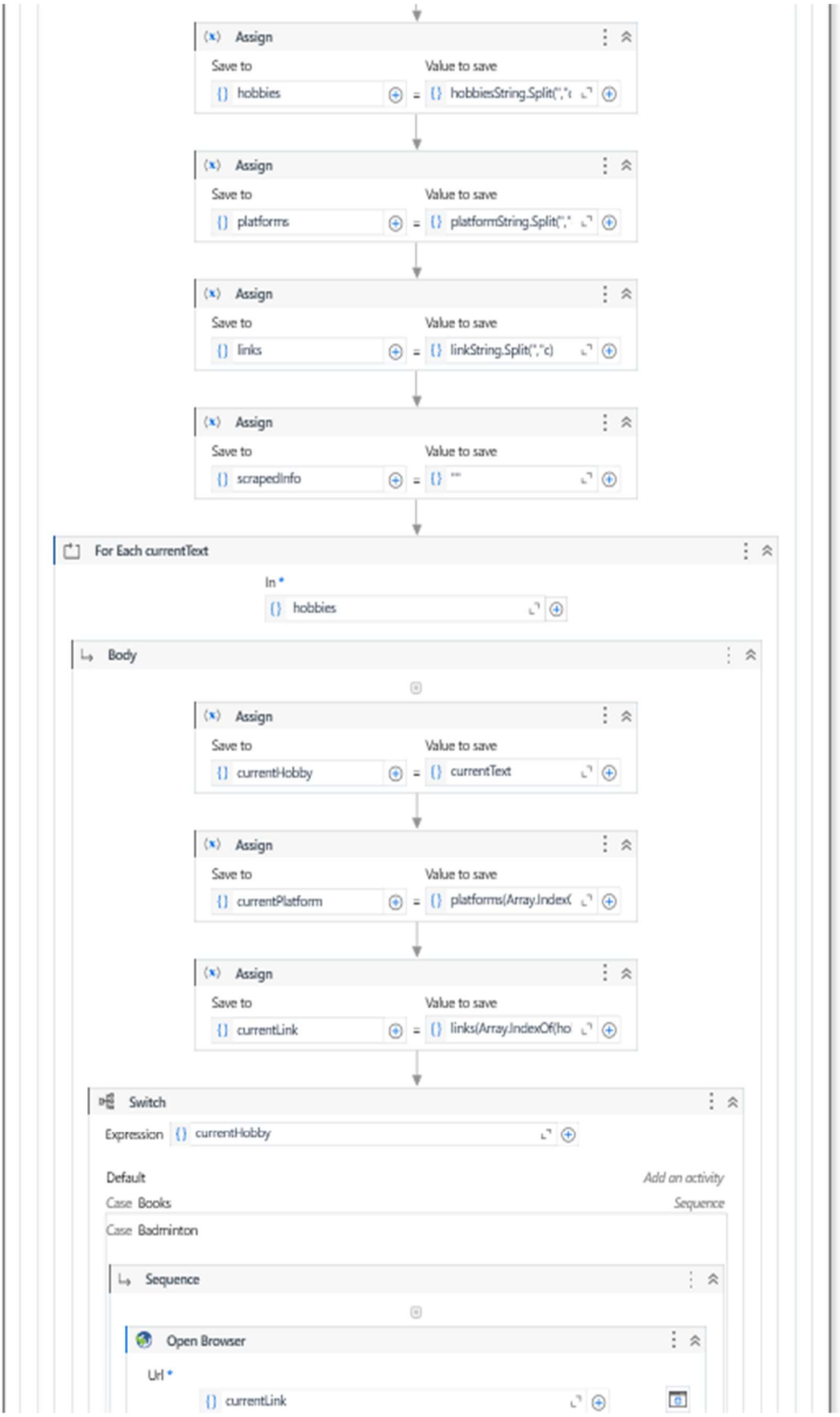
This system efficiently reads user data, including hobbies and links, from an Excel file and processes it to fetch relevant content from Reddit. By automating the data extraction, aggregation, and communication steps, the project not only reduces manual effort but also ensures a consistent and error-free workflow. The ability to dynamically adapt to user-specific input ensures the system's versatility, making it scalable for future integrations with additional platforms and data sources.

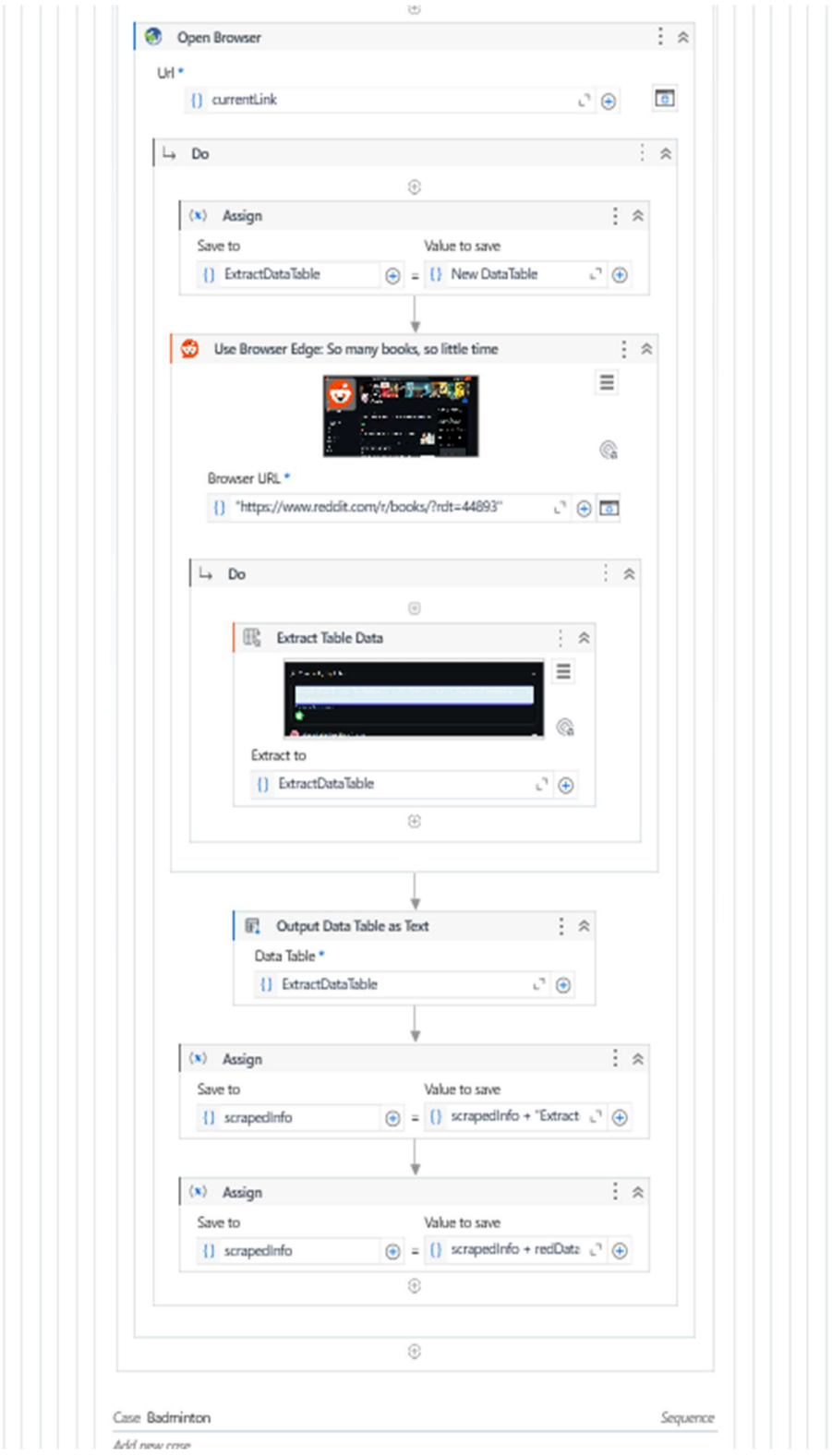
In conclusion, this project serves as a valuable example of how automation can bridge the gap between vast online data and personalized user needs. While its current focus is on Reddit, the potential for expansion and integration with other platforms makes it a promising foundation for future development. With its user-centric design and efficient automation, this project reaffirms the significance of RPA in enhancing productivity and personalization.

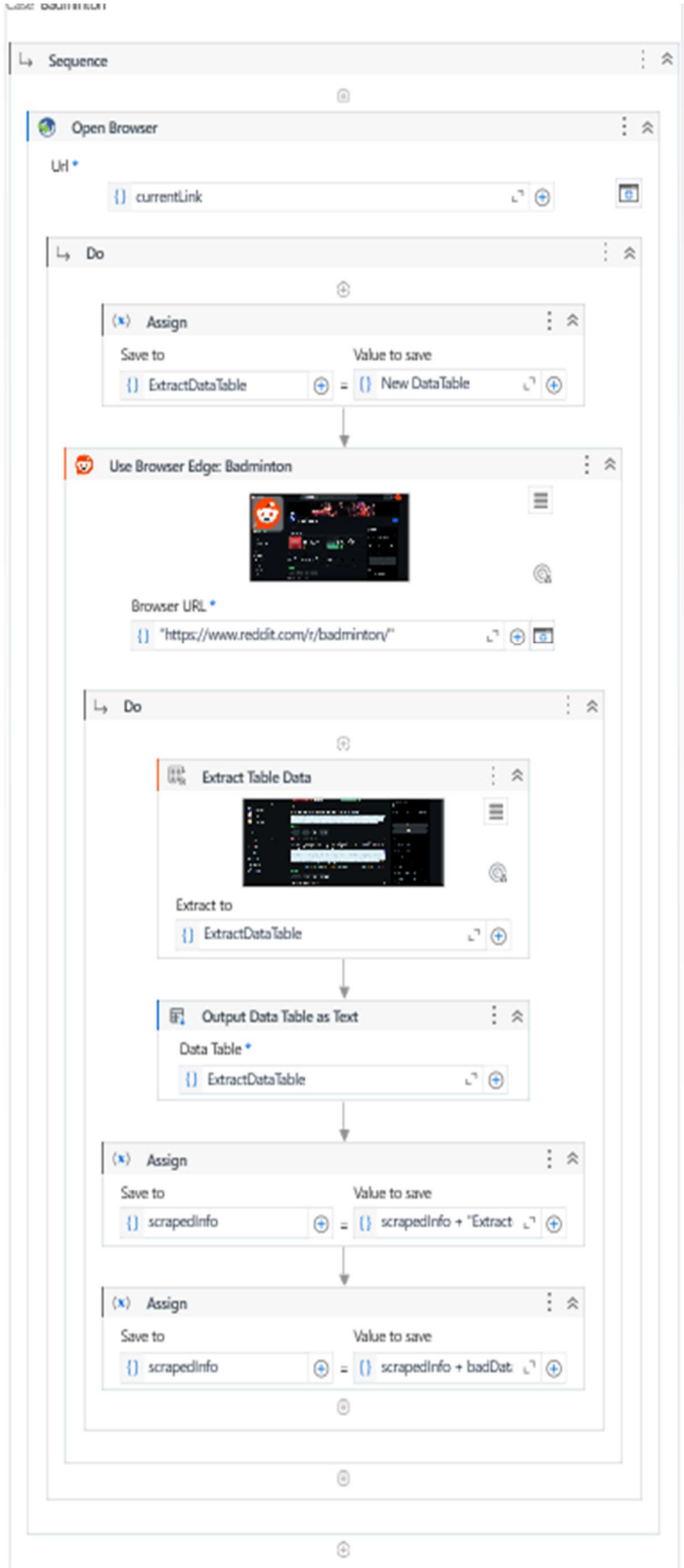
APPENDIX

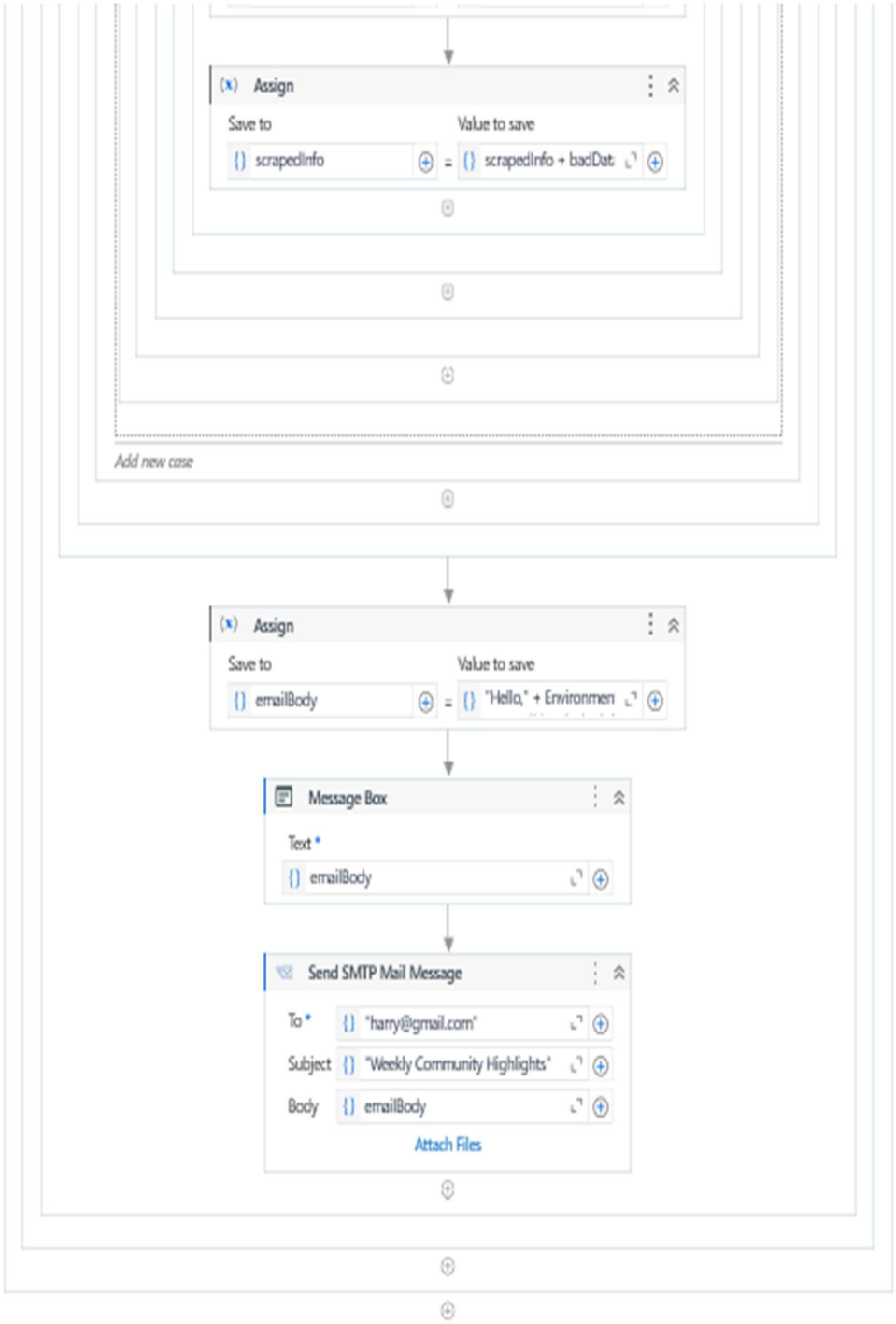
PROCESS WORK FLOW

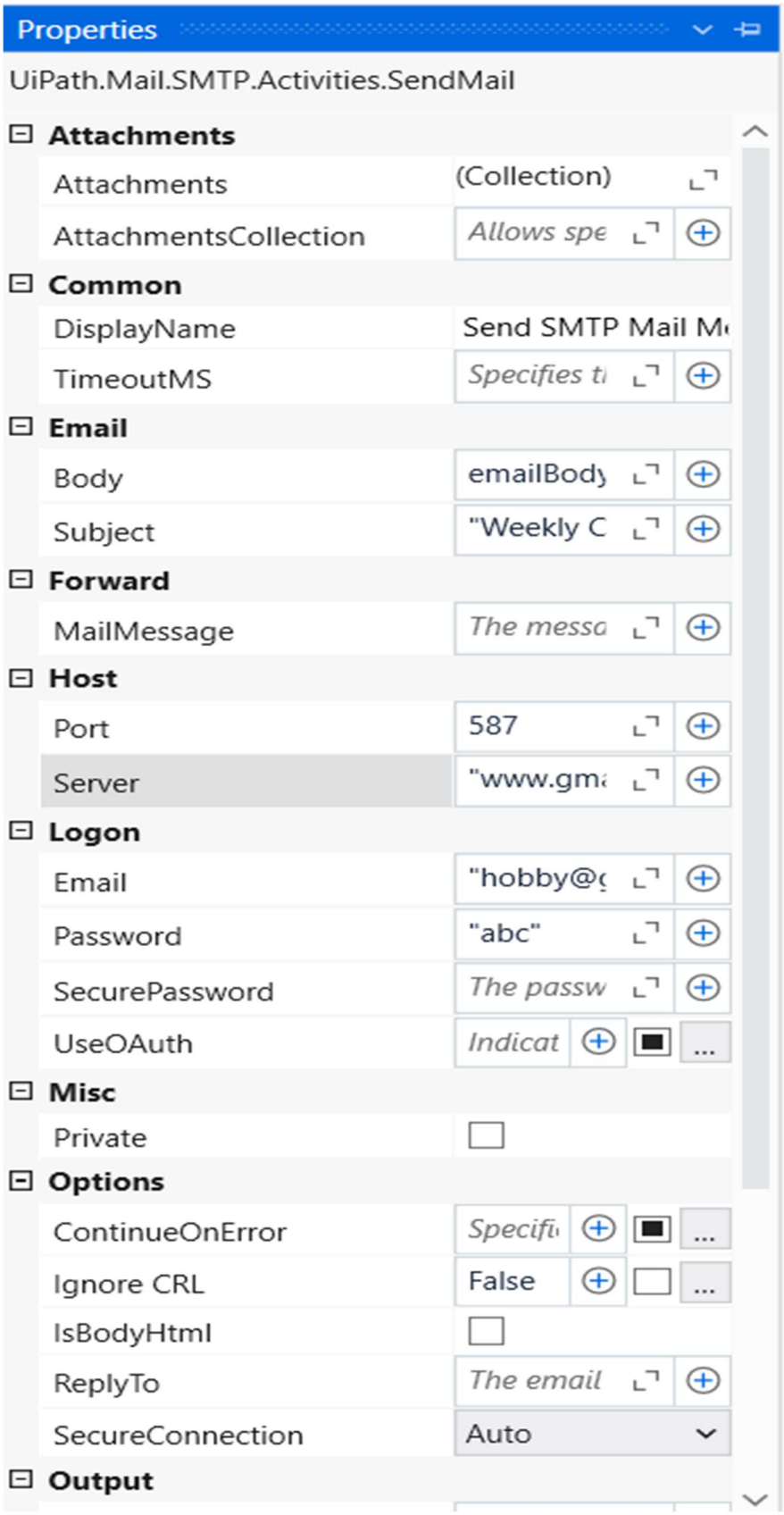



Assign
Save to
Value to save
hobbies
hobbiesString.Split(",
Assign
Save to
Value to save
platforms
platformString.Split(",
Assign
Save to
Value to save
links
linkString.Split(",'c)
Assign
Save to
Value to save
scrapedInfo
""
For Each currentText
In *
hobbies
Body
Assign
Save to
Value to save
currentHobby
currentText
Assign
Save to
Value to save
currentPlatform
platforms(Array.Index(
Assign
Save to
Value to save
currentLink
links(Array.IndexOf(ho
Switch
Expression
currentHobby
Default
Add an activity
Case Books
Sequence
Case Badminton
Sequence
Open Browser
Url *
currentLink


Open Browser
Url *
currentLink
Do
Assign
Save to
Value to save
ExtractDataTable
New DataTable
Use Browser Edge: So many books, so little time
Browser URL *
"https://www.reddit.com/r/books/?rdt=44893"
Do
Extract Table Data
Extract to
ExtractDataTable
Output Data Table as Text
Data Table *
ExtractDataTable
Assign
Save to
Value to save
scrapedInfo
scrapedInfo + "Extract
Assign
Save to
Value to save
scrapedInfo
scrapedInfo + redData
Case Badminton
Sequence


Sequence
Open Browser
Url *
currentLink
Do
Assign
Save to
Value to save
ExtractDataTable
=
New DataTable
Use Browser Edge: Badminton
Browser URL *
"https://www.reddit.com/r/badminton/"
Do
Extract Table Data
Extract to
ExtractDataTable
Output Data Table as Text
Data Table *
ExtractDataTable
Assign
Save to
Value to save
scrapedInfo
=
scrapedInfo + "Extract
Assign
Save to
Value to save
scrapedInfo
=
scrapedInfo + badDat


Assign
Save to
Value to save
scrapedInfo
scrapedInfo + badDat
Add new case
Assign
Save to
Value to save
emailBody
"Hello," + Environmen
Message Box
Text *
emailBody
Send SMTP Mail Message
To *
"harry@gmail.com"
Subject
"Weekly Community Highlights"
Body
emailBody
Attach Files


Properties
UiPath.Mail.SMTP.Activities.SendMail
Attachments
Attachments
(Collection)
AttachmentsCollection
Allows spe
Common
DisplayName
Send SMTP Mail M
TimeoutMS
Specifies ti
Email
Body
emailBody
Subject
"Weekly C
Forward
MailMessage
The messa
Host
Port
587
Server
"www.gm
Logon
Email
"hobby@
Password
"abc"
SecurePassword
The passw
UseOAuth
Indicat
Misc
Private
Options
ContinueOnError
Specifi
Ignore CRL
False
IsBodyHtml
ReplyTo
The email
SecureConnection
Auto
Output

REFERENCES

The following resources were instrumental in guiding the development and implementation of this project. They provide foundational knowledge and practical insights into the technologies and methodologies employed:

1. UiPath Documentation: Email Activities Guide

- This documentation offered detailed instructions on configuring and using UiPath's email activities, such as fetching emails via IMAP and dispatching them via SMTP. It provided step-by-step guidance on setting up these functionalities securely and efficiently, forming the backbone of the email integration module.

2. OpenAI API Documentation: ChatGPT API Usage and Features

- The OpenAI API documentation served as a key reference for understanding the capabilities of the GPT models, including text generation, response customization, and handling API requests. It also provided essential details about authorization, payload structure, and error handling, enabling seamless integration with UiPath workflows.

3. Literature on RPA and AI Integration

- Various research papers, online articles, and case studies on RPA and AI integration offered insights into best practices, potential challenges, and real-world applications. These sources helped in designing a system that combines the strengths of both technologies for maximum efficiency and scalability