

**Certificate**

This is to certify that

19202A0047– Shourish Kothawale

19202A0011 – Shubham Jadhav

19202A0054 – Shishir Nath

19202A0006 – Piyush Nirbhavane

**Data Scraper for Google Maps**

A partial fulfilment of the requirement of the third-year diploma in Information Technology affiliated to Maharashtra State Board of Technical Education, Mumbai for the Academic year 2021 - 2022.

Internal Examiner Project guide

Head of department Principal

**Abstract**

Information is paramount nowadays and is seemingly growing day by day. If correct information is delivered quickly, it can be a highly effective in situations. New technology is being discovered daily to deliver data as quickly and efficiently as possible. To provide a solution to this problem we will create a data/web scraper for Google Maps for providing different kinds of data like name, phone, website etc. Web Scrapers allow to find specific data on the internet and then export them into spreadsheets like Microsoft Excel sheet or a CSV file. Web Scrapers are created to extract personalised data hence this project will focus on extracting data through the Google Maps website. Most of the times these scrapers are used to extract data from HTML documents, but sometimes they can even extract data through CSS and JavaScript. Web Scrapers are of many types from self-built to pre-built, and can be available in Software, and Browser extension formats. These web scrapers are built in many different programming languages, but python seems to be the most popular of them all as it has many different libraries that are perfect for the job. Scrapers not only have many functions, but they also have many applications. They can be used in business analytics as it can help analyse competition. Market Research is one of the major uses of the scrapers as the data is specific and can be used to identify various trends in the market.

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Title** | **Pg. No** |
| 1 | Introduction | **1** |
| 2 | Problem Statement | **2** |
| 3 | Features/Functionalities of the Project | **3** |
| 4 | Review of Literature | **5** |
| 5 | Modules/Users/Stakeholders | **8** |
| 6 | Plan of Work | **9** |
| 7 | Algorithm/ER Diagram/Circuit Diagram | **10** |
| 8 | Design (DFD & UML Diagrams) | **11** |
| 9 | Technology Used | **17** |
| 10 | Pros and Cons | **18** |
| 11 | Future Scope | **20** |
| 12 | Applications | **21** |
| 13 | Conclusion | **22** |
| 14 | References | **23** |

**List of Tables and Figures**

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Title** | **Pg. No** |
| 1 | Table (1) Market Survey | **5** |
| 2 | Fig. (1) Timeline Chart | **9** |
| 3 | Table (2) Keyword/URL/Location | **10** |
| 4 | Table (3) Restaurant Search | **10** |
| 5 | Fig (2) DFD Level 0 | **11** |
| 6 | Fig (3) DFD Level 1 | **12** |
| 7 | Fig (4) DFD Level 2 | **13** |
| 8 | Fig (5) Flowchart | **14** |
| 9 | Fig (6) Use Case Diagram | **15** |
| 10 | Fig (7) Sequence Diagram | **16** |

1. **Introduction**

Data Scraping is also known as web scraping. It is the process to import information from a website to a spreadsheet or local file. It can help get data easily, and also help to channel data to other websites. Data Scraping can be used in many places in the modern day, from price monitoring, market research to even Sentiment Analysis. Web Scraping has many applications and has a wide variety of usage. Google Maps is a website full of information. It contains the location of a place and even pinpoints it, and it has many different attributes for places like contact numbers, websites, contact number, rating, review count, timings etc. It is a place which contains a lot of data, and efficient scraping of the website can be done helping many causes.

The main aim of the application is to extract the data as quickly and efficiently as possible. Since getting information is important for many reasons this app will be very helpful in doing so. The app will help with this as it with be able to pick out the appropriate data from a page with unstructured data and then will do it way quicker than someone sitting and manually noting down the data in an excel sheet which might take hours.

The application will have an array of functionalities, which starts off with searching for the place using an URL. This URL has to be specific to that of the place they want to search so they can do this by providing the google maps URL for their preferred place. The next feature that will be available in the application is the use of keywords. The user can simply enter a keyword like “Pizza places Mumbai” and get the attributes such as the address (location), review count, rating, timings, contact number, and website of the places that show in the result of the place. This can help gather data quickly from the google maps page as it will be able to scan all the relevant data on the place. Another way to search will be simply through providing the location name, like “Mumbai”, “Paris” to search for places related to it. The next functionality that will be in the application is the filter option. Users will have an option to either keep the data with a blank phone number field or to remove the data. This helps get more personalised results. After searching all the data this data has to be saved. To do this we will eb allowing the user to import the data into two formats. The first format which will be used is Microsoft Excel spreadsheet which has a .xlsx extension. The second import format will be CSV with a .csv extension. The only prerequisite for the user will be that they have to prepare the excel sheet with its name beforehand. There will be buttons to execute the import.

With data increasing daily it is becoming more and more difficult to extract the appropriate data online. Since information is so important nowadays this project will surely play its part to accomplish its main goal. Sites with abundant unstructured data can be easily broken down extracting the appropriate data within a matter of seconds. With Google maps being online for so many years accumulating a lot of data, it will be a website that will help in extracting data online. This data can then be used for many purposes, and the use of the application is limitless.

1. **Problem Statement**

In today’s world information plays a very important part. Practically, everything we do requires information. With the world advancing information is growing and expanding at an extraordinary pace. With this comes the problem of unstructured data everywhere, along with redundant data. Every second new data is being created, for example with Google over 40,000 search queries are made every second, that’s a whopping 1.2 trillion searches per year. Around 44 zettabytes With so much data it is becoming more and more difficult to find data that is appropriate to the cause and in turn is making it difficult to find it. This can cause a major crisis for many people as the world works on data, from businesses to science data is integral. To perform appropriate searches data scrapers can be used. Data scrapers help to find only the appropriate data in a website so that you don’t have to search for the data on your own manually.

After looking through various data scraper one common thing was, they were subscription based, and they only had one option to search by that was using keywords. But with our Google Maps Web Scraper we will be provided to search for places using location, URL, Keyword and even a special search option for restaurants since they have different attributes compared to other types of location with generic characteristics.

Google maps contains a lot of data about a place from location to opening times, popular hours etc. These types of data can be used by businesses to analyze and formulate the correct strategy. Not only businesses but many other people who like to travel can search for data on famous places to plan out their next trip. Data Scrappers are not limited to a specific group of people, they can be used by everyone to find information about their liking.

**3. Features / Functionalities of Project**

Overall, the working of the application will be simple and easy to use as that is the main aim of the project. All the user has to do in the application is to enter the keywords in the text box or a URL and mention the excel sheet which has been already prepared. Then after clicking the search button the data will be transferred into the excel sheet saving all the information. Users will be able to view the data after it has been saved in the Excel sheet. The application will also contain a filter operation which will remove the data with blank records and more. To Implement this and develop the project we will have to use and IDE which is an Integrated Development Environment. For this project we will be using Jupyter Notebook and PyCharm since they are free to use and have many modern-day features.

1. **Extract data into Excel and CSV files:**

After Searching for the data using the different search options the data will be temporary displayed in table format in the application. But to save the data we will require to save it in a separate file like Excel or CSV. In order to do this, we will be providing a functionality. With the simple click of a button the file will be saved in the preferred format that the user selects. Excel is a professional spread sheet application, so it is well formatted and user friendly. While CSV files or Comma Separated Values files are unstructured, they are easier to use in programming than Excel, hence both formats are provided.

1. **Filter Operations:**

After doing thorough market research we found out that many of the web scrapers provide data with blank fields. This causes inconsistent data. To combat this, we will be providing options to either keep the blank fields or to simply remove them. Other options for filtering include blank mobile number, email and website.

1. **Search Place using URL:**

Our first method to search Google Maps will be through the use of an URL. The user must provide a URL for a particular search and then the data for that part will b scrapped.

1. **Search Place using Keyword:**

The next search method will be through keywords. The user for example can enter “Pizza Places in Mumbai” and get data regarding it. This makes the data more personalised.

1. **Search Place using Location:**

The next way to search would be through location name. This means that the locations such as England, India, London, New York etc. can be mentioned. It will provide personalised information about the place like Quick Facts.

1. **Special Tab for Restaurants:**

A Special search option is dedicated to restaurants as the structure for restaurants is slightly different from that of other places. Additional options such as descriptions, tags and things like menu details and place an order are only specific to restaurants. Even though restaurants can be searched through the keywords tab, the restaurant-specific tab will provide better results.

**4. Review of Literature**

For Market research we looked through the internet for different data scrapers and also looked for specific Google Maps data scrapers. We looked at the features of each app and noted them down along with their obvious disadvantages. We looked through web scrapers both paid and unpaid. We downloaded all the apps and tested each of them.  
  
**Apps we looked through:**

* Web Scraper- Free Web Scraping
* DigLead: Business Lead Scraper
* Google Map Business Scrapper
* Google Map Scrapper Listing
* Leads Extractor - Google Maps Scraper

|  |  |  |  |
| --- | --- | --- | --- |
| Logo | App Name | Features | Disadvantages |
|  | Web Scraper - Free Web Scraping | \* Data scraping from multiple pages;  \* Multiple data extraction types (text, images, URLs, and more);  \* Scraping data from dynamic pages (JavaScript + AJAX, infinite scroll);  \* Browsing scraped data;  \* Exporting scraped data from a website to Excel | The GUI for this web extension is not user friendly, and it is difficult to use for the layman. |
|  | DigLead: Business Lead Scraper | Main Features:  -Targeted List Generator  -Webpage or Website Email Scraper  -Google My Business Listings Scraper (Google Maps/GMB Scraper)  -Yelp Lead Extractor  -Export/Download all scraped leads in Excel Spreadsheet which can be share via email or open with spreadsheet viewer. | Even if installing the app is free it contains a subscription fee to use it. Also, it has limited functionality. |
|  | Google Map Business Scrapper | instantly generate a ton of leads, anywhere in the world. This Google maps extractor tool extracts **name, address, phone number, website, social media links, reviews, current marketing pixels** and more from companies. | This app is paid so not everyone will afford it. |
| \*No Image\* | Google Map Scrapper Listing | Software Extract All Detail from Google Business  Name  Phone  Email – if email available on website for particular Business Website  Website  Rating  Review  Images – with Live Link and Also Download images into Local Computer  Business Closing Hours  Software Download Business images into Local Computer with Business Name Folder separately  Software Search Google Business Contact Location Wise Filter.  100% Real Time Extract Business Contact with Advance Export into Excel, CSV and Text Format. | This app is paid at a high price of $22 |
|  | Leads Extractor - Google Maps Scraper | \* Find leads for your business in any industry across all countries  \* Qualify leads faster than ever before  \* Increase productivity by reducing hours spent looking for leads | Limited number of options as they are only limited to a combo box. |

Table (1) Market Survey

The observations from this research were that the results from the applications generated many blank fields. The reason for this was the data that was requested was never present in the structure of the data. Hence, we decided to add unique personalized tabs for each type of points of interests, like location, restaurants, which contain different attributes in its structure.

**5. Modules / Users / Stakeholders**

Users:

Naive User: The naive users are those users which interact with the application without having any knowledge about how it works. They just interact with the system to do their daily activities. This user will be the one that uses the Data Scraper for data extraction. The user will be performing the search queries through various ways. From entering keywords, URLs, Locations. The users also will be the ones to import the data into Excel or CSV.

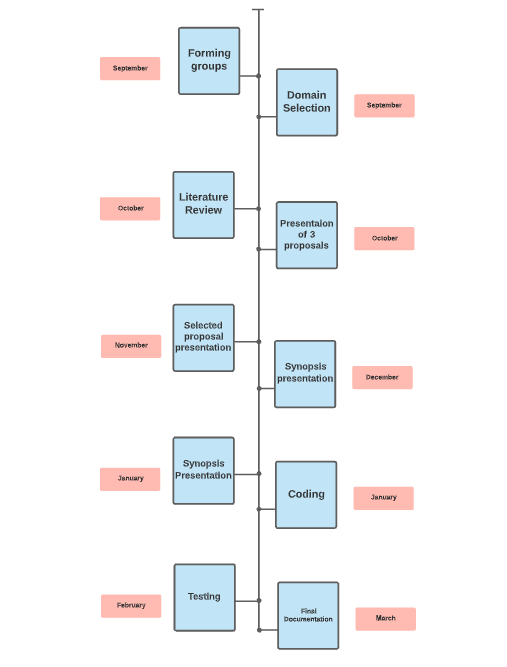
**6. Plan of Work**

Fig. (1) Timeline Chart

**7. Algorithm / ER Diagram / Circuit Diagram**

**Table Representation for Excel and CSV Output:**

Keyword/URL/ Location:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Rating | Address | Website | Number |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Table (2) Keyword/URL/Location

Restaurant:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | Rating | Review Count | Price Rating | Tags | Address | Open hours | Website | Number |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

Table (3) Restaurant Search

**8. Design (DFD & UML Diagrams)**

**8.1 DFD Diagrams**

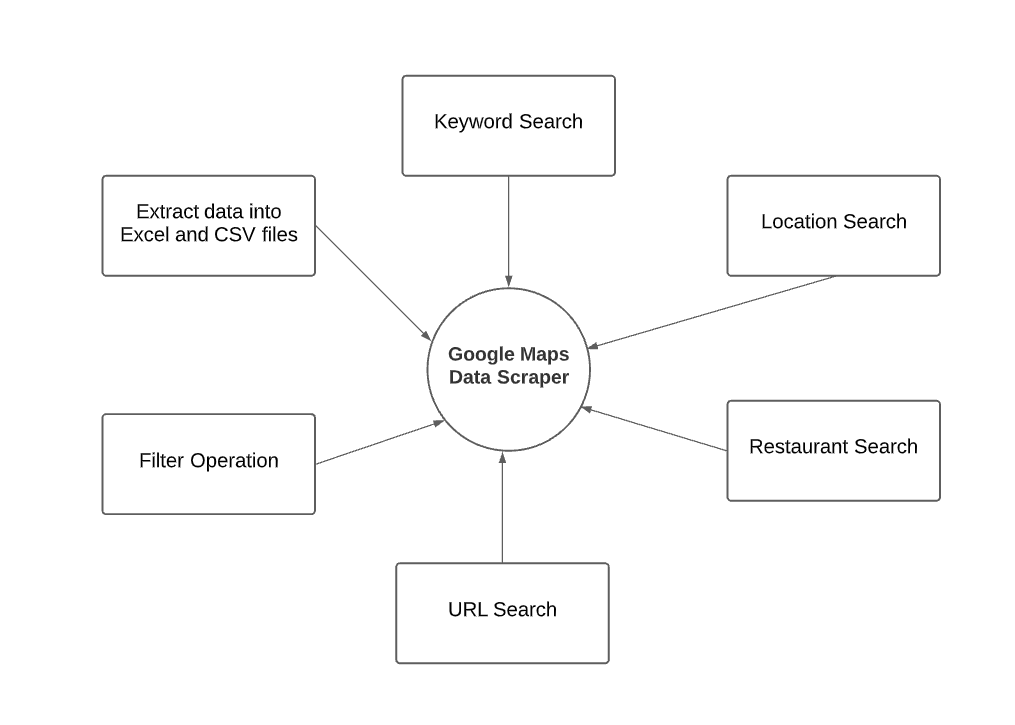
**8.1.1 DFD Level 0**

Fig (2) DFD Level 0

**8.1.2 DFD Level 1**

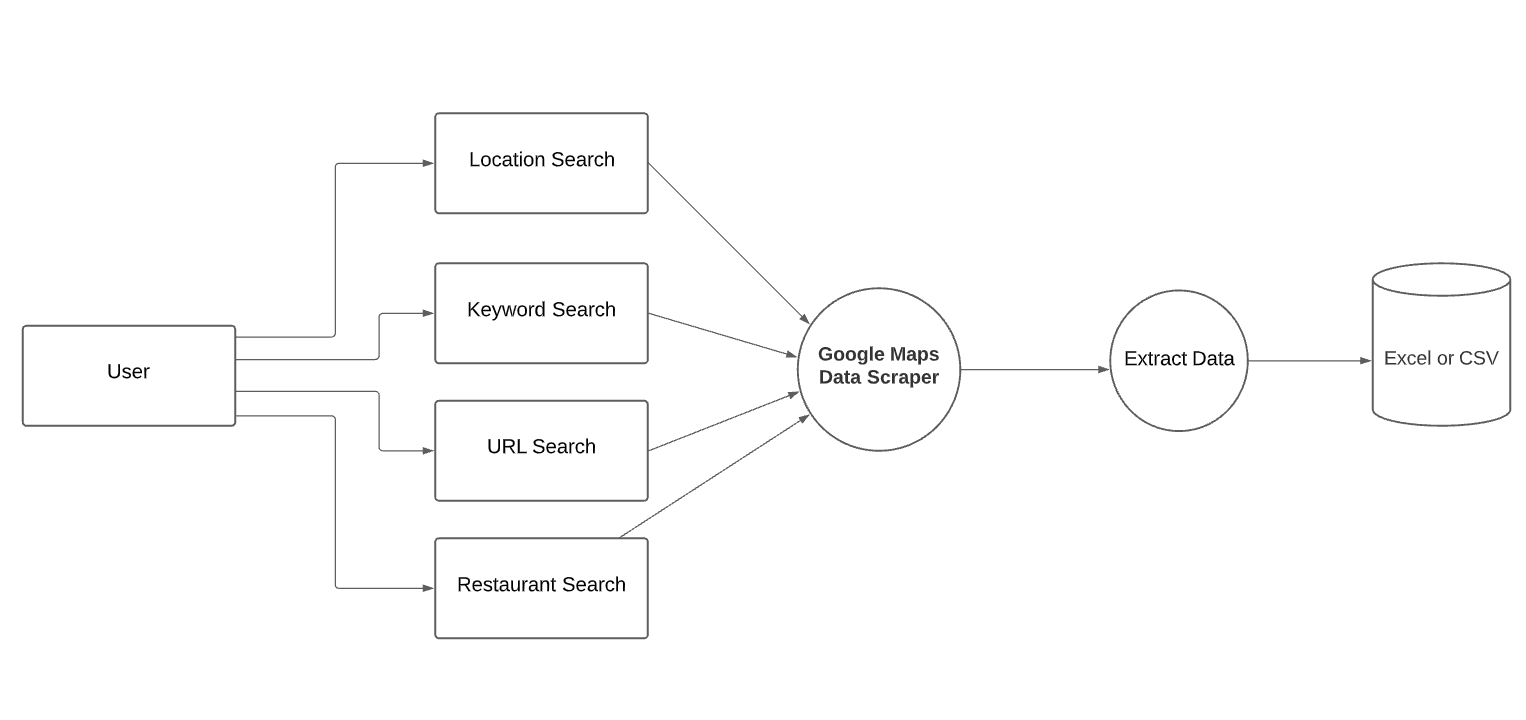
****

Fig (3) DFD Level 1

**8.1.3 DFD Level 2**

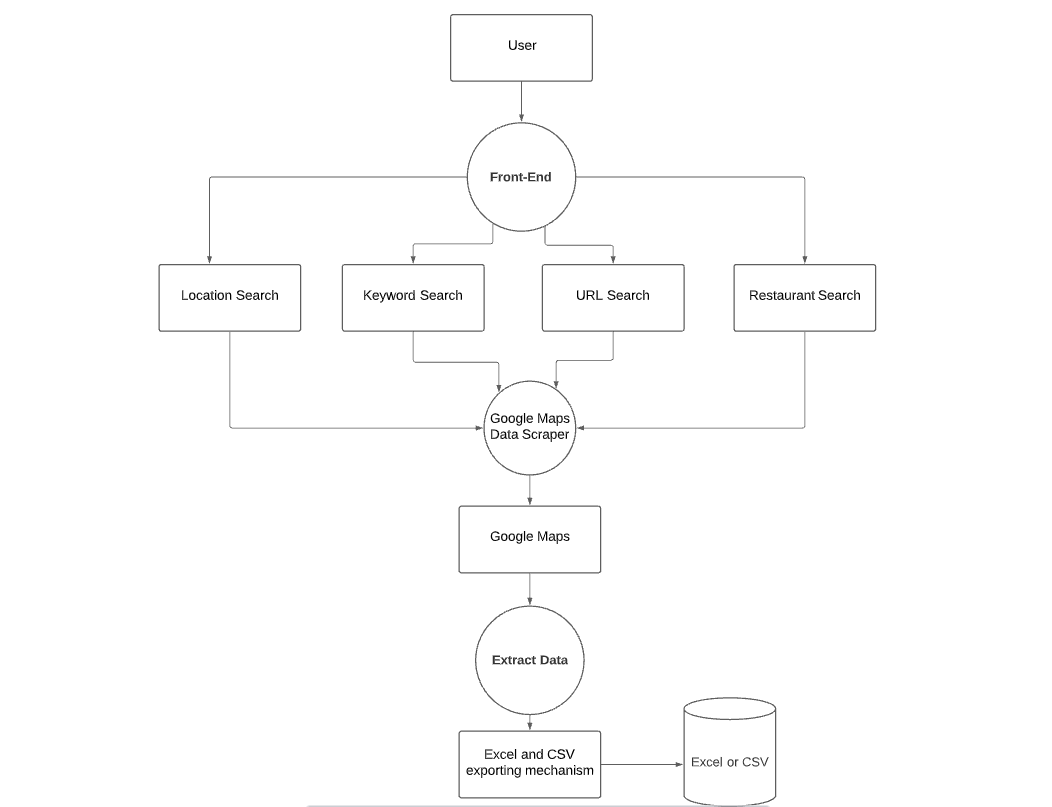


Fig (4) DFD Level 2

**8.2 UML Diagrams**

**8.2.1 Flowchart**

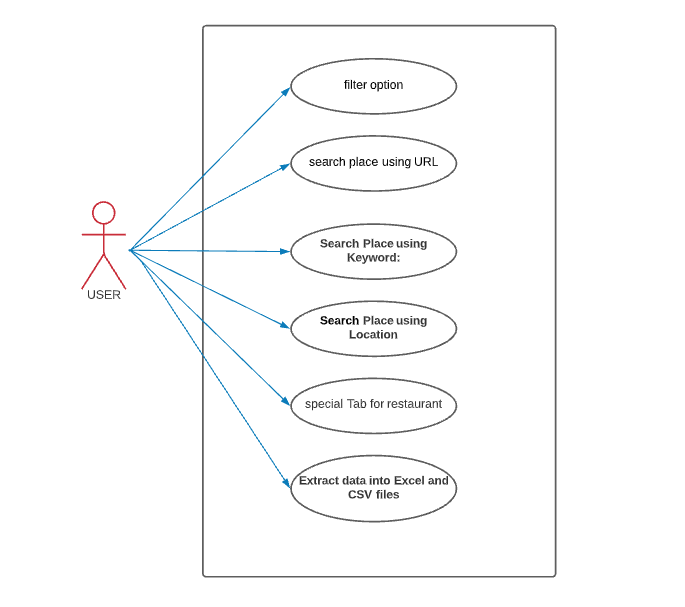
**8.2.2 Use Case Diagram**

Fig (5) Flowchart

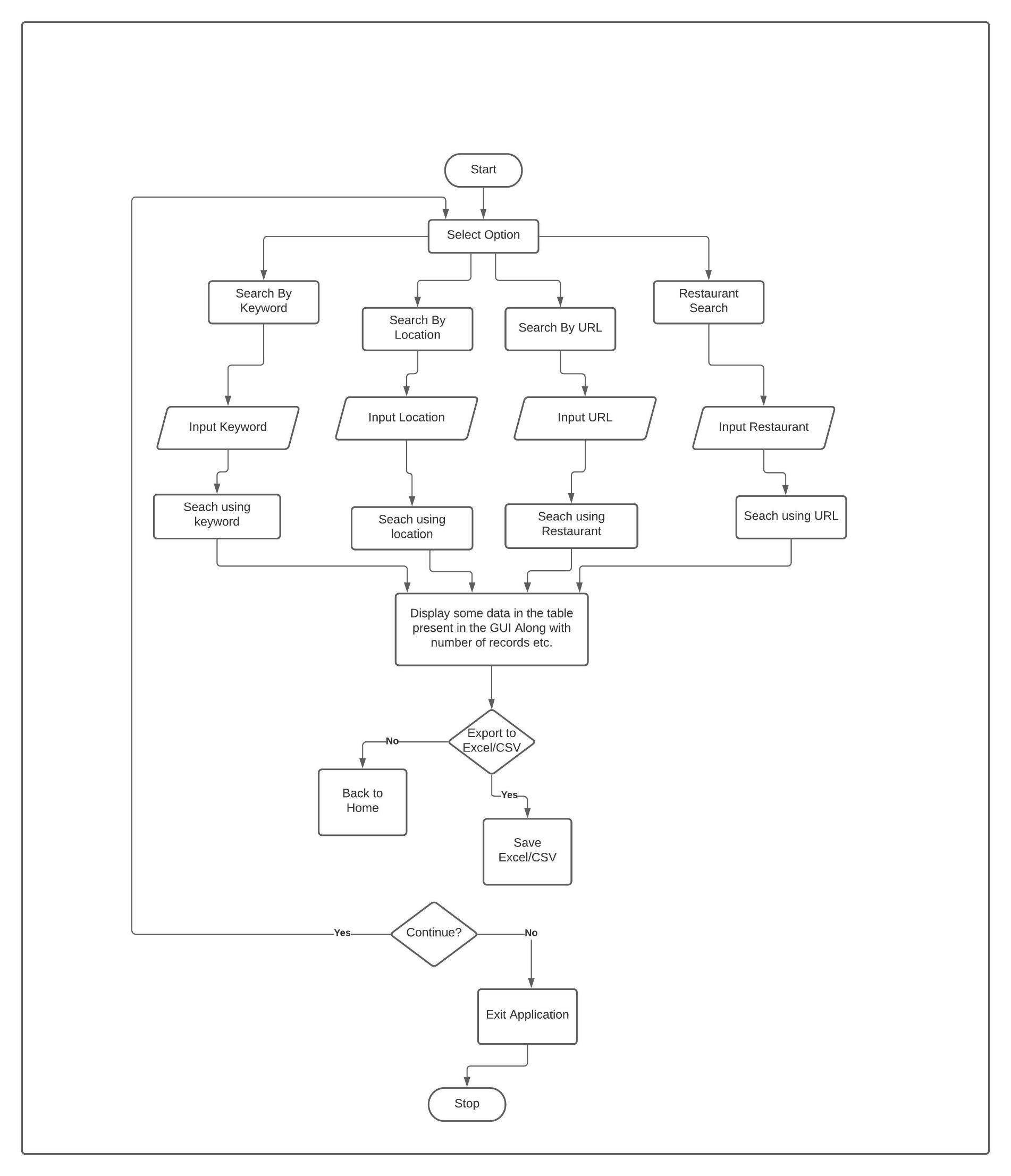


Fig (6) Use Case Diagram

**8.2.3 Sequence Diagram**

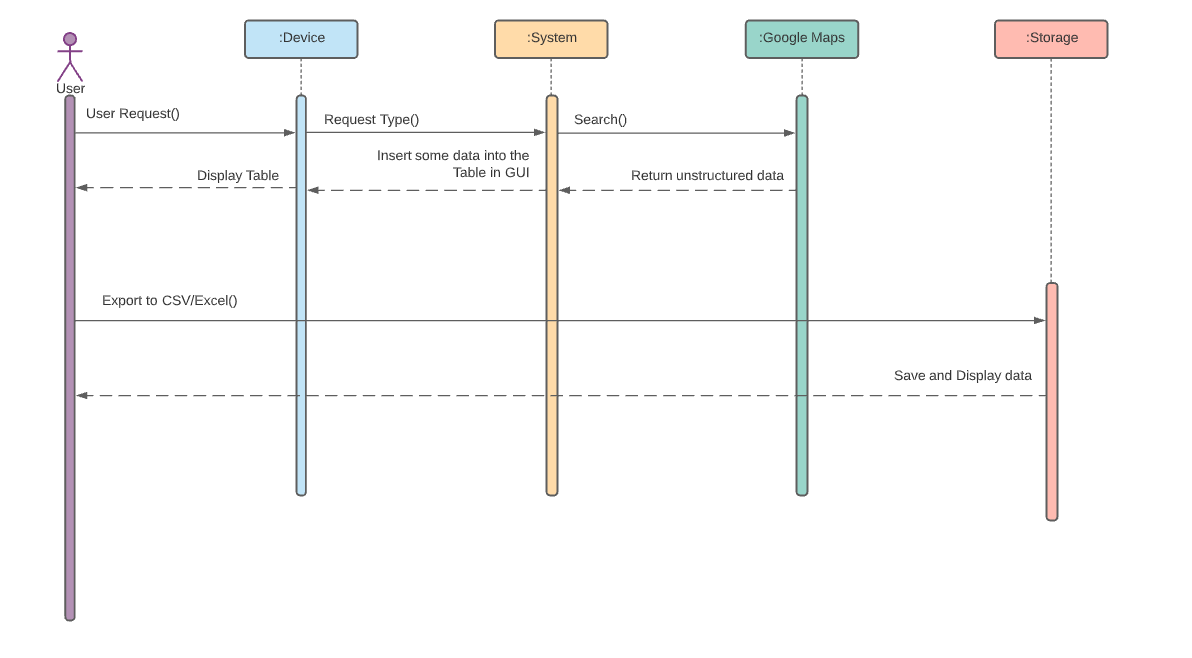
****

Fig (7) Sequence Diagram

**9. Technology Used**

For Implementing this we will be using various technologies for back-end like Python and front-end technologies like Tkinter a Python library. We will be using Microsoft Excel to export the data and store it so that it can later be used. The reason for using python is that it consists of many libraries that are perfect for this kind of application. Selenium, Beautiful Soup, Pandas, Openpyxl are a few of the libraries that will be used for this project. For searching the internet, we will be using the Google Chrome browser which will help us to find the specific Google Maps website, hence we will be using the Chrome Web driver to create an instance of the browser.

* Hardware Requirement: -

The only hardware requirement for this project would be a laptop with minimum requirements of a 3.3GHz or faster 64-bit dual-core processor. The RAM required for this would be a minimum of 4GB but 8GB is recommended.

* Software Requirement: -
* Front End: -

Tkinter will be used for front-end. Which is a python library used to create a GUI

* Back End: -
  + Selenium: Selenium is a web testing library. It is used to automate browser activities.
  + BeautifulSoup: Beautiful Soup is a Python package for parsing HTML and XML documents. It creates parse trees that is helpful to extract the data easily.
  + Pandas: Pandas is a library used for data manipulation and analysis. It is used to extract the data and store it in the desired format.
  + openpyxl – Used to work with Microsoft Excel
  + os and time standard libraries
  + ChromeDriver for using Selenium and creating an instance for the Chrome browser.
  + Tkinter for front-end

1. **Pros & Cons**

**Advantages of Web Scraping:**

1. Automation

The first and most important benefit of web scraping is developing tools that have simplified data retrieval from different websites to only a few clicks.

1. Cost-Effective

Web scraping services provide an essential service at a competitive cost. The data will have to be collected back from websites and analyzed so that the internet functions regularly. Web scraping services manage to do this in a cost-effective and budget-friendly manner.

1. Low Maintenance and High Speed

Web Scraping does have a very low maintenance cost associated with it over a while. In that way, it helps plan the budget accurately. Also, scraping web saves a lot of time, as it can do a day’s manual work in a few hours.

1. Data Accuracy

Simple errors in data extraction can lead to major issues. Hence it is needed to ensure that the data is correct. Data scraping is not only a fast process, but it’s accurate too. This reputation helps while collecting important data such as sales price, financial data to name a few.

1. Easy to Implement

Once a website scraping service starts collecting data, you can rest assured that you are getting data from not just a single page but from the whole domain. With a onetime investment, it can have a high volume of data.

1. Easy Integration

The ease with which an API can be integrated into a developer’s application is one of its most appealing features.

1. Customization

A web scraping API allows you to personalize it and use its capabilities to its full potential to achieve all of your scraping goals, from API calls and geotargeting to dedicated accounts and custom scrapers.

**The Disadvantages of Web Scraping:**

1. Difficult to Analyze

For those who are not much tech-savvy and aren’t an expert, web scrapers can be confusing. Even though it’s not a major issue.

1. Not all locations of Google Maps have the same attributes for example restaurants, hence the personalized web scrapers won't be able to search all the attributes together for certain results.
2. **Future Scope**

* The Google Maps Data Scraper can be enhanced by making it scrape data from CSS and JavaScript as well for more accurate results.
* The Data Scraper can be enhanced with A.I and M.L to get better results.
* The functionality of all the search options could be merged into one by just entering a single word instead of URL and location.
* Location Service can be added to it to make it more location specific.
* The Application could also be improved to meet requirements such as ISO.

1. **Applications**

* Lead Generation

One incredibly popular use of web scraping is lead generation. In short, web scraping is used by many companies to collect contact information about potential customers or clients. This is incredibly common in the business-to-business space, where potential customers will post their business information publicly online.

* Real Estate Listing Scraping

Many real estate agents use web scraping to populate their database of available properties for sale or for rent.

* Data Analysis

You might want to collect and analyse data related to a specific category from multiple websites. The category might be real estate, automobiles, electronic gadgets, industrial equipment, business contacts, marketing etc.

* Academic Research

Data is an integral part of any research, be it academic, marketing or scientific. A Web Scraper will help you gather structured data from multiple sources in the Internet with ease.

* Training and Testing Data for Machine Learning Projects

Web Scraping helps you to gather data for testing / training your Machine Learning models. Quality of your machine learning models depends on the quality of training data used and when the data is not readily available you can employ web scraping to collect it from various websites.

1. **Conclusion**

To conclude, the Google Maps data scraper will offer a variety of personalised search features, at a free cost. Since information is vital nowadays it will help in the cause. It will contain various search parameters from, keywords, location, to even URLs. Due to restaurants having different structures we will be providing a unique feature only known to our app that provides special searches for restaurants. The users will later also be able to export the data in to Excel or CSV formats in their prepared sheets. Use of Python and it’s libraries will be vital for the completion of the project.

The main aim of this project is to make it easier and quicker to find data. Hence this we scraper helps as it can save time and even makes it easier to view data. This data scraper can not only be useful in business fields but can be used for various other analytical and statistical reasons.

1. **References**

[https://chrome.google.com/webstore/detail/web-scraper-free-web- scra/jnhgnonknehpejjnehehllkliplmbmhn?hl=en](https://chrome.google.com/webstore/detail/web-scraper-free-web-scra/jnhgnonknehpejjnehehllkliplmbmhn?hl=en) - Web Scraper-Free Web scraping extension

<https://play.google.com/store/apps/details?id=com.lead.dig&hl=en_IN&gl=US> - DigLead Business Lead Scrapper

<https://chrome.google.com/webstore/detail/google-map-business-scrap/lfeafcjdmideekklmmioaaaonahmhioe?hl=en> – Google Map Business Scrapper Extension

<https://chrome.google.com/webstore/detail/leads-extractor-google-ma/gaojpolkokdnlbdnejallmpmcigpnhan?hl=en> – Leads Extractor

<https://www.targetinternet.com/what-is-data-scraping-and-how-can-you-use-it/> - what is data scrapping and how to use it

<https://www.edureka.co/blog/web-scraping-with-python/> - Web Scrapping Libraries and brief introduction

<https://medium.com/swlh/scraping-google-maps-using-selenium-3cec08eb6a92> - How to scrape from google Chrome

<https://www.crummy.com/software/BeautifulSoup/bs4/doc/> - Beautiful Soup Documentation

<https://www.selenium.dev/documentation/> - Selenium Documentation

<https://chromedriver.chromium.org/downloads> - Chrome Driver

<https://developers.google.com/maps> - Google maps API Docs

<https://www.youtube.com/watch?v=HChq5_7yTGk&list=PL3JVwFmb_BnQrh3CFjyxD4vfyZ7ovvDLK> – Google Maps API and Python

<https://pypi.org/project/googlemaps/> - PyPi googlemaps python client

<https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/> - what is web scraping

<https://www.zyte.com/blog/web-scraping-requirement-gathering/> - Web scraping steps

<https://www.lucidchart.com/pages/> - Lucidchart