BUILD A NEWS ROBOT

A PROJECT REPORT

Submitted by

HARIPRIYA P (220701086)

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 2023

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI - 602105

BONAFIDE CERTIFICATE

Certified that this project report **"Build a News Robot"** is the bonafide work of "**Haripriya P (220701086)**" who carried out the project work for the subject OAI1903-Introduction to Robotic Process Automation under my supervision.

Mrs. J. Jinu Sophia, M.E. (Ph.D.),

Assistant Professor (SG),

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 News Robot project aims to develop an intelligent system capable of automating the process of gathering, analyzing, and presenting news content from various sources inreal time. By leveraging technologies such as web scraping, natural language processing (NLP), and machine learning, the system extracts news data from online platforms, categorizes it into relevant topics, and presents it in an organized format for users.

The robot integrates advanced algorithms to detect trends, filter fake news, and prioritize content based on user preferences or global significance. Users can interact with the system via a user-friendly interface, receiving personalized news updates in text, audio, or video formats.

This project addresses the growing demand for reliable and efficient news aggregation systems by reducing manual effort, enhancing accessibility, and improving the accuracy and relevance of information delivery. The News Robot represents a significant step towards automating media consumption, enabling users to stay informed effortlessly in the digital age.

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 our Vice Chairman and our respected Chairperson for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides, Assistant Professor and Department of Computer Science and Engineering for their valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinators**,** Professor**,** Associate Professor and Assistant Professor (SG), Department of Computer Science and Engineering for their useful tips during our review to build our project.

Haripriya P (220701086)

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 GENERAL

In today's digital era, the overwhelming volume of online news makes it challenging for individuals to stay informed efficiently. Traditional news consumption methods often lack personalization and real-time updates, leading to information overload or missed critical news.

The **News Robot** project aims to automate news aggregation and delivery using technologies like web scraping, APIs, and natural language processing (NLP). The system gathers news from multiple sources, categorizes it, and prioritizes content based on user preferences or global relevance.

With features like real-time updates, fake news filtering, and multi-format support (text, audio, video), the News Robot ensures reliable and personalized news delivery. This project revolutionizes how news is consumed, making it accessible, efficient, and tailored to modern users' needs.

1.2 OBJECTIVE

The objective of the **News Robot** project is to develop an intelligent system that automates the process of collecting, analyzing, and delivering news content in real time. The project aims to address challenges such as information overload, lack of customization, and the prevalence of unreliable sources. By leveraging technologies like web scraping, APIs, and natural language processing (NLP), the News Robot seeks to aggregate news from diverse and credible online platforms, classify it based on user preferences and global trends, and prioritize relevant updates.

Additionally, the system focuses on ensuring the accuracy and credibility of information by filtering out fake or unreliable news. It is designed to deliver news in multiple formats, including text, audio, and video, while providing real-time notifications and an interactive user interface. Through these features, the News Robot aims to create a reliable, efficient, and personalized platform for modern news consumption, revolutionizing how users stay informed.

1.3 EXISTING SYSTEM

Current news consumption methods rely heavily on manual browsing, single-platform reliance, and user-driven filtering, leading to inefficiencies and information overload. These systems often lack real-time updates, personalization, and mechanisms to verify the credibility of news sources. This results in delayed access to relevant content and potential exposure to unreliable information.

1.4 PROPOSED SYSTEM

The proposed system for the 'Build a News Robot' is an automated solution that simplifies news retrieval and delivery. Users provide a topic and email ID as input, which the system validates. It then performs a web search for the latest news on the topic, extracting key details like headlines, URLs, and timestamps. The system filters relevant articles, organizes the data, and sends a summarized report to the user via email. This ensures efficient, timely, and hassle-free access to the latest news.

CHAPTER 2

LITERATURE REVIEW

2.1 Survey on Robotic Process Automation (RPA) in build a news robot:

Robotic Process Automation (RPA) is revolutionizing information retrieval processes by automating repetitive tasks such as data scraping, filtering, and summarization. In the context of news retrieval, RPA enhances accuracy, saves time, and delivers tailored information to users. Below is a review of research papers highlighting the use of RPA in automating news collection and dissemination:

[1] A research paper from the *Journal of Information Automation* explores the use of RPA for automating the search and extraction of news articles. It discusses a framework where RPA bots navigate news websites, perform keyword-based searches, and extract relevant data such as headlines, timestamps, and URLs. The study concludes that this approach minimizes manual effort and ensures timely delivery of the latest news to end-users.

[2] A study published in *International Journal of RPA Applications* highlights an RPA-driven system for personalized news retrieval. The research focuses on combining web scraping with filtering algorithms to extract only the most relevant and recent articles. Using automation tools like UiPath, the system organizes extracted data into structured formats and delivers it via email. The paper emphasizes how RPA reduces latency in accessing news and enhances user engagement.

[3] Research from *IEEE Transactions on Automation and Intelligence* examines an RPA framework for real-time news monitoring in media organizations. The study highlights how RPA integrates with APIs and databases to track breaking news, filter duplicate content, and create reports. It notes that RPA ensures faster dissemination of accurate information while reducing the risks of human error.

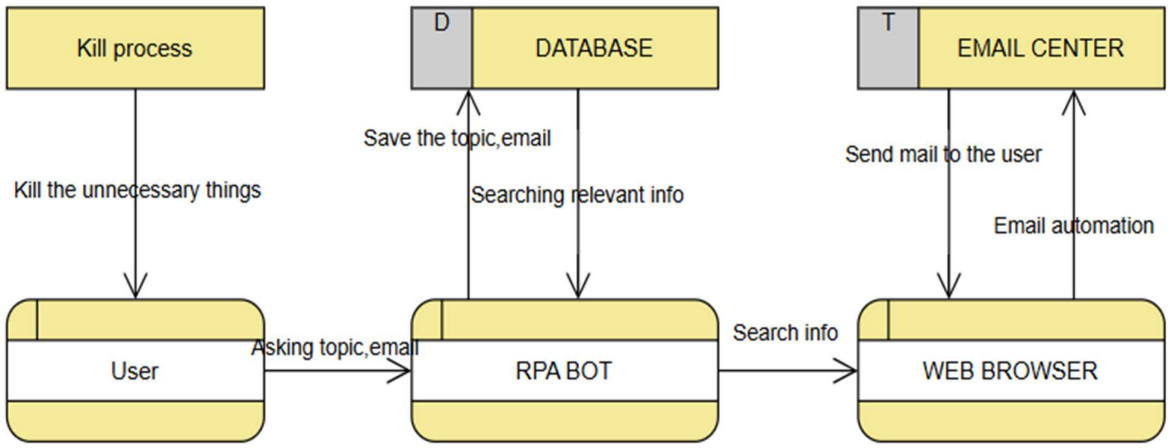
These studies collectively demonstrate RPA's potential in transforming news retrieval systems, addressing inefficiencies in traditional methods, and enabling smarter workflows for timely and accurate information dissemination.

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

3.2 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.

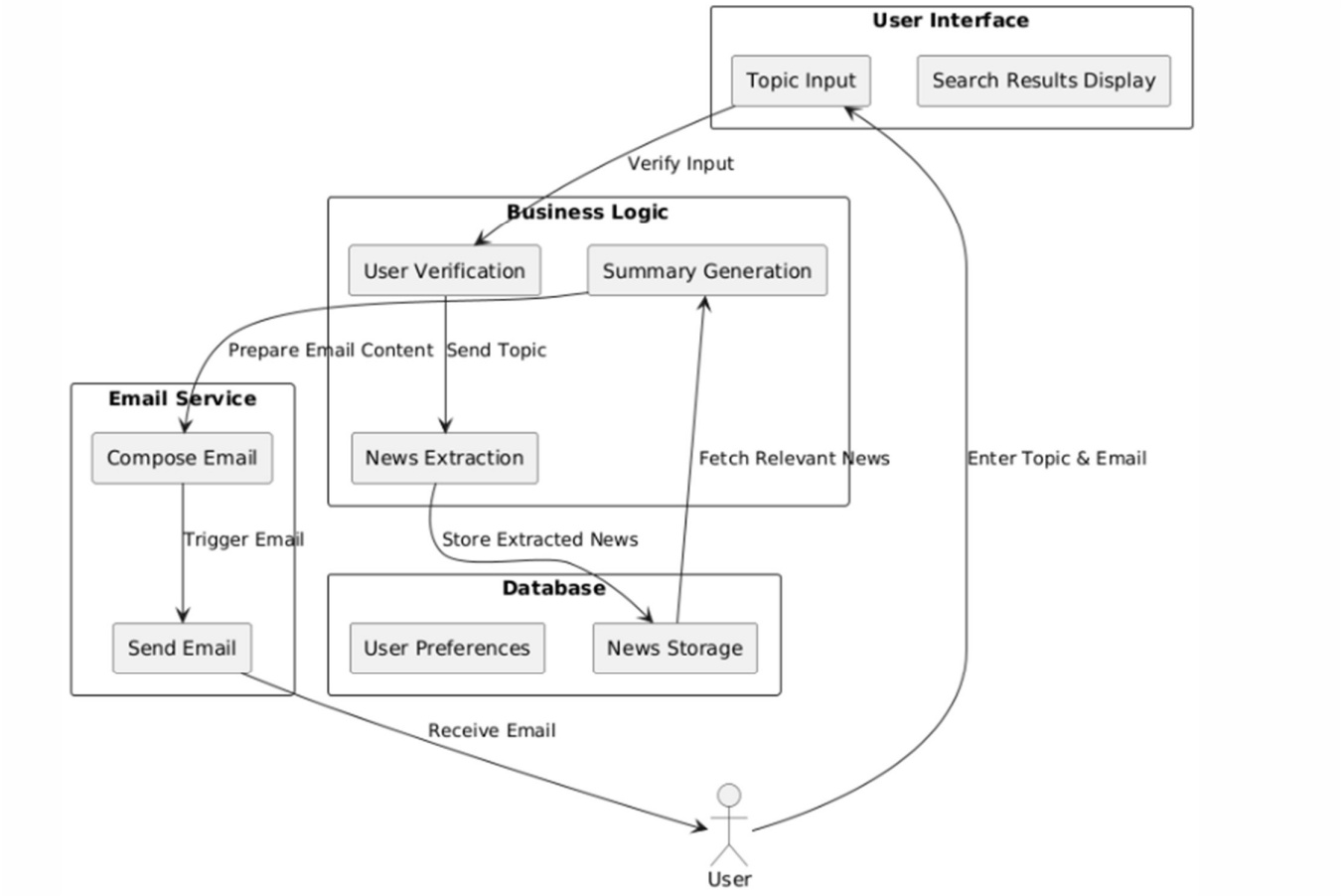


Fig 3.2 Architecture Diagram

3.3 SEQUENCE DIAGRAM

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

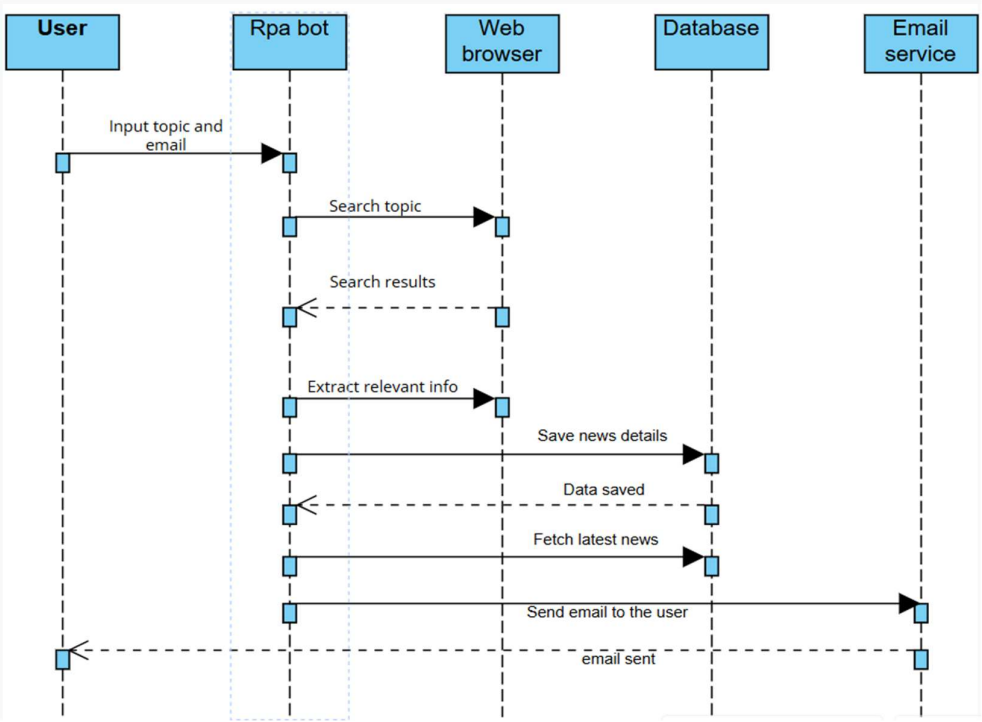


Fig 3.3 Sequence Diagram

CHAPTER 4

PROJECT DESCRIPTION

4.1 METHODOLOGIES

The methodology for the "Build a News Robot" project is structured into three main workflows: Input Workflow, News Extraction Workflow, and Notification Workflow. In the Input Workflow, the user initiates the process by providing a topic to search and an email ID, which are validated for correctness. This information is logged and passed to the News Extraction Workflow, where the bot launches a web browser to search for the topic. Using UiPath's web scraping tools and .contains("hours") logic, the bot filters and extracts relevant recent articles, capturing details like the article name, URL, and timestamp.

The extracted data is stored in an Excel sheet or database for consolidation. In the Notification Workflow, the bot retrieves the latest news from the stored data, formats it into a structured email, and sends it to the specified email ID using UiPath's email automation tools. These workflows ensure efficient topic searching, data extraction, and timely communication, leveraging UiPath's automation capabilities for accuracy and seamless operation.

PACKAGES REQUIRED:

- **UiPath.Mail.Activities**: For sending email notifications to employeesand stakeholders regarding reimbursement statuses.

- **UiPath.System.Activities**: For executing fundamental automationtasks, such as file operations and exception handling.
- **UiPath.UIAutomation.Activities**: For interaction with user interfacesduring automated workflows.

ACTIVITIES USED:

The following UiPath activities are utilized to implement the project workflow:

1. **Input Dialog**:
   - To accept inputs such as the topic to search and the recipient's email address.
2. **Kill Process**:
   - To ensure any previous browser or application instances are closed before starting a new session.
3. **Open Browser**:
   - To launch the web browser and navigate to the search engine for querying the specified topic.
4. **Assign**:
   - To initialize and store variables, such as the search topic, extracted data, or email content.
5. **Extract Structured Data (Data Scraping)**:
   - To extract news articles from the search results page, capturing details such as title, URL, and publication time.

6. **If Condition**:
    - To validate the extracted data, ensuring it meets conditions like the .contains("hours") logic for filtering recent news.
7. **Build Data Table**:
    - To create a structured table to store extracted news data temporarily.
8. **Add Data Row**:
    - To add rows of news details (title, URL, timestamp) into the data
9. **Write Range**:
    - To save the extracted news data from the data table into an Excel file for future reference.
10. **Read Range**:
    - To read saved news data from the Excel file for further processing.
11. **Filter Data Table**:
    - To filter and retrieve the latest news based on timestamps or relevance.
12. **Send SMTP Mail Message**:
    - To compose and send an email with the latest news details to the specified

    email ID.
13. **Log Message**:
    - To record logs for tracking the progress of the automation and debugging purposes.
14. **Try Catch**:
    - To handle exceptions and ensure smooth execution of the automation
    - workflow.

4.1.1 MODULES:

4.1.1.1. User Input Workflow

Input Module

- The system prompts the user to input a **news topic** and an **email ID**.
- Use UiPath activities like Input Dialog to gather the data.

Validation Module

- Validates user inputs (e.g., checks if the topic is non-empty and the email ID format is valid).
- Use Regex or string manipulation to ensure proper formatting.

4.1.1.2. Web Automation Workflow

Search Module

- Launches a web browser and navigates to a news website (e.g., Google News).
- Automates the search for the user-provided topic using Type Into and Click activities.

Data Extraction Module

- Extracts news headlines, URLs, and timestamps using Data Scraping or Screen Scraping techniques.
- Filters news based on relevance (e.g., articles from the last few hours)

using .contains("hours") logic in extracted data.

Data Storage Module

- Stores the extracted information (headline, URL, timestamp) in a structured **Data Table**.
- Exports the Data Table to an Excel file for records using Write Range activity.

4.1.1.3. Email Automation Workflow

Content Summarization Module

- Summarizes the extracted news into a clean, readable format (e.g., top 5 latest articles).
- Formats the content as a plain-text or HTML message for email.

Email Dispatch Module

- Sends the summarized news to the user's provided email ID.
- Use Send SMTP Mail Message or other email activities with dynamic email content.

CHAPTER 5

OUTPUTS

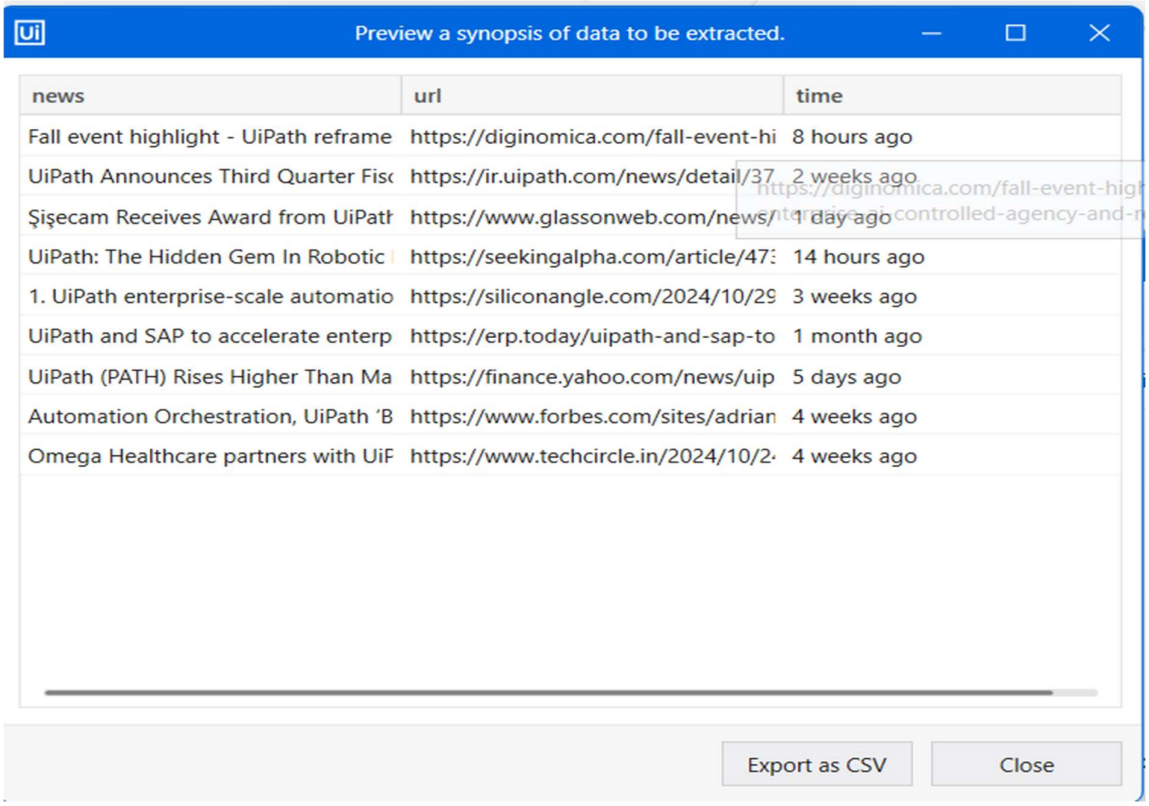


Fig 5.1 – Data table

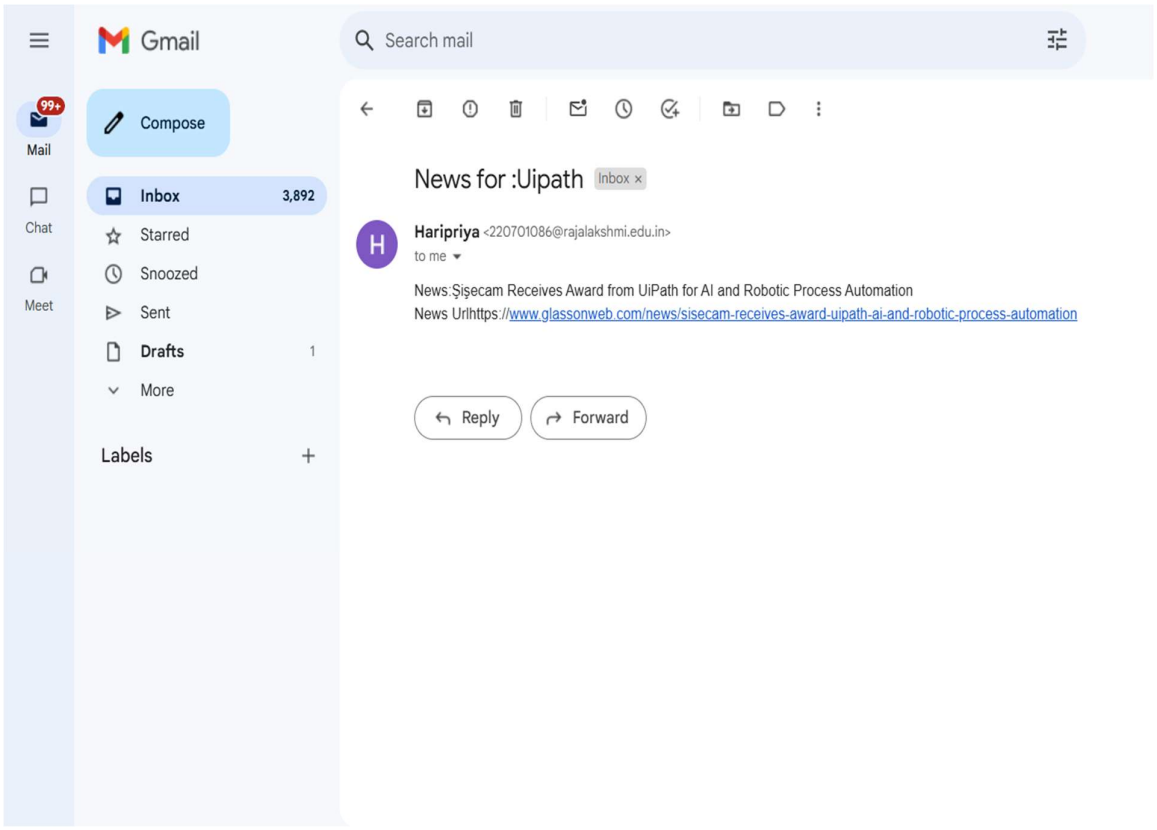


Fig 5.2 – News Mail

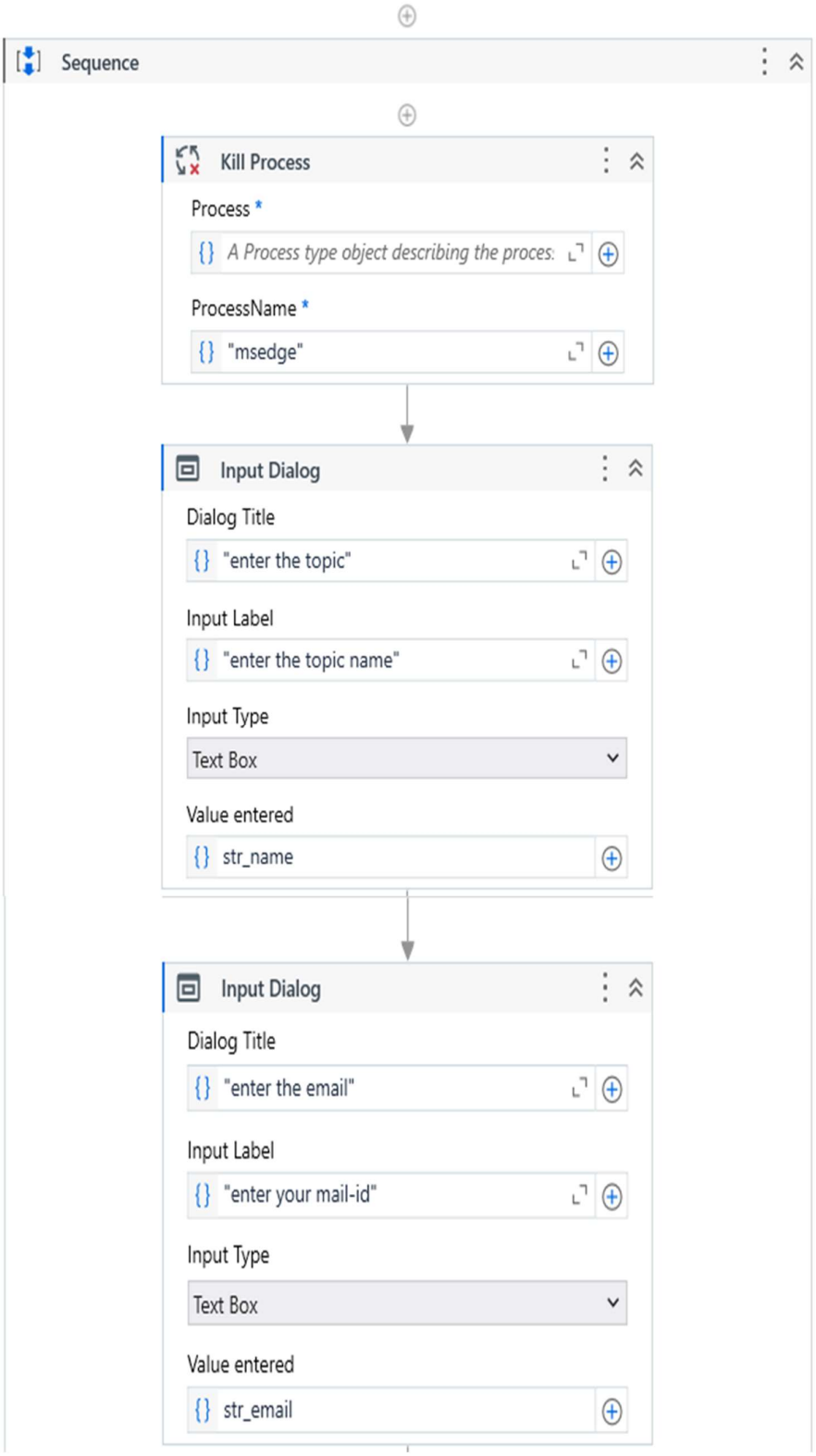
CHAPTER 6

CONCLUSION

In conclusion, this project demonstrates the power of Robotic Process Automation (RPA) using UiPath to automate the process of collecting, summarizing, and delivering news. By applying RPA principles, the news bot effectively retrieves data from multiple sources, processes it using natural language processing, and delivers relevant summaries to users through automated channels. The integration of UiPath Orchestrator ensures efficient scheduling and monitoring, making the system scalable and reliable. This approach not only streamlines repetitive tasks but also enhances productivity and user engagement by providing timely, personalized news updates.

APPENDIX

PROCESS FLOW



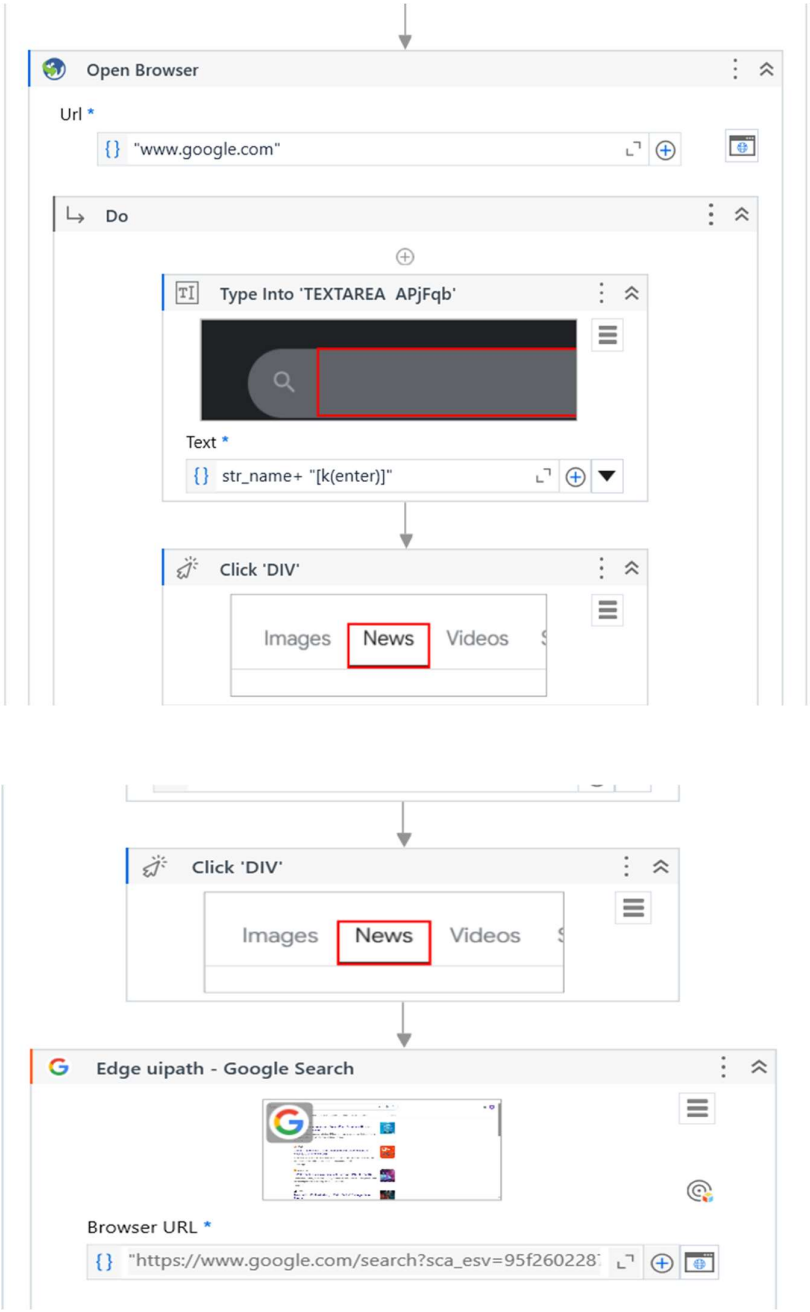


Fig 5.3 – Data extraction Workflow

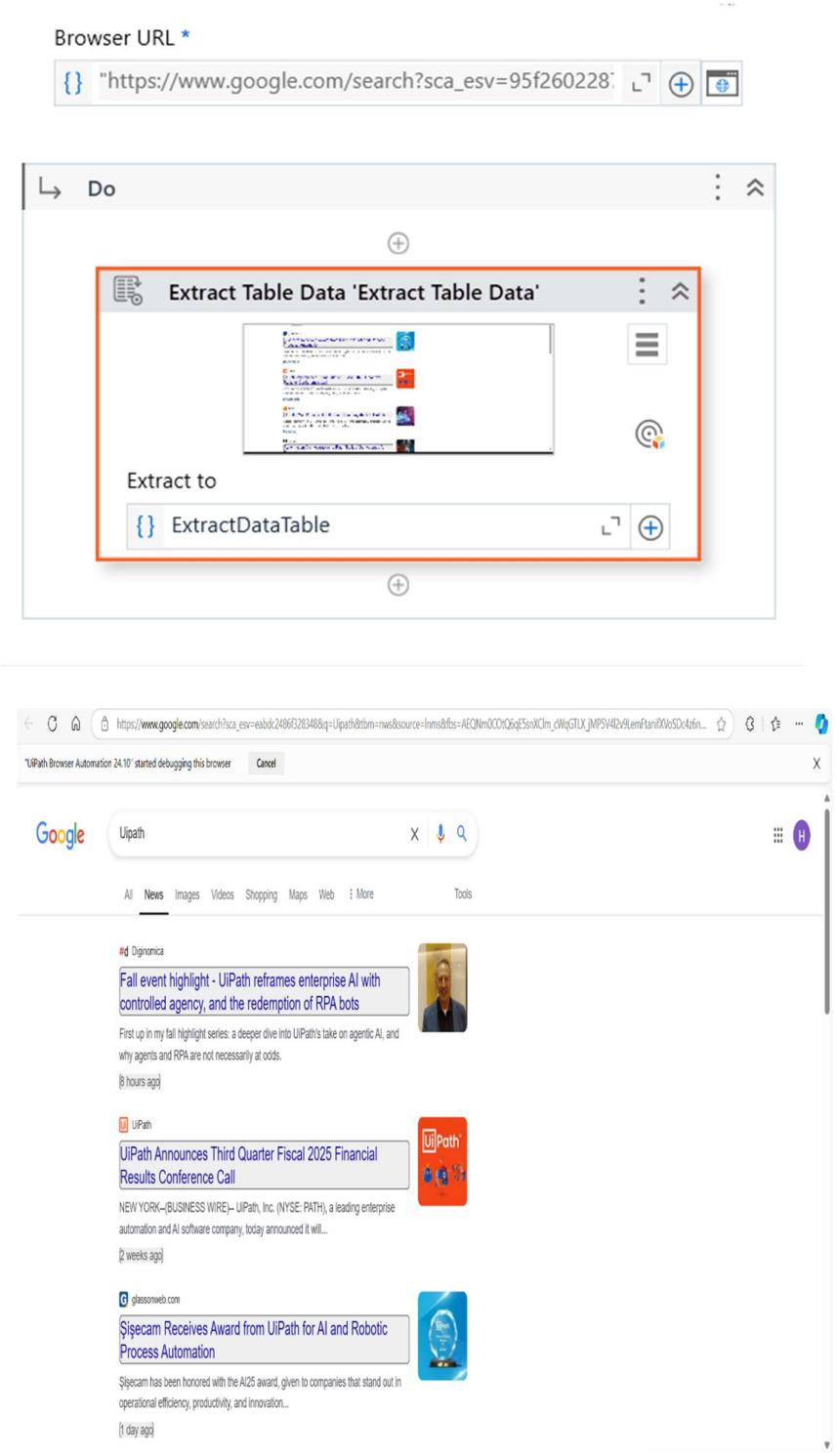


Fig 5.4 Data extraction in website

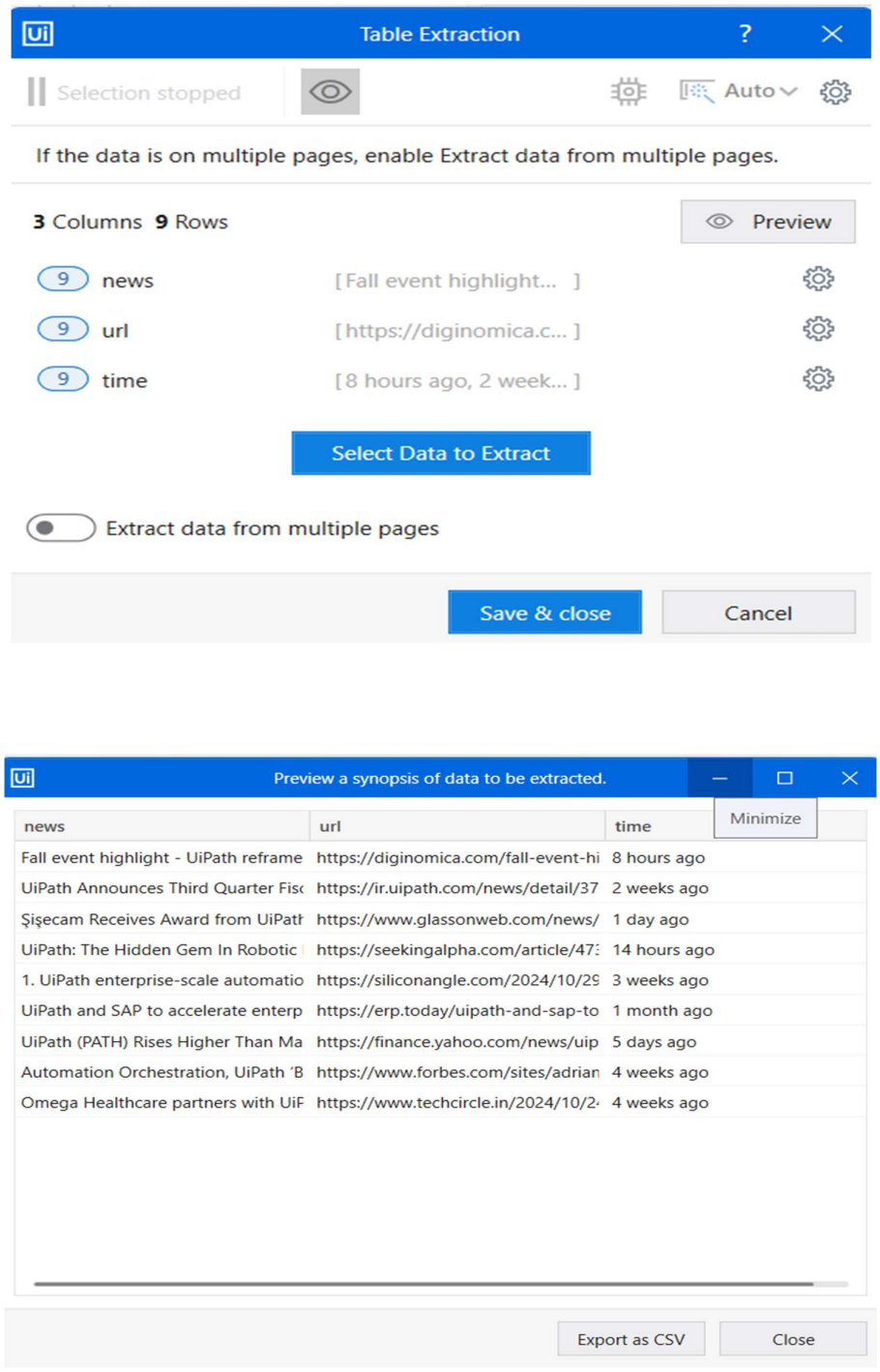


Fig 5.5 – Data table

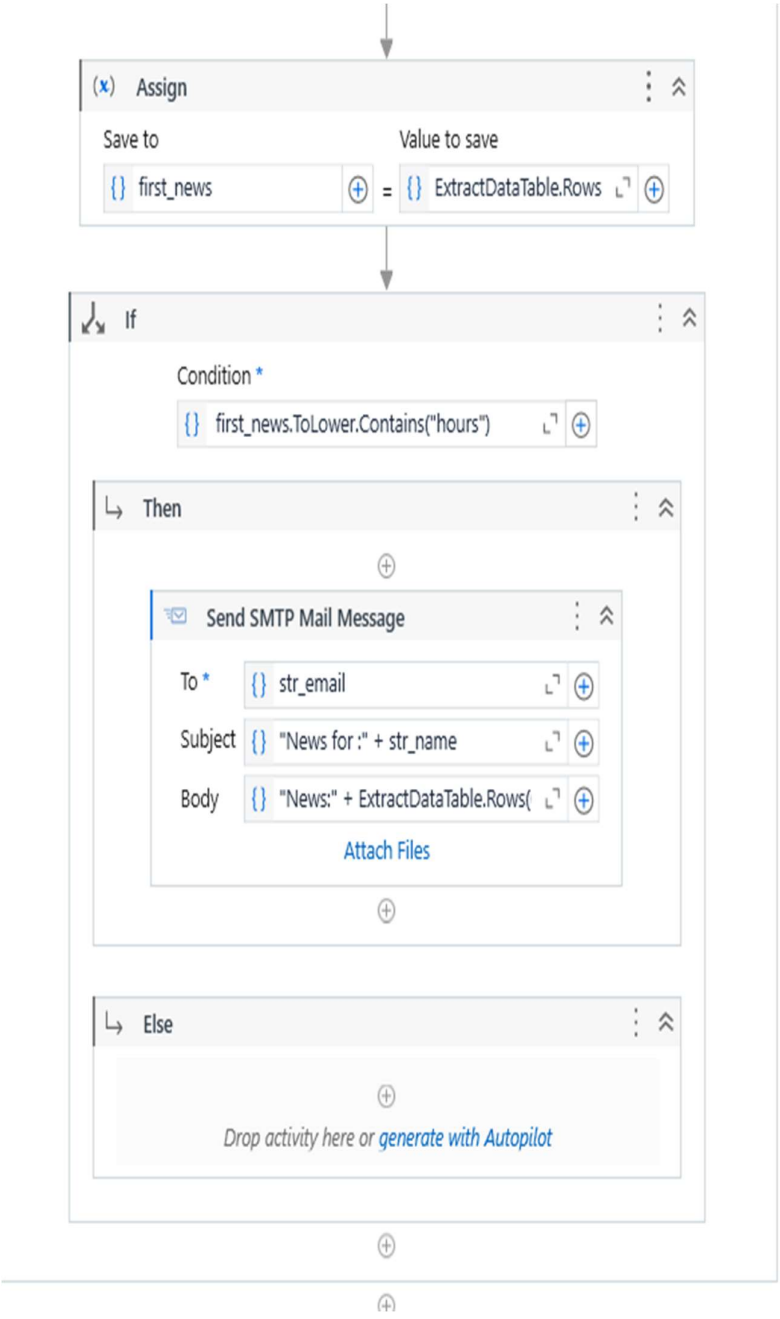


Fig 5.6 E-mail automation Workflow

REFERENCES

[1] Kuppusamy, Palanivel & Joseph K, Suresh. (2020). Automating Expense Reimbursement Processes with Robotic Process Automation. 3775.

[2] Patil, Dr & Mane, Vinod & Patil, Dr. (2019). Transforming Business Processes through Robotic Process Automation (RPA). International Journal of Innovative Technology and Exploring Engineering. 8. 3757-3760. https://doi.org/10.35940/ijitee.K2148.0981119.

[3] Elkhatat, A.M., Elsaid, K. & Almeer, S. Implementing RPA for Business Process Optimization: Case Studies and Applications. Int J Bus Autom 19, 17 (2023). https://doi.org/10.1007/s40979-023-00140-5.

[4] H. Alamleh, A. A. S. AlQahtani and A. ElSaid, "Enhancing Financial Workflows Using Robotic Process Automation," 2023 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 2023, pp. 154-158. https://doi.org/10.1109/SIEDS58326.2023.10137767.

[5] Tomáš Foltýnek, Norman Meuschke, and Bela Gipp. 2019. Robotic Process Automation in Data Management: 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. “Automation in Financial Systems: Taxonomy, Tools, and Techniques,” 19th National Convention on Business Process Automation, 2018.