SearchScrapeAssistant Application Using RPA Tool

S. Adityeshwar Goud[1], Sindhuja Ramala[1], PavanTeja Modugu[1], Ravi Kumar[1(✉)], Amit Lathigara[2], and M. Vinay Kumar Redy[2]

[1] Department of Information, Technology, Vardhaman College of Engineering (Autonomous), Hyderabad, India
ravikumar.e@gmail.com
[2] Department of Information, RK University, Rajkot, Gujarat, India

Abstract. The goal is to develop a solution using UiPath Studio's Robotic Process Automation (RPA) capabilities to scrape relevant data from LinkedIn profiles and pages efficiently and accurately. The solution should be able to navigate through LinkedIn's web interface, search for specific profiles or keywords, ex-tract desired information from profiles and pages, and store the extracted data for further analysis or integration with other systems. The specific challenges to address include Authenticating and accessing LinkedIn's website, Navigating and interacting with LinkedIn's user interface, Extracting structured data, Handling pagination and multiple search results, Error handling and resilience. By addressing these challenges, the goal is to develop a UiPath Studio automation workflow that can extract relevant data from LinkedIn web pages efficiently and reliably. The extracted data can then be used for various business purposes, providing valuable insights and streamlining processes that rely on LinkedIn data. And also to develop the solution in such a way that if the user doesn't want anything from LinkedIn then it will ask for any other query that the user wants to know and navigates that input to the chat GPT and finds the answer for the query and displays it to the user.

Keywords: Robotic Process Automation (RPA) · UiPath · Web Scrapping · Data Extraction · Linkedin

1 Introduction

Web data extraction involves the retrieval of information from websites and web pages [12]. It is an integral part of many businesses' operations [11], as it can be used for data analysis, market research, competitor analysis, lead generation, and more [16]. Using software robots or bots, Robotic Process Automation is a technology that enables businesses to automate tasks that are repetitive and rule-based [1, 2]. RPA bots can be used to interact with different systems and applications [4], including web browsers, to perform tasks efficiently and accurately [3]. LinkedIn is a treasure trove of data for businesses of all sizes [15]. From recruitment to market research, there is no shortage of ways to leverage LinkedIn data to your advantage. However, manually extracting data

X. Cheng (Ed): BROADNETS 2024, LNICST 602, pp. 154–166, 2025.
https://doi.org/10.1007/978-3-031-81171-5_16

from LinkedIn can be time-consuming and error-prone. Web data extraction, also known as web scraping, is the process of automating the extraction of data from websites [16]. By integrating web data extraction with. ChatGPT, a large language model developed by OpenAI, you can enhance the conversational abilities of the AI model and make it more intelligent and context-aware [4]. UiPath Studio is a leading RPA tool that can be used to automate web data extraction from LinkedIn [6]. By combining the power of web data extraction, RPA, ChatGPT integration, and UiPath Studio for documentation, organizations can gain valuable insights from LinkedIn data to drive business growth and innovation [4]. In this paper, we will discuss how to use web data extraction, RPA, ChatGPT Integration, and UiPath Studio to automate data collection from LinkedIn [17]. We will also provide examples of how this can be used to improve decision-making processes, enhance customer interactions, and gain valuable insight [14].

1.1 Motivation

LinkedIn is a popular professional networking platform with over 830 million members worldwide. It is a valuable source of data for businesses of all sizes, as it contains information on a wide range of topics, including employee profiles, company pages, job postings, and industry trends. However, manually extracting data from LinkedIn can be time-consuming and error-prone. This is where Robotic Process Automation (RPA) can help. RPA is a technology enabling businesses to automate repetitive and rule-based tasks through the utilization of software robots or bots. RPA bots can be used to interact with different systems and applications, including web browsers, to perform tasks efficiently and accurately. UiPath Studio is a leading RPA tool that can be used to automate web data extraction from LinkedIn. By developing a UiPath Studio RPA solution to scrape data from LinkedIn, we can address the following challenges:

- Authentication and access.
- Navigation and interaction.
- Structured data extraction.
- Pagination and multiple search results.
- Error handling and resilience.

1.2 Contribution

We developed a solution that uses UiPath Studio to automate the process of extracting data from LinkedIn [12, 16]. Our solution can efficiently and accurately scrape relevant data from LinkedIn profiles and pages, such as job titles, work experience, education, and skills [6]. It can also navigate through LinkedIn's web interface, search for specific profiles or keywords, and store the extracted data for further analysis or integration with other systems [18]. Our solution is also flexible enough to be adapted to meet the specific needs of each user [4]. For example, if a user does not want to extract data from LinkedIn, they can instead provide a query to ChatGPT, which will then find and display the answer to the query [17]. We believe that our solution has the potential to revolutionize the way businesses use LinkedIn data. By automating the process of data extraction, our solution can save businesses time and money, and help them to gain valuable insights that can drive theirsuccess [5].

1.3 *M*ethodology

1.3.1 UiPath

UiPath, previously recognized as Desk Over since 2005, initially created automation libraries and SDKs for corporations such as IBM, Google, and Microsoft. It was in 2012 when they redirected their attention to the RPA market, partnering with BPO providers like Cognizant, Accenture, and Deloitte. They further enhanced their tools to cater to various industries including BFSI, healthcare, telecom, media, and retail. UiPath is headquartered in Bucharest, [6] UiPath's RPA platform consists of three core components: UiPath Studio for process design, [20] UiPath Robot for task automation, and UiPath Orchestrator for process management. They offer both Community and Enterprise Editions, catering to users from beginners to large enterprises [11]. Recognizing the importance of training, UiPath provides a comprehensive online learning platform to enhance users' understanding of their tools and services. They offer various services, training programs, and certifications to facilitate product integration into customer enterprises. This platform covers everything needed for deploying, operating, and scaling robotic factories from a centralized dashboard. In addition to their core offerings, UiPath's Automation Testing product, TestSuite, combines Test Manager, StudioPro, Orchestrator, and Robots [13]. This suite effectively tackles typical software testing obstacles, such as regression testing, assessing business logic, and conducting combinatorial testing, aligning seamlessly with the requirements of software testing and development.

Pros of UiPath:
Strong Technology Partnerships: UiPath's extensive network of 250+ technology partners enables seamless integrations with major enterprise products, enhancing its adaptability and versatility for businesses. Active Developer Community: With a community of 750,000 developers, UiPath fosters collaborative innovation and resource sharing in its marketplace, enriching its pool of automation and AI tools.

Operational Excellence: UiPath's commitment to operational excellence is evident in its dedicated resources for helping users scale RPA solutions. They offer free and paid versions, ample learning resources, online training, and certifications, empowering users to maximize the potential of their RPA tools, including TestSuite for software.

Cons of Uipath

- pricing,
- customer support
- product upgrades and
- deployment

1.3.2 LinkedIn

LinkedIn is a professional networking platform with over 830 million members worldwide [18]. It is a valuable source of data for businesses of all sizes, as it contains information about professionals, companies, and industries. However, manually extracting data from LinkedIn can be time-consuming and error-prone.

Web data extraction, also known as web scraping, is the process of automating the extraction of data from websites [12]. By integrating web data extraction with Robotic Process Automation (RPA) and ChatGPT [11], a large language model developed by OpenAI [19], organizations can automate the process of collecting and analyzing data from LinkedIn to gain valuable insights. In this paper, we will discuss how to use web data extraction, RPA, and ChatGPT integration to automate data collection from LinkedIn. We will also provide examples of how this can be used to improve decision-making processes, enhance customer interactions, and gain valuable insights.Web data extraction, RPA, and ChatGPT integration can be used to automate the process of collecting data from LinkedIn.

- Identifying potential customers or leads
- Tracking industry trends
- Analyzing competitor activity

Understanding employee sentiment

2 Literature Review

Web data extraction, often referred to as web scraping, involves the retrieval of information from websites and web pages. This practice is fundamental to numerous business operations, serving purposes such as data analysis, market research, competitor assessment, lead generation, and more. Robotic Process Automation (RPA) stands as a technological solution enabling enterprises to automate repetitive and rule-based assignments through the utilization of software robots or bots [7]. These RPA bots possess the capability to engage with diverse systems and applications, encompassing web browsers, in order to execute tasks with precision and efficiency. LinkedIn is a treasure trove of data for businesses of all sizes. From recruitment to market research, there is no shortage of ways to leverage LinkedIn data to your advantage. However, manually extracting data from LinkedIn can be time-consuming and error-prone [10].

2.1 Web Data Extraction

Web data extraction is the process of extracting data from websites and web pages [12]. It can be used to collect a wide variety of data, such as product information, contact information, financial data, and social media data [16]. Web data extraction can be used for a variety of purposes, such as:

- Data analysis: Web data extraction can be used to collect data for analysis and reporting. For example, a business might use web data extraction to collect data on competitor pricing or product reviews.
- Market research: Web data extraction serves as a valuable method for gathering information to support market research objectives. For instance, a company could employ web data extraction to compile data related to customer preferences or prevailing trends
- within the industry.

- Competitive analysis of data: Web data extraction can be used to collect data for competitive analysis purposes. For example, a business might use web data extraction to collect data on competitor products or pricing strategies.
- Lead generation: Web data extraction can be used to generate leads for sales and marketing purposes. For example, a business might use web data extraction to collect contact information for potential customers.

2.2 Robotic Process Automation

Robotic Process Automation is a technology that enables businesses to automate repetitive and rule-based tasks through the use of software robots or bots. RPA bots can be used to interact with different systems and applications, [14] including web browsers, to perform tasks efficiently and accurately.

Automation using RPA encompasses a diverse range of tasks, including:

- Data entry: RPA bots can be used to enter data into different systems and applications. For example, an RPA bot could be used to enter customer order information into a CRM system.
- Data processing: RPA bots can be used to process data in different ways. For example, an RPA bot could be used to process data from a web form and then store it in a database.
- Reporting: RPA bots can be used to generate reports from different systems and applications. For example, an RPA bot could be used to generate a daily sales report from a CRM system [9].

2.3 LinkedIn Data Extraction

LinkedIn is a professional networking platform with over 830 million members worldwide. It is a valuable source of data for businesses of all sizes, as it contains information about professionals, companies, and industries. However, manually extracting data from LinkedIn can be time-consuming and error-prone. Web data extraction can be used to automate the process of collecting data from LinkedIn. LinkedIn data extraction can be used to collect a variety of data [8], such as:

- Professional profiles: LinkedIn data extraction can be used to collect data on professional profiles, such as name, job title, company, education, and skills.
- Company profiles: LinkedIn data extraction can be used to collect data on company profiles, such as company name, industry, location, and number of employees.
- Industry trends: LinkedIn data extraction can be used to collect data on industry trends, such as popular job titles, skills, and companies.

Benefits of Automating LinkedIn Data Extraction:
There are several benefits to automating LinkedIn data extraction:

- Save time: Automating LinkedIn data extraction can save businesses a significant amount of time.
- Reduce errors: Automating LinkedIn data extraction can help to reduce errors that can occur when manually extracting data.

- Gain insights: Automating LinkedIn data extraction can help businesses to gain valuable insights from LinkedIn data. For example, a business might use LinkedIn data extraction to identify potential customers, track industry trends, or analyze competitor activity.

2.4 Issues in Existing System

Previously, we only had the process of data extraction from LinkedIn, without any additional advantages [12]. However, with our new model, we have added several more features to make it easier for users to access more than just data extraction. We have added various activities such as Input Dialog, Message Box, Type Into, Click, Data Scraping, Write Range, and integration with ChatGPT using HTTP Request, Deserialize JSON, and related activities [4].

3 Proposed Model

The proposed model is a web automation framework that combines web automation with UiPath activities to interact with websites and extract data. The framework also integrates with ChatGPT to enhance its conversational abilities and make it more context-aware.

The framework leverages web browsing capabilities to navigate websites, perform user actions, and scroll through pages. It can also retrieve HTML content and interact with web elements. This enables users to interact with websites programmatically. Data scraping is a core feature of the framework, enabling users to extract specific data from web pages. By leveraging web scraping, screen scraping, and HTML parsing techniques, the framework can retrieve text, images, links, tables, and other relevant information from websites. This functionality is highly useful for gathering data from diverse online sources. Another key aspect of the framework is chat integration, which enables users to interact with the framework through chat applications. By integrating seamlessly with popular messaging platforms, the framework can receive commands, respond to queries, and provide requested information from the web. This integration facilitates a user-friendly and familiar interface for interacting with the framework (Fig. 1).

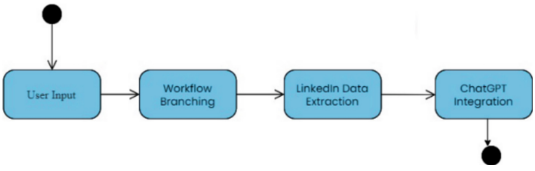

Fig. 1. Brief Workflow

3.1 Proposed Solution

Our bot automates the process of searching for companies on LinkedIn based on user input and extracting relevant data from the search results. This eliminates the need for manual effort, saving time and increasing efficiency.

Benefits:
The proposed system offers several benefits over the existing system, including:

- Increased efficiency: Our bot automates the data extraction process, saving users time and effort.
- Improved accuracy: Our bot is less prone to errors than manual data extraction.
- Greater flexibility: Our bot can be customized to meet the specific needs of each user.
- Enhanced capabilities: Our bot's integration with ChatGPT enables it to perform more complex tasks, such as understanding and responding to contextual queries.

4 Design

This documentation offers an extensive step-by-step guide to seamlessly integrate a bot with ChatGPT through UiPath Studio. The guide is divided into two primary sections, namely, extracting web data from LinkedIn and incorporating it with ChatGPT. Each individual step is meticulously explained to ensure a thorough comprehension of the entire procedure. The workflow begins with creating a project and initializing a sequence. An input dialog is included to capture user input, and the associated variable is defined. The workflow then proceeds to an if-else activity to handle different scenarios based on the user's response.

If the user selects "yes," the following actions are performed:

- The browser window is attached for further processing.
- A message box is displayed to the user.
- The user's desired domain for searching on LinkedIn is collected.
- The LinkedIn URL is entered into the search bar.
- The user's search query is entered into the LinkedIn search bar.
- The jobs button is selected on the screen.
- The available companies on the LinkedIn page are extracted.
- The extracted data is pasted into an Excel or CSV file.

If the user does not select "yes," the following actions are performed:

- A message box is displayed to the user.
- The user input is collected.
- Values are assigned for the API key and the API code is imported from the ChatGPT webpage.
- The user input and API key are sent to ChatGPT.
- The response received from ChatGPT is processed.
- Relevant information is displayed in the output panel.

Further processing involves extracting specific information from the JSON response and displaying the text output obtained from ChatGPT. Finally, a message box is used to present the output to the user.

Overall, this documentation aims to guide users through the step-by-step implementation and customization of the UiPath workflow for integrating a bot with ChatGPT. This enables enhanced conversational abilities and leverages web data extraction from LinkedIn. By following this documentation, users can efficiently implement and customize the workflow according to their specific requirements, providing an intelligent and interactive bot experience for their users (Fig. 2).

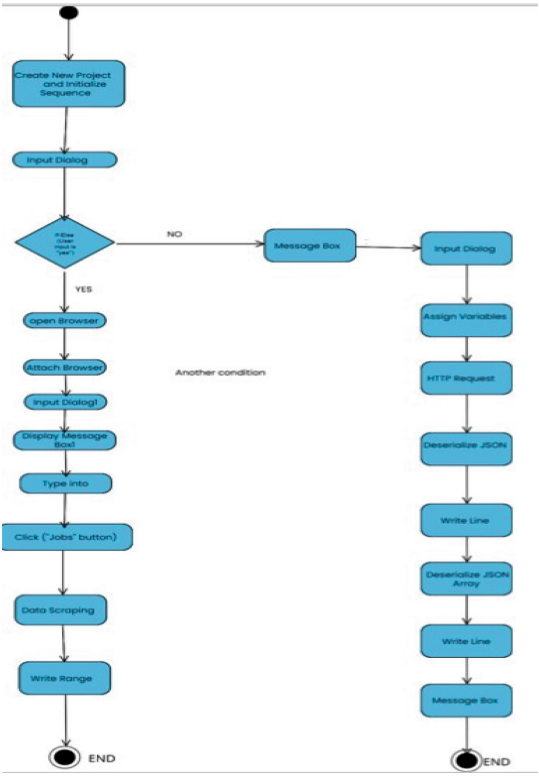

Fig. 2. Detailed Workflow

5 Execution and Result

See Figs. 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12.

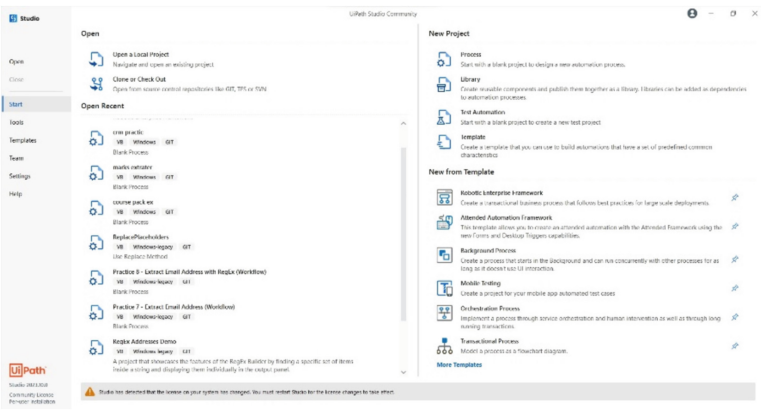

Fig. 3. The home page

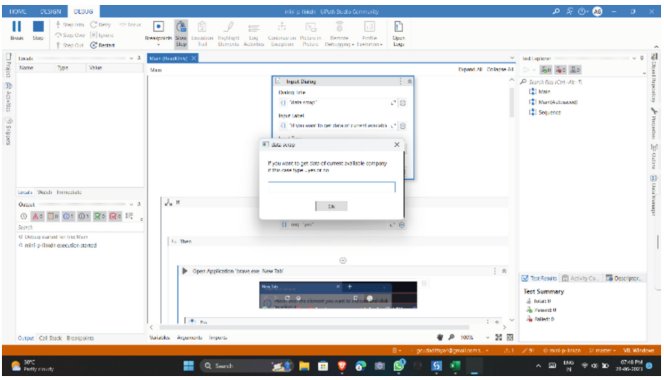

Fig. 4. Asking the user need

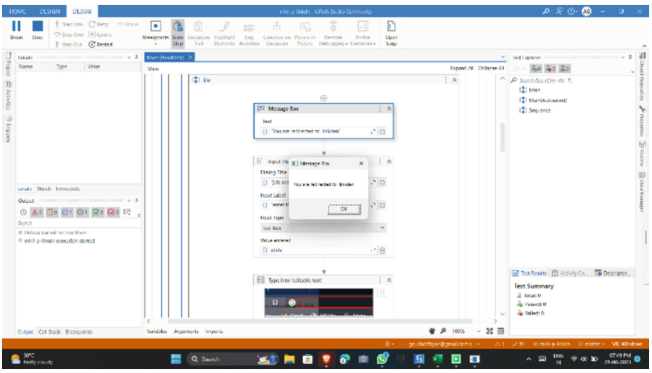

Fig. 5. Choose yes,the bot redirect to Linkdin

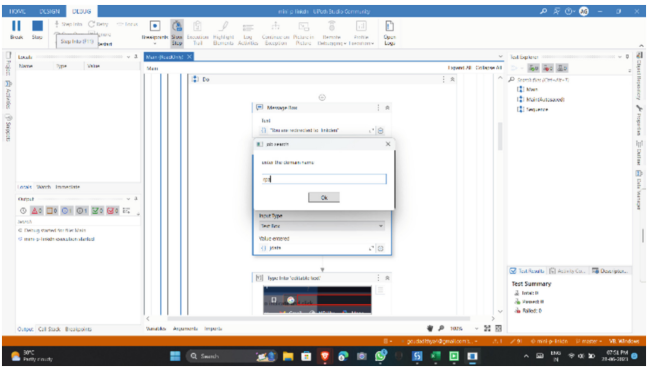

Fig. 6. Choosing the required domain.

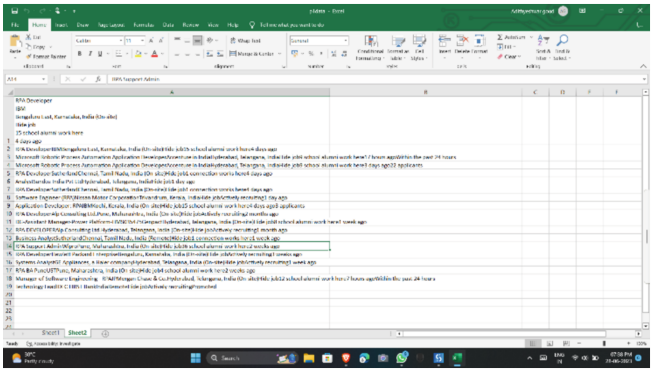

Fig. 7. Extracted data

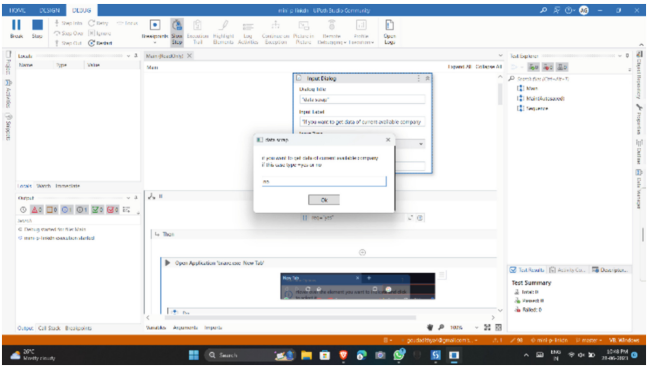

Fig. 8. Choosing no as input

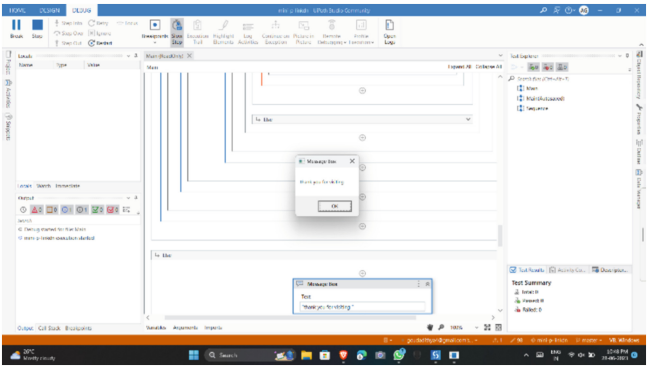

Fig. 9. Displaying message box

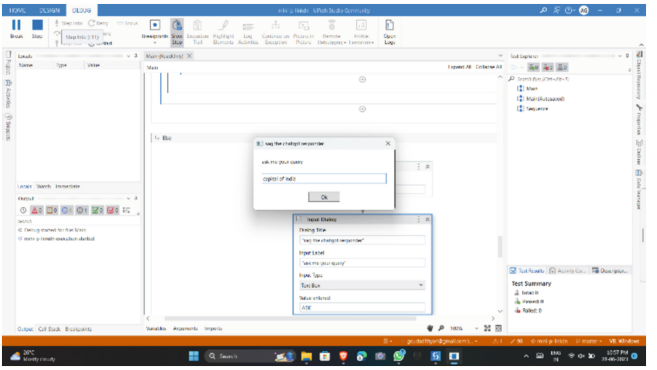

Fig. 10. Asking query to bot

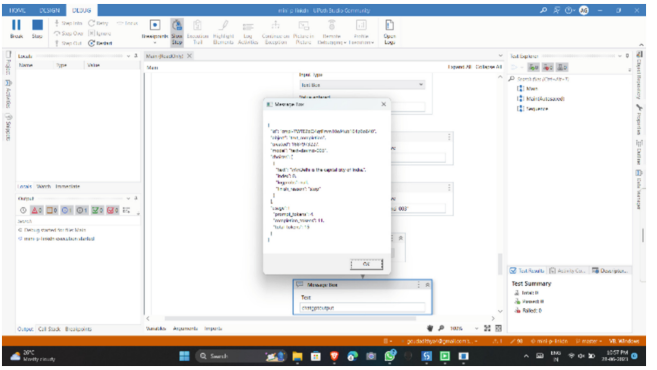

Fig. 11. The output before deserializing

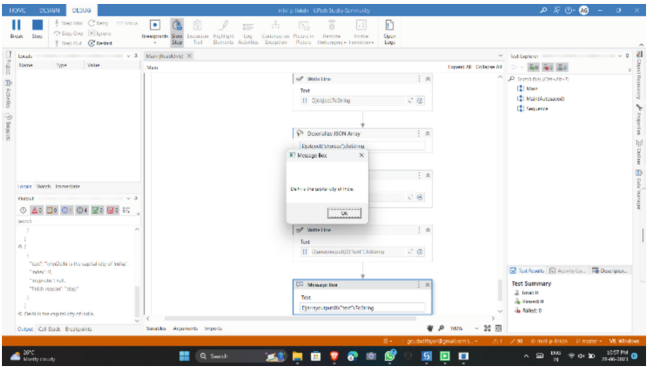

Fig. 12. The Final output

6 Conclusion

The proposed web automation framework provides a powerful and versatile solution for integrating a bot with ChatGPT and extracting data from LinkedIn. The framework's user-friendliness and adaptability make it possible to tailor it to the unique requirements of various users.It can also be integrated with popular RPA tools and chat applications. The framework has several potential applications in the real world. For example, it can be used to develop chatbots that can answer customer queries, generate leads, or provide insights into market trends. The provided framework is versatile, serving purposes such as automating tasks like data extraction, web scraping and content generation. In essence, it offers substantial utility for businesses and organizations of various scales, enhancing operational efficiency, cutting_expenses and extracting valuable insights from data. The potential for transformation in the way businesses and organizations engage with their customers and data is limitless, thanks to the ever-evolving nature of the framework, which is continuously enriched with novel functionalities and features.

References

1. Andrade, D.: Challenges of automated software testing with robotic process automation rpa - a comparative analysis of UiPath and automation anywhere. Int. J. Intell. Comput. Res. (IJICR) **11**(1), 1066 (2020)
2. Ketkar, Y., Gawade, S.: Effectiveness of robotic process automation for data mining using UiPath. In: 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), pp. 1–6. IEEE (2021)
3. Axmann, B., Harmoko, H.: Process & software selection for robotic process automation (RPA). Hrcak Croatian J. Librariansh. Inf. Sci. Museol. **55**(1), 73–86 (2022)
4. Khan, S.P., Khan, R.: RPA using UiPATH in the context of next generation automation. In: Rawat, R., Chakrawarti, R.K., Sarangi, S.K., Choudhary, R., Gadwal, A.S., Bhardwaj, V. (eds.) Next Generation Automation: Challenges, Opportunities and Solutions, pp. 477–494. Wiley (2023)
5. Yadav, N., Panda, S.P.: A path forward for automation in robotic process automation projects: potential process selection strategies. In: 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), pp. 160–165. IEEE (2022)
6. UiPath. (n.d.). UiPath: Robotic process automation (RPA) platform. Retrieved from https://www.uipath.com/
7. Automation Anywhere. (n.d.). Automation Anywhere: Robotic process automation (RPA) platform. Retrieved from https://www.automationanywhere.com/
8. Robocorp. (n.d.). Robocorp: Open-source robotic process automation (RPA) platform. Retrieved from https://robocorp.com
9. Somayya, M., Rajesh, M.H., Durgesh, K.J.: The Future Digital Work Force: Robotic Process Automation (RPA), TECSI LaboratóriodeTecnologiaeSistemasdeInformação - FEA/USP (2019). https://www.redalyc.org/jatsRepo/2032/203261541001/html/index.html
10. ScienceDaily, Big Data, for better or worst: 90% of world's data generated over last two years (2013). https://www.sciencedaily.com/releases/2013/05/130522085217.htm
11. https://docs.uipath.com/studio/docs/about-data-scraping
12. https://docs.uipath.com/studio/docs/example-of-using-data-scraping
13. https://docs.uipath.com/activities/docs/write-csv-file

14. https://www2.deloitte.com/content/dam/Deloitte/fr/Documents/servicesfinanciers/publications/deloitte_globalrobotics-survey-2018-full-report.pdf
15. https://ro.linkedin.com/jobs/search?keywords=IT&location=Bucharest%2C%20Bucharest%2C%20Romania&trk=homepagejobseeker_jobssearchbar_searchsubmit&redirect=false&position=1&pageNum=0
16. https://www.uipath.com/developers/video-tutorials/web-data-extraction-automation
17. https://www.mckinsey.com/industries/technology-media-and-telecommunications/ourinsights/the-social-econom
18. https://www.linkedin.com
19. https://openai.com/blog/chatgpt
20. https://www.uipath.com/product/orchestrator