A Major Project Report

On

SCRAPE YOUR NEED

Submitted in Partial Fulfillment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

(Computer Science and Engineering)

Submitted by

Himanshi Chauhan (160970101022)

Simran Sharma (160970101049)

Shubham Singh Negi (160970101048)

Under the Guidance of

Ms. Pragya Baluni

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

THDC INSTITUTE OF HYDROPOWER ENGINEERING & TECHNOLOGY

TEHRI, UTTARAKHAND

(Uttarakhand Technical University, Dehradun)

2016-2020

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled "**SCRAPE YOUR NEED"** in partial fulfillment of the requirement for the award of degree of **Bachelor of Technology** and submitted in Department of Computer Science and Engineering of THDC Institute of Hydropower Engineering & Technology, Tehri, is an authentic record of my own work carried out. Under the supervision of Mr. Pragya Baluni, Assistant Professor, Department of Computer Science and Engineering, THDC Institute of Hydropower Engineering & Technology, Tehri.

The matter presented in this report has not been submitted by me anywhere for the award of any other degree of this or any other institute.

Himanshi Chauhan (160970101022)
Simran Sharma (160970101049)
Shubham Singh Negi (160970101048)

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

Date:

(Ms. Pragya Baluni) (**Mr. Ashish Joshi**)
Project Guide Co-Guide

HOD Project In-charge

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B.Tech. Minor Project undertaken during B. Tech. Final Year. We owe special debt of gratitude to Professor Ms. Pragya Baluni, Department of Computer Science and Engineering, THDC Institute of Hydropower Engineering & Technology, Tehri for her constant support and guidance throughout the course of our work. Her sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day. We also take the opportunity to acknowledge the contribution of Professor Ms. Pragya Baluni support and assistance during the development of the project.

We also do not like to miss the opportunity to acknowledge the contribution of all We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Himanshi Chauhan (160970101022)
Simran Sharma (160970101049)
Shubham Singh Negi (160970101048)

ABSTRACT

Web Scraping (also termed Screen Scraping, Web Data Extraction, Web Harvesting etc.) is a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to a local file in your computer or to a database in table (spreadsheet) format.

Data displayed by most websites can only be viewed using a web browser. They do not offer the functionality to save a copy of this data for personal use. The only option then is to manually copy and paste the data - a very tedious job which can take many hours or sometimes days to complete. Web Scraping is the technique of automating this process, so that instead of manually copying the data from websites, the Web Scraping software will perform the same task within a fraction of the time.

The proposed system for this project “Scrape Your Need” is a web scraper that is able to access and extract information Housing, Education and Job at various locations from websites using a standalone application as an interface for user interaction. The extracted details are then stored in a csv file, as the standalone application allows the user to search through and query the saved findings.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 About python IDLE 3.7

IDLE (Integrated Development and Learning Environment) is an integrated development environment (IDE) for Python. The Python installer for Windows contains the IDLE module by default.

IDLE is not available by default in Python distributions for Linux. It needs to be installed using the respective package managers. For example, in case of Ubuntu:

$ sudo apt-get install idle

IDLE can be used to execute a single statement just like Python Shell and also to create, modify and execute Python scripts. IDLE provides a fully-featured text editor to create Python scripts that includes features like syntax highlighting, auto completion and smart indent. It also has a debugger with stepping and breakpoints features.

To start IDLE interactive shell, search for the IDLE icon in the start menu and double click on it.

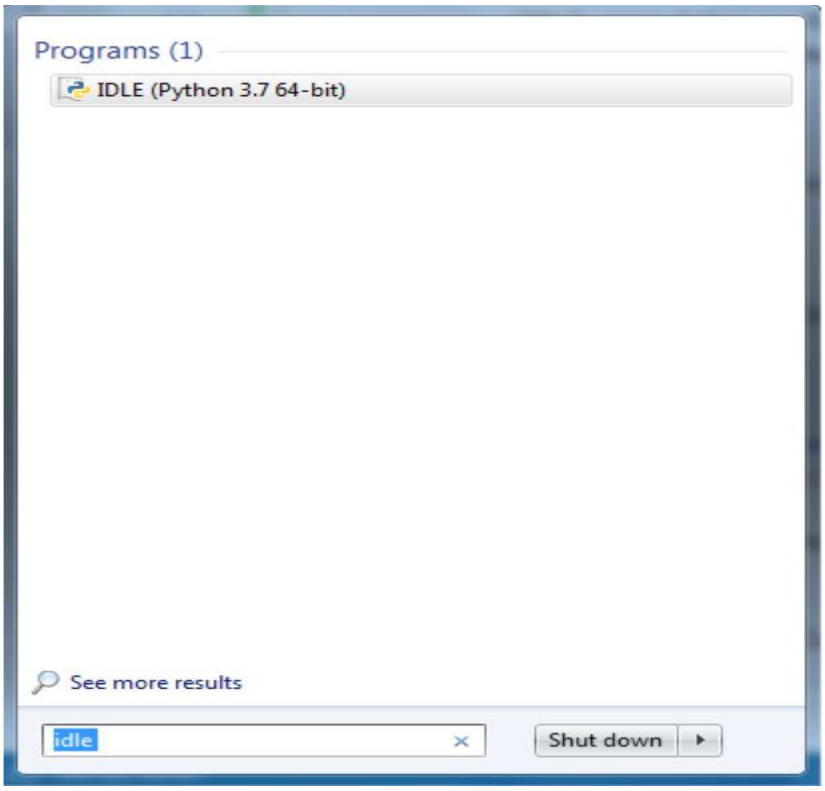


Fig1.1.1 Open Python IDLE [1]

This will open IDLE, where you can write Python code and execute it as shown below.

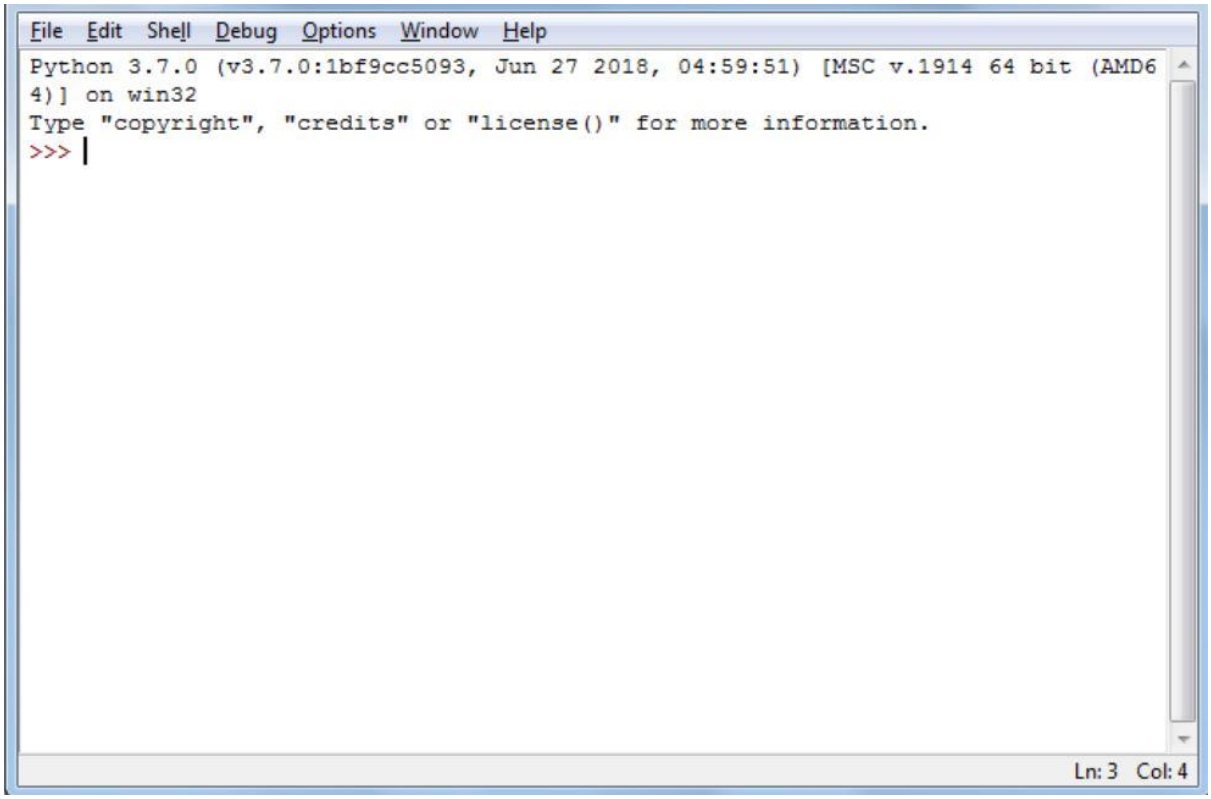


Fig1.1.2 Execute Program on Python IDLE [1]

To execute a Python script, create a new file by selecting File -> New File from the menu.

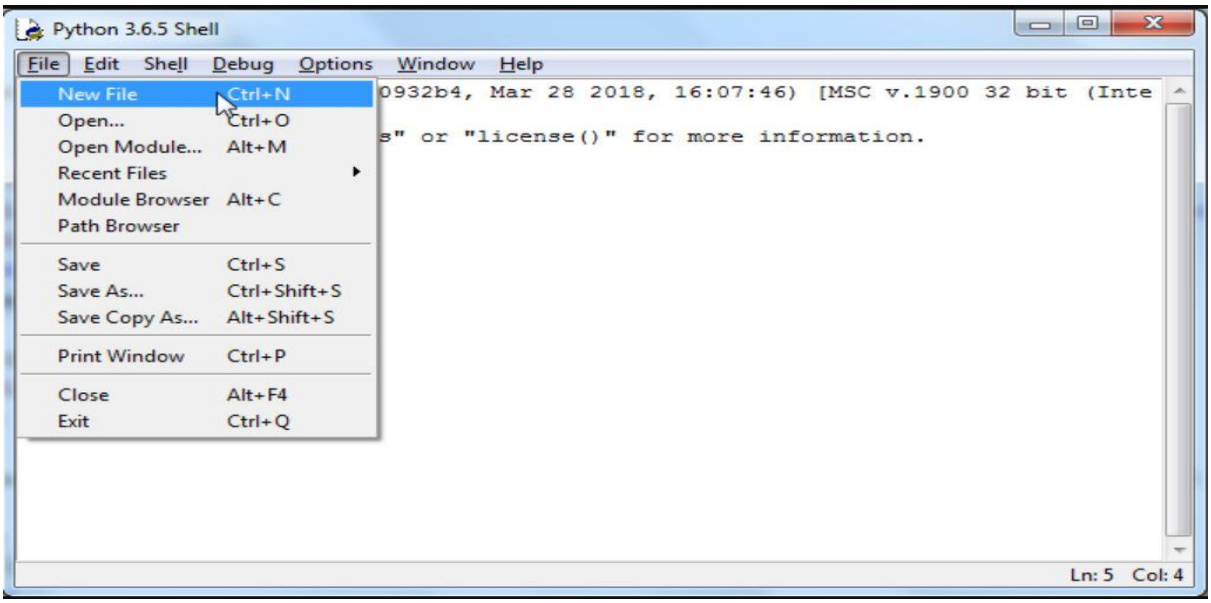


Fig 1.1.3 Save python file [1]

Enter multiple statements and save the file with extension .py using File -> Save. For example, save the following code as hello.py.

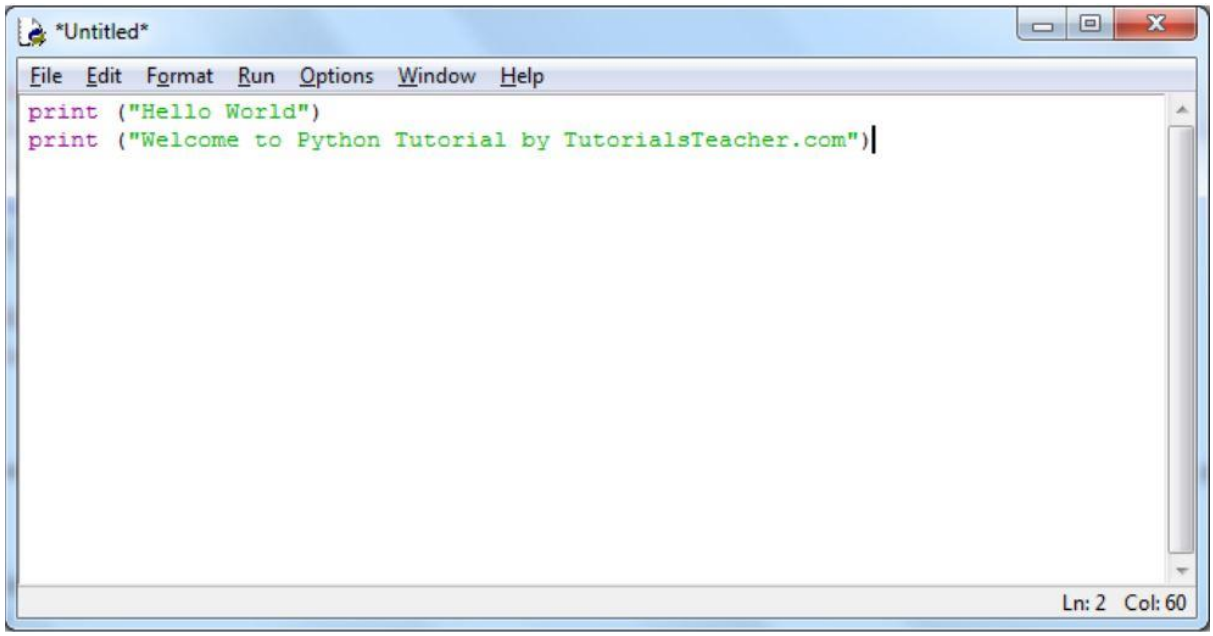


Fig 1.1.4 Python Script in IDLE [1]

Now, press F5 to run the script in the editor window. The IDLE shell will show the output.

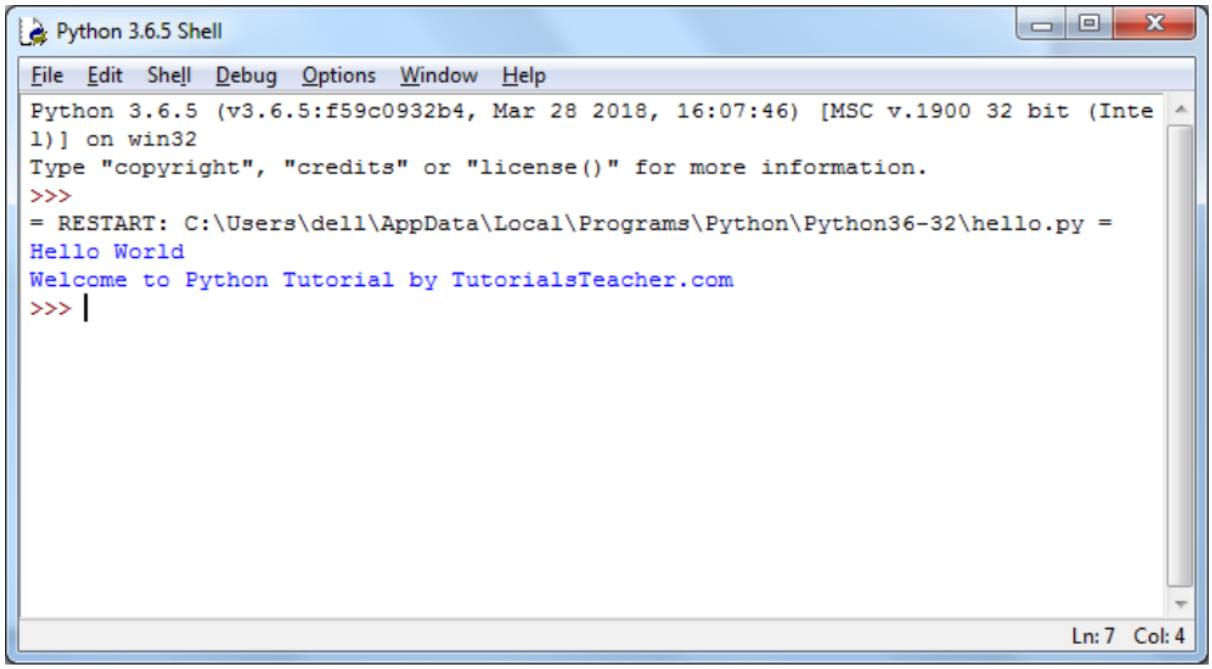


Fig 1.1.5 Python Script Execution Result in IDLE [1]

Thus, it is easy to write, test and run Python scripts in IDLE.

1.2 About Web Scraping

• Web Scraping (also termed Screen Scraping, Web Data Extraction, Web Harvesting etc.) is a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to a local file in your computer or to a database in table (spreadsheet) format.

• Data displayed by most websites can only be viewed using a web browser. They do not offer the functionality to save a copy of this data for personal use. The only option then is to manually copy and paste the data - a very tedious job which can take many hours or sometimes days to complete. Web Scraping is the technique of automating this process, so that instead of manually copying the data from websites, the Web Scraping software will perform the same task within a fraction of the time.

Fig1.2.1 Web Scrapping

Web scraping is used to collect large information from websites. But why does someone have to collect such large data from websites? To know about this, let's look at the applications of web scraping:

• **Price Comparison**: Services such as ParseHub use web scraping to collect data from online shopping websites and use it to compare the prices of products.

• **Email address gathering**: Many companies that use email as a medium for marketing, use web scraping to collect email ID and then send bulk emails.

• **Social Media Scraping**: Web scraping is used to collect data from Social Media websites such as Twitter to find out what's trending.

• **Research and Development**: Web scraping is used to collect a large set of data (Statistics, General Information, Temperature, etc.) from websites, which are analysed and used to carry out Surveys or for R&D.

• **Job listings**: Details regarding job openings, interviews are collected from different websites and then listed in one place so that it is easily accessible to the user.

Web scraping is an automated method used to extract large amounts of data from websites. The data on the websites are unstructured. Web scraping helps collect these unstructured data and store it in a structured form. There are different ways to scrape websites such as online Services, APIs or writing your own code.

1.2.1 How does web Scraping Work?

When you run the code for web scraping, a request is sent to the URL that you have mentioned. As a response to the request, the server sends the data and allows you to read the HTML or XML page. The code then, parses the HTML or XML page, finds the data and extracts it.

To extract data using web scraping with python, you need to follow these basic steps:

1. Find the URL that you want to scrape
2. Inspecting the Page
3. Find the data you want to extract
4. Write the code
5. Run the code and extract the data
6. Store the data in the required format

1.2 Scrape Your Need

The proposed system for this project “Scrap Your Need” is a web scraper that is able to access and extract information of Housing, Education and Job at various locations from websites using a standalone application as an interface for user interaction. The extracted details are then stored in a csv file, as the standalone application allows the user to search through and query the saved findings.

CHAPTER 2

PROJECT REVIEW

2.1 Requirements

2.1.1 Python Modules

2.1.1. Beautifulsoup

Beautiful Soup is a Python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup). It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.

It is available for Python 2.7 and Python 3.

Installing Beautiful Soup

If you're using a recent version of Debian or Ubuntu Linux, you can install Beautiful Soup with the system package manager:

```
$ apt-get install python-bs4 (for Python 2)

$ apt-get install python3-bs4 (for Python 3)
```

Beautiful Soup 4 is published through PyPi, so if you can't install it with the system packager, you can install it with easy_install or pip. The package name is beautifulsoup4, and the same package works on Python 2 and Python 3. Make sure you use the right version of pip or easy_install for your Python version (these may be named pip3 and easy_install3 respectively if you're using Python 3).

```
$ easy_install beautifulsoup4

$ pip install beautifulsoup4
```

2.1.1.2 Request

Requests is a Python module that you can use to send all kinds of HTTP requests. It is an easyto-use library with a lot of features ranging from passing parameters in URLs to sending custom headers and SSL Verification. In this tutorial, you will learn how to use this library to send simple HTTP requests in Python.

Requests allow you to send HTTP/1.1 requests. You can add headers, form data, multipart files, and parameters with simple Python dictionaries, and access the response data in the same way.

Installing the Requests module

To install requests, simply:

```
$ pip install requests
```

Or, if you absolutely must:

```
$ easy_install requests
```

2.1.1.3 Html5lib

html5lib is a pure-python library for parsing HTML. It is designed to conform to the WHATWG HTML specification, as implemented by all major web browsers.

Installation

html5lib works on CPython 2.7+, CPython 3.4+ and PyPy. To install it, use:

```
$ pip install html5lib
```

2.1.1.4 Tkinter

Tk was developed as a GUI extension for the Tcl scripting language by John Ousterhout. The first release was in 1991. Tk proved as extremely successful in the 1990's, because it is easier to learn and to use than other toolkits. So it is no wonder that many programmers wanted to use Tk independently of Tcl. That's why bindings for lots of other programming languages have been developed, including Perl, Ada (called TASH), Python (called Tkinter), Ruby, and Common Lisp.

Tk provides the following widgets:

• button

• canvas

• checkbutton

• combobox

• entry

• frame

• label

• labelframe

• listbox

- menu
- menubutton
- message
- notebook
- tk_optionMenu
- panedwindow
- progressbar
- radiobutton
- scale
- scrollbar
- separator

CHAPTER 3

METHODOLOGY USED

3.1 Project Management

Software Development Life Cycle (SDLC) is a process used by the software industry to design, develop and test high quality software. The SDLC aims to produce a high-quality software that meets or exceeds customer expectations, reaches completion within times and cost estimates.

• SDLC is the acronym of Software Development Life Cycle.
• It is also called as Software Development Process.
• SDLC is a framework defining tasks performed at each step in the software development process.
• ISO/IEC 12207 is an international standard for software life-cycle processes. It aims to be the standard that defines all the tasks required for developing and maintaining software.

What is SDLC?
SDLC is a process followed for a software project, within a software organization. It consists of a detailed plan describing how to develop, maintain, replace and alter or enhance specific software. The life cycle defines a methodology for improving the quality of software and the overall development process.
The following figure is a graphical representation of the various stages of a typical SDLC.

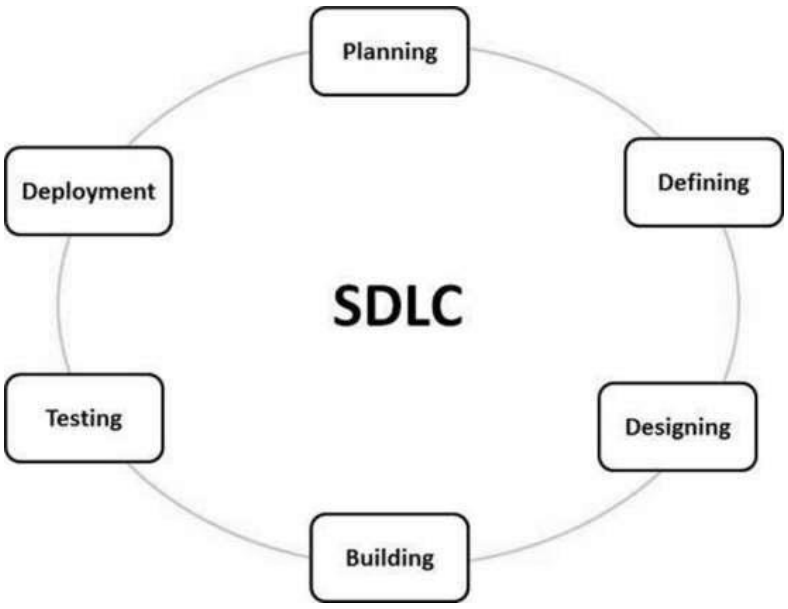


Fig 3.1.1 SDLC

A typical Software Development Life Cycle consists of the following stages -

Stage 1: Planning and Requirement Analysis

Requirement analysis is the most important and fundamental stage in SDLC. It is performed by the senior members of the team with inputs from the customer, the sales department, market surveys and domain experts in the industry. This information is then used to plan the basic project approach and to conduct product feasibility study in the economical, operational and technical areas.

Planning for the quality assurance requirements and identification of the risks associated with the project is also done in the planning stage. The outcome of the technical feasibility study is to define the various technical approaches that can be followed to implement the project successfully with minimum risks.

Stage 2: Defining Requirements

Once the requirement analysis is done the next step is to clearly define and document the product requirements and get them approved from the customer or the market analysts. This is done through an SRS (Software Requirement Specification) document which consists of all the product requirements to be designed and developed during the project life cycle.

Stage 3: Designing the Product Architecture

SRS is the reference for product architects to come out with the best architecture for the product to be developed. Based on the requirements specified in SRS, usually more than one design approach for the product architecture is proposed and documented in a DDS - Design Document Specification.

This DDS is reviewed by all the important stakeholders and based on various parameters as risk assessment, product robustness, design modularity, budget and time constraints, the best design approach is selected for the product.

A design approach clearly defines all the architectural modules of the product along with its communication and data flow representation with the external and third party modules (if any). The internal design of all the modules of the proposed architecture should be clearly defined with the minutest of the details in DDS.

Stage 4: Building or Developing the Product

In this stage of SDLC the actual development starts and the product is built. The programming code is generated as per DDS during this stage. If the design is performed in a detailed and organized manner, code generation can be accomplished without much hassle.

Developers must follow the coding guidelines defined by their organization and programming tools like compilers, interpreters, debuggers, etc. are used to generate the code. Different highlevel programming languages such as C, C++, Pascal, Java and PHP are used for coding. The programming language is chosen with respect to the type of software being developed.

Stage 5: Testing the Product

This stage is usually a subset of all the stages as in the modern SDLC models, the testing activities are mostly involved in all the stages of SDLC. However, this stage refers to the testing only stage of the product where product defects are reported, tracked, fixed and retested, until the product reaches the quality standards defined in the SRS.

Stage 6: Deployment in the Market and Maintenance

Once the product is tested and ready to be deployed it is released formally in the appropriate market. Sometimes product deployment happens in stages as per the business strategy of that organization. The product may first be released in a limited segment and tested in the real business environment (UAT- User acceptance testing).

Then based on the feedback, the product may be released as it is or with suggested enhancements in the targeting market segment. After the product is released in the market, its maintenance is done for the existing customer base.

3.2 CSV

A CSV file (Comma Separated Values file) is a type of plain text file that uses specific structuring to arrange tabular data. Because it’s a plain text file, it can contain only actual text data—in other words, printable ASCII or Unicode characters.

The structure of a CSV file is given away by its name. Normally, CSV files use a comma to separate each specific data value. Here’s what that structure looks like:

column 1 name, column 2 name, column 3 name first row data 1, first row data 2, first row data 3 second row data 1, second row data 2, second row data 3 ... Notice how each piece of data is separated by a

comma. Normally, the first line identifies each piece of data—in other words, the name of a data column. Every subsequent line after that is actual data and is limited only by file size constraints.
In general, the separator character is called a delimiter, and the comma is not the only one used. Other popular delimiters include the tab (\t), colon (:) and semi-colon (;) characters. Properly parsing a CSV file requires us to know which delimiter is being used.

Where Do CSV Files Come From?

CSV files are normally created by programs that handle large amounts of data. They are a convenient way to export data from spreadsheets and databases as well as import or use it in other programs. For example, you might export the results of a data mining program to a CSV file and then import that into a spreadsheet to analyse the data, generate graphs for a presentation, or prepare a report for publication.

CSV files are very easy to work with programmatically. Any language that supports text file input and string manipulation (like Python) can work with CSV files directly.

Parsing CSV Files with Python's Built-in CSV Library

The csv library provides functionality to both read from and write to CSV files. Designed to work out of the box with Excel-generated CSV files, it is easily adapted to work with a variety of CSV formats. The csv library contains objects and other code to read, write, and process data from and to CSV files.
Reading CSV Files
With csv Reading from a CSV file is done using the reader object. The CSV file is opened as a text file with Python's built-in open() function, which returns a file object. This is then passed to the reader, which does the heavy lifting.

CHAPTER 4

SCRAPE YOUR NEED

"Scrape Your Need" is a web scraper that is able to access and extract information of different types of housing, education and jobs at various locations from websites using a standalone application as an interface for user interaction. The extracted details are then stored in a csv file, as the standalone application allows the user to search through and query the saved findings.

4.1. WORKING MODEL

There are mainly two ways to extract data from a website:
• Use the API of the website (if it exists). For example, Facebook has the Facebook Graph API which allows retrieval of data posted on Facebook.
• Access the HTML of the webpage and extract useful information/data from it. This technique is called web scraping or web harvesting or web data extraction.

Steps involved in web scraping:

1. Send a HTTP request to the URL of the webpage you want to access. The server responds to the request by returning the HTML content of the webpage. For this task, we will use a third-party HTTP library for python requests.

2. Once we have accessed the HTML content, we are left with the task of parsing the data. Since most of the HTML data is nested, we cannot extract data simply through string processing. One needs a parser which can create a nested/tree structure of the HTML data.
There are many HTML parser libraries available but the most advanced one is html5lib.

3. Now, all we need to do is navigating and searching the parse tree that we created, i.e. tree traversal. For this task, we will be using another third-party python library, Beautiful Soup. It is a Python library for pulling data out of HTML and XML files.

How user interact with Job Scrapper

When the application is executed you will first be directed to the home screen which have five buttons:

(i) Housing: It directs you to the page where you can search for the house according to the search options of title (as rent and sale), location (Delhi, Gurugram, Bangalore, Pune etc.).

(ii) Education: It directs you to the page where you can search for the college and school in which you location you wanted to find out.

(iii) Job: It directs you to the page where you can search for the job vacancies on the bases of private and government.

(iv) About us: Gives The information about the developers.

(v) Quit: Closes the application.

4.2. MODULE OF SCRAP YOUR NEED

4.2.1. GUI

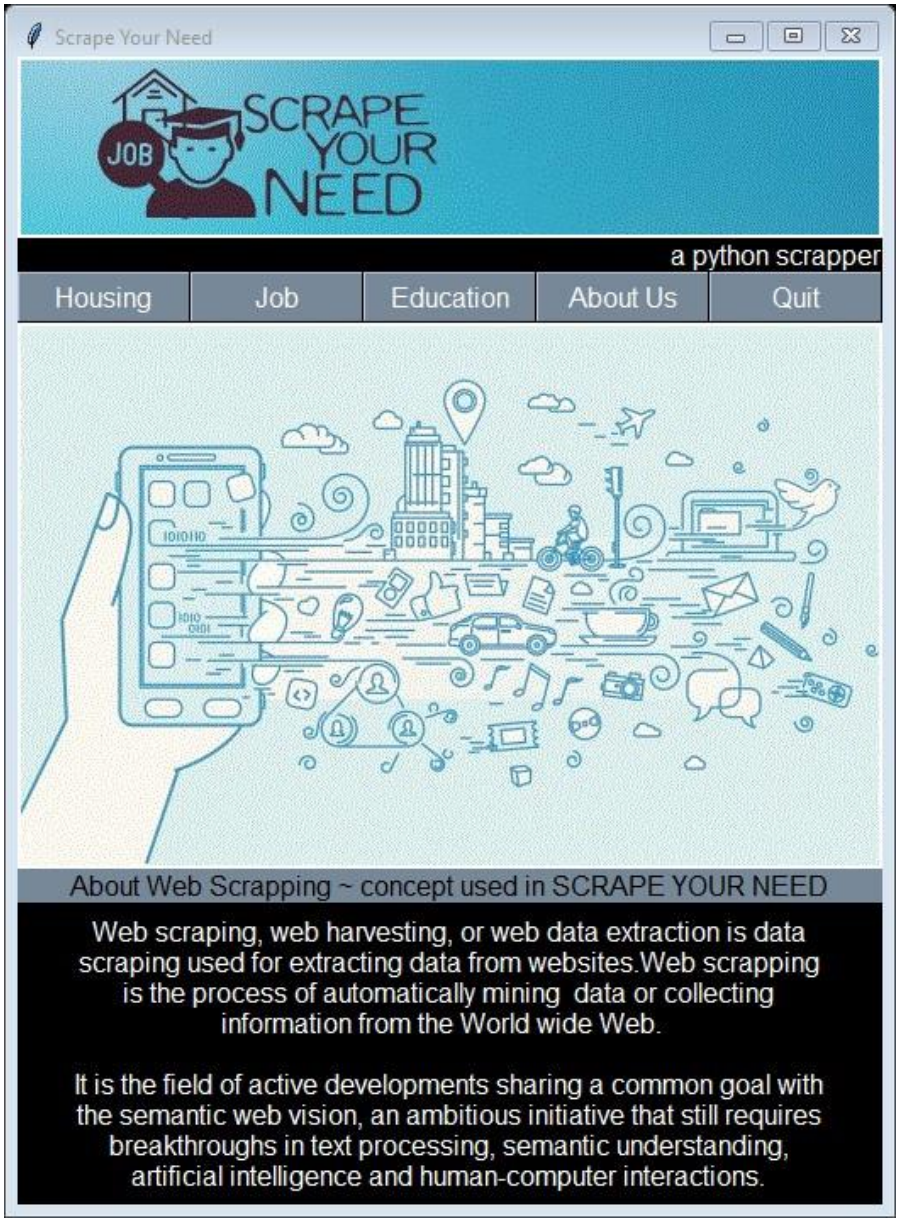


Fig4.2.1.1 Scrape Your Need

Python offers multiple options for developing GUI (Graphical User Interface). Out of all the GUI methods, tkinter is most commonly used method. It is a standard Python interface to the Tk GUI toolkit shipped with Python. Python with tkinter outputs the fastest and easiest way to create the GUI applications. Creating a GUI using tkinter is an easy task. There are a number of widgets which we have used in our application such as Button, Text field, title, label etc.

4.2.2. HOUSING

Owning a house is on the wishlist of most people. This wish, however, can lead to trouble, if people stretch their finances to fulfil it. But postponing the purchase till one accumulates the necessary funds can also be a costly proposition— besides being frustrating.

So, what does one do? Well, there's a third option: affordable housing. A starter home— affordable, smaller—may not offer all that you had fancied in a house, but purchasing it can be a financially sound move.

Hence, here the Housing option of 'Scrape Your Need' gives you the facility to search for your dream house either for rent or sale on the bases of location (Delhi, Gurugram, Bangalore, Pune etc.) of your choice.

4.2.3. EDUCATION

There are many critical issues surrounding access, quality and costs of information and knowledge over the Internet as well as on provision of content and learning material. As it becomes clearer that the growth of Internet offers real opportunities for improving access and transfer of knowledge and information from universities and colleges to a wide range of users.

So here the Education option gives you the flexibility to search for different schools and colleges according to the location (Delhi, Gurugram, Bangalore, Pune etc.).

4.2.4. JOB

There are thousands of job sites on the web, but the best job boards and job search engine sites have search:" # tools that are quick and easy to use and allow you to search based on the type of job you're looking for, your location, and other criteria.

There are also sites that focus on certain types of positions or match you with employers. These sites are worth incorporating into your job search, because not all employers list on every website, even though it may seem that way. Don't limit yourself to just one job website, because each job site only lists jobs from particular websites or companies.

It directs you to the page where you can search for the private and government job vacancies, according to the search option location (Delhi, Gurugram, Bangalore, Pune etc.).

4.2.5. SCRAPPING SCRIPT

Web scraping, also called web data mining or web harvesting, is the process of constructing an agent which can extract, parse, download and organize useful information from the web automatically.

In this Module we prepare a programming script in with we prettify the scrapped URL with the help of BeautifulSoup. we had written three scripts for three different job searching Platform that is Indeed, Simply Hired and Times job.

In script first we prettify the data or information that we scrapped and the select the tags that we want to scarp for our application and finally the Scrapped data is stored in the form of CSV (comma separated value) file.

4.2.6. RESULT MODULE

In this Module the scrapped data which is stored on csv file is being fetched to the show result page in the form of table with the attributes that is we scrapped from the sites such as:

- **JOB**
  - Private
    - Title , Organization , Posted , Location , Description
  - Government
    - Title , Organization , Posted , Location , Description
- **HOUSING**
  - Rent
    - Title , Location , Price , Detail
  - Sale
    - Title , Location , Price , Detail
- **EDUCATION**
  - School
    - Name , Location
  - College
    - Name , Location , Course , Degree

CHAPTER 5

RESULT

5.1. HOME SCREEN

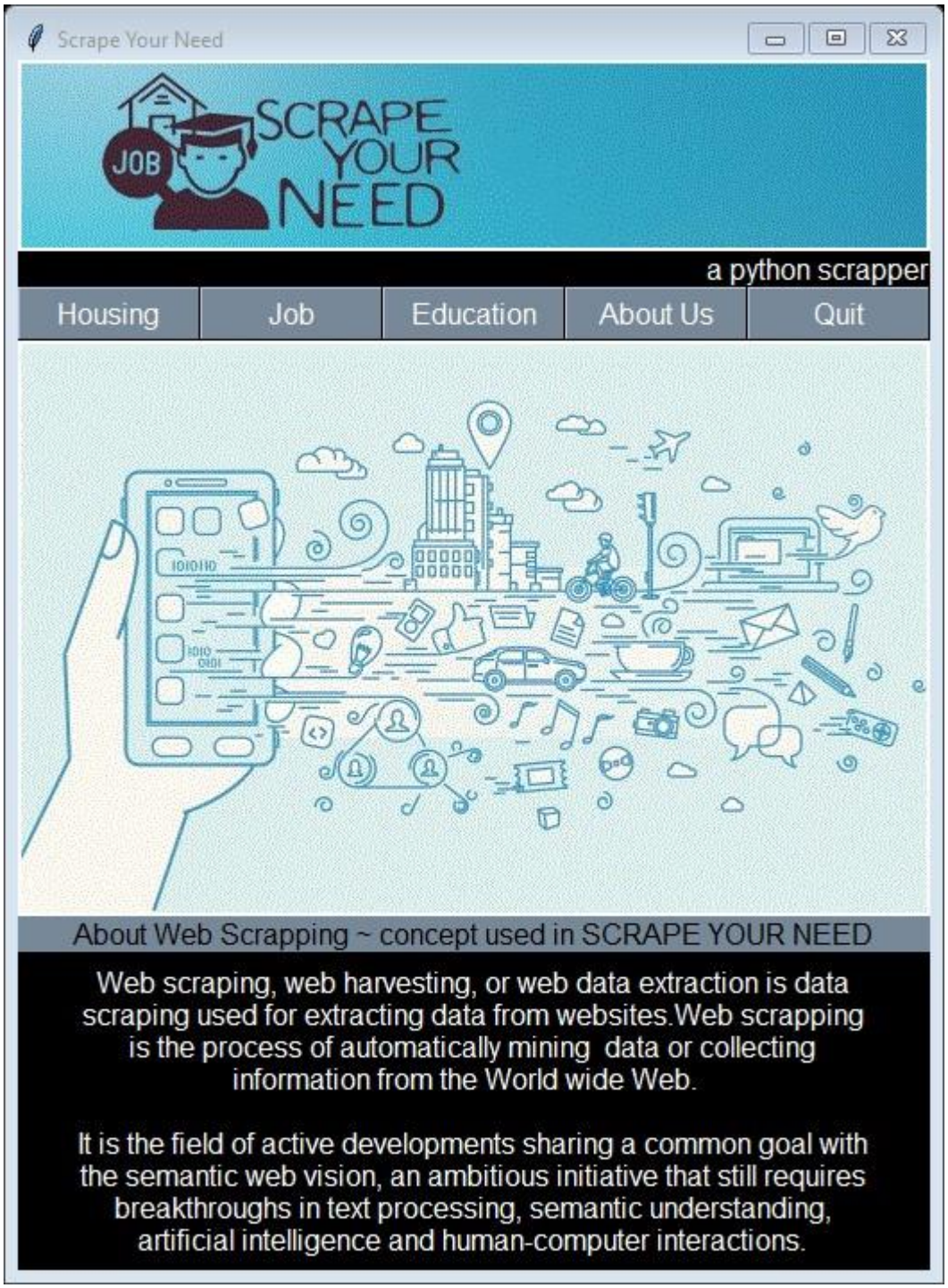


Fig 5.1.1 Home Screen

5.2. HOUSING

Fig5.2.1 Housing (Rent)

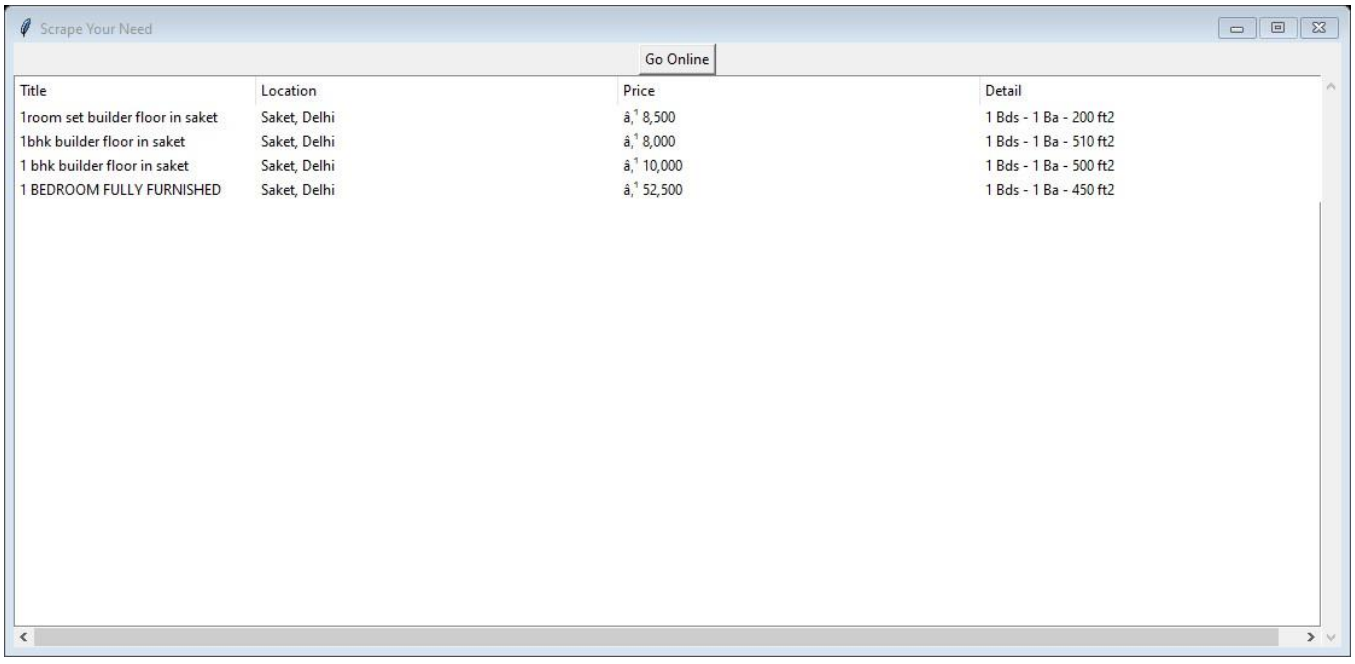


Fig5.2.2 Housing Result (Rent)

Fig5.2.3 Housing (Sale)

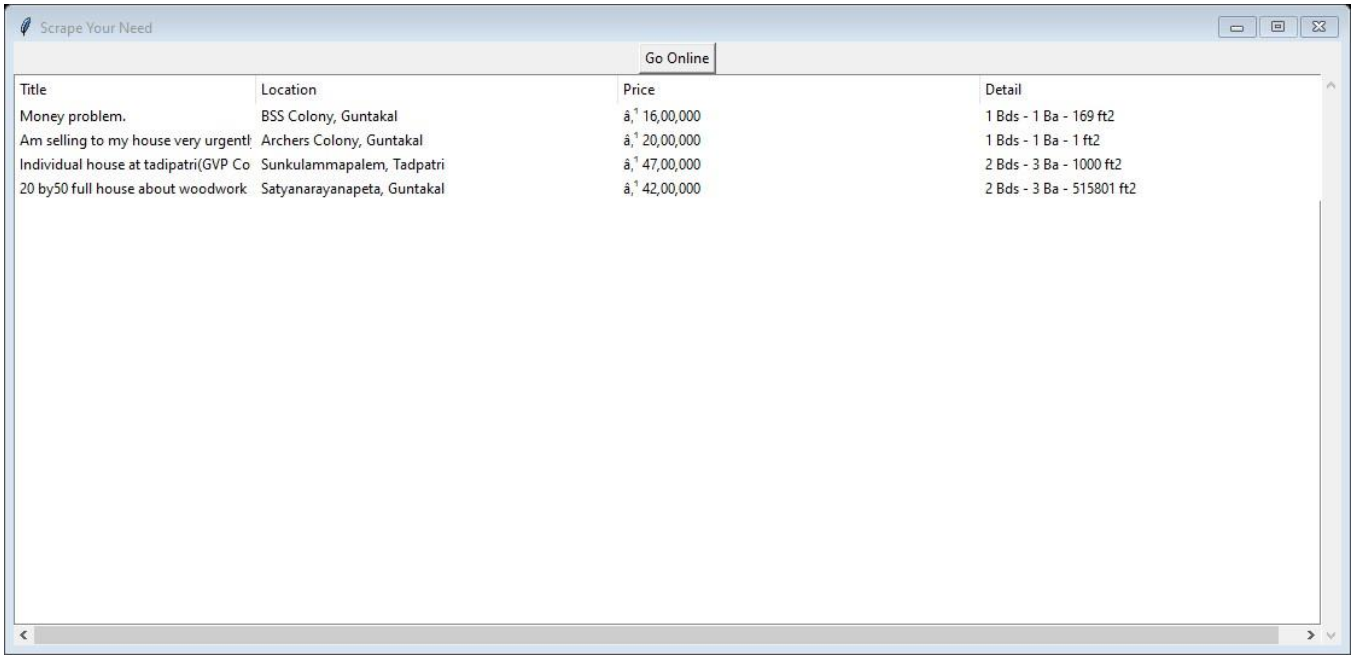


Fig5.2.4 Housing Result (Sale)

5.3. EDUCATION

Fig5.3.1 Education (College)

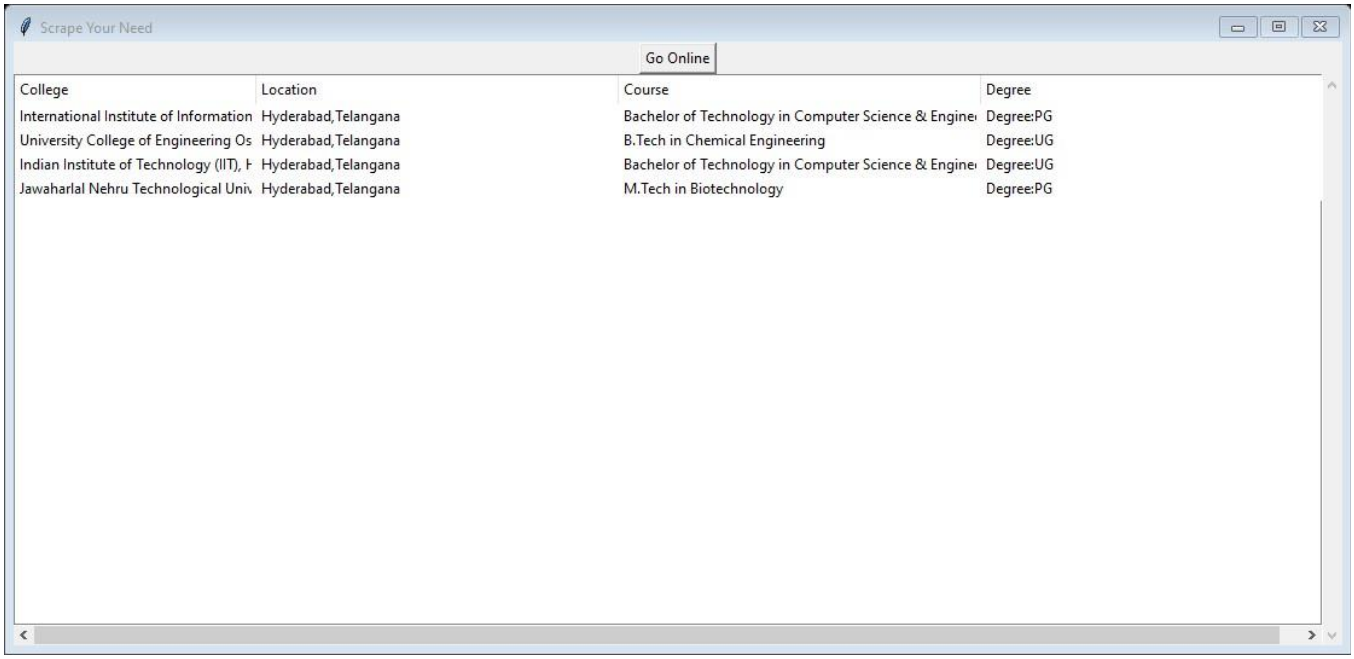


Fig5.3.2 Education Result (College)

Fig5.3.3 Education (School)

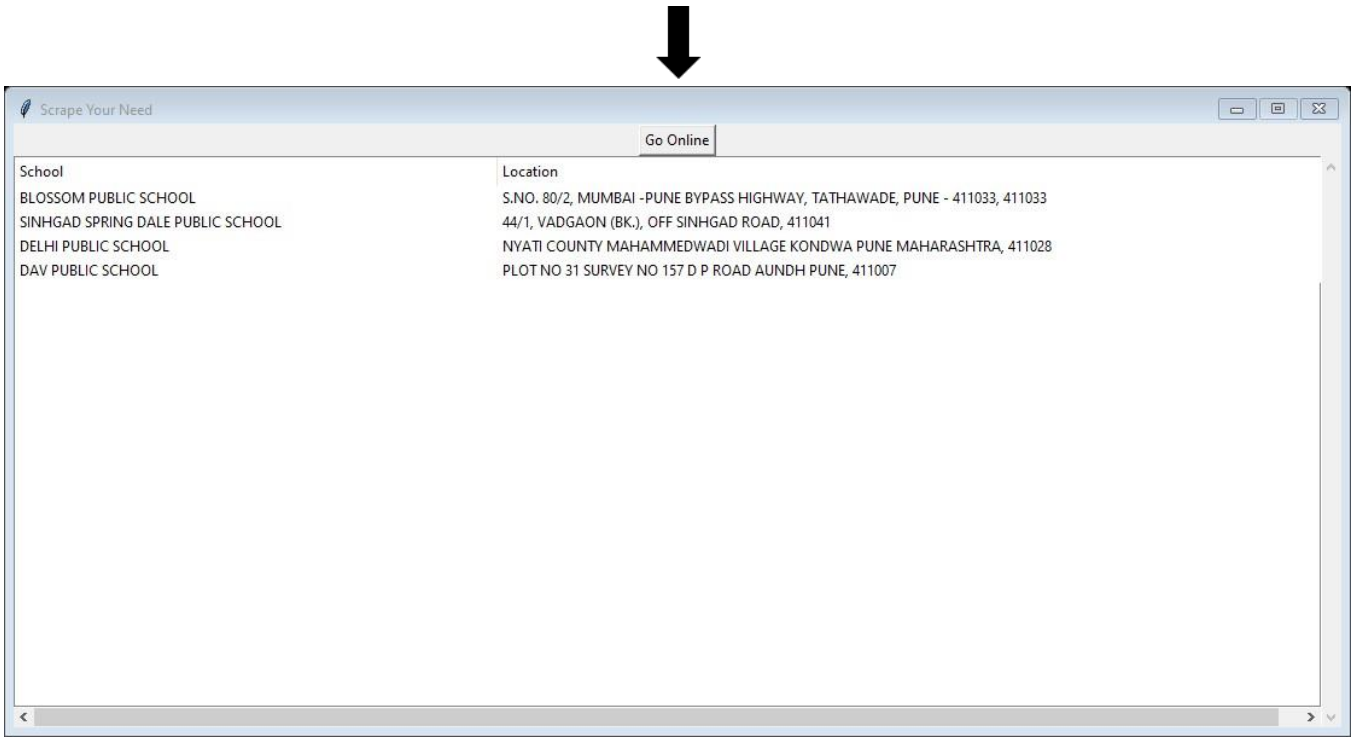


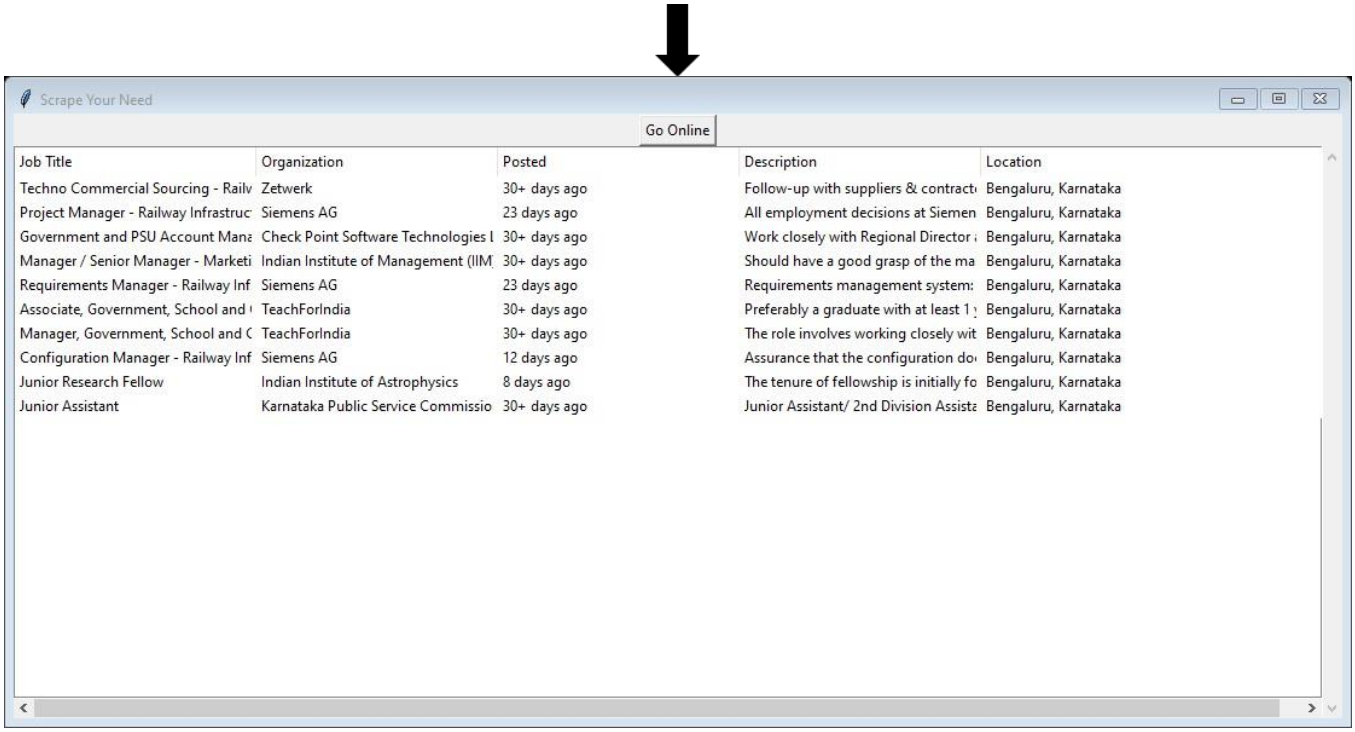
Fig5.3.4 Education Result (School)

There are many critical issues surrounding access, quality and costs of information and knowledge over the Internet as well as on provision of content and learning material. As it becomes clearer that the growth of Internet offers real opportunities for improving access and transfer of knowledge and information from universities and colleges to a wide range of users.

So here the Education option gives you the flexibility to search for different schools and colleges according to the location (Delhi, Gurugram, Bangalore, Pune etc.).

5.4. JOB

Fig 5.4.1 Job (Government)



Job Title	Organization	Posted	Description	Location
Techno Commercial Sourcing - Railv	Zetwerk	30+ days ago	Follow-up with suppliers & contract	Bengaluru, Karnataka
Project Manager - Railway Infrastruc	Siemens AG	23 days ago	All employment decisions at Siemen	Bengaluru, Karnataka
Government and PSU Account Mana	Check Point Software Technologies l	30+ days ago	Work closely with Regional Director	Bengaluru, Karnataka
Manager / Senior Manager - Marketi	Indian Institute of Management (IIM	30+ days ago	Should have a good grasp of the ma	Bengaluru, Karnataka
Requirements Manager - Railway Inf	Siemens AG	23 days ago	Requirements management system:	Bengaluru, Karnataka
Associate, Government, School and	TeachForIndia	30+ days ago	Preferably a graduate with at least 1	Bengaluru, Karnataka
Manager, Government, School and C	TeachForIndia	30+ days ago	The role involves working closely wit	Bengaluru, Karnataka
Configuration Manager - Railway Inf	Siemens AG	12 days ago	Assurance that the configuration do	Bengaluru, Karnataka
Junior Research Fellow	Indian Institute of Astrophysics	8 days ago	The tenure of fellowship is initially fo	Bengaluru, Karnataka
Junior Assistant	Karnataka Public Service Commissio	30+ days ago	Junior Assistant/ 2nd Division Assista	Bengaluru, Karnataka

Fig 5.4.2 Job Result (Government)

Fig 5.4.3 Job (Private)

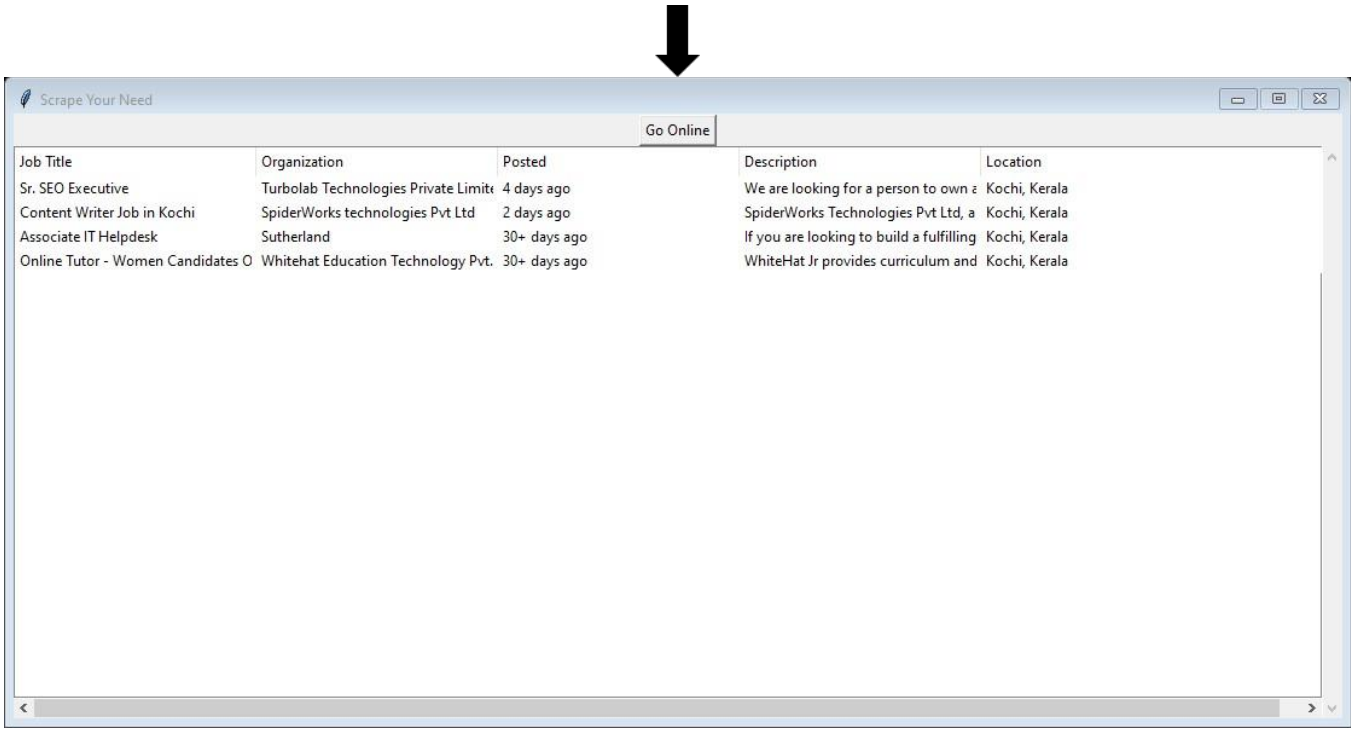


Fig 5.4.4 Job Result (Private)

There are thousands of job sites on the web, but the best job boards and job search engine sites have search:" # tools that are quick and easy to use and allow you to search based on the type of job you're looking for, your location, and other criteria.

There are also sites that focus on certain types of positions or match you with employers. These sites are worth incorporating into your job search, because not all employers list on every website, even though it may seem that way. Don't limit yourself to just one job website, because each job site only lists jobs from particular websites or companies.

It directs you to the page where you can search for the private and government job vacancies, according to the search option location (Delhi, Gurugram, Bangalore, Pune etc.).

CHAPTER 6

CONCLUSION

Web scraping services provide an essential service at a low cost. It is paramount that data is collected back from websites and analyzed so that the internet functions regularly. Web scraping services do the job in an efficient and budget friendly manner.

Once a web scraping services deploys the proper mechanism to extract data, you are assured that you are not only getting data from a single page but from the entire domain. This means that with just a onetime investment, a lot of data can be collected.

One aspect that is often overlooked when installing new services is the maintenance cost. Long term maintenance costs can cause the project budget to spiral out of control. Thankfully, web scraping technologies need very little to no maintenance over a long period. Another characteristic that must also be mentioned is the speed with which web scraping services do their job. A job that could take a person week is finished in a matter of hours.

The web scraping services are not only fast, they are accurate too. Simple errors in data extraction can cause major mistakes later on. Accurate extraction of any type of data is thus very important. In websites that deal in pricing data, sales prices, real estate numbers or any kind of financial data, the accuracy is extremely important.

Needs like Housing, education and Jobs displayed by different websites can only be viewed using a web browser. Scrape Your Need is the Standalone Application of automating this process that is, it scraps these needs from different websites and present it on the single platform which saves time and increase the opportunities for an individual.

CHAPTER 7

FUTURE SCOPE

Online network, which provides users with a central platform for classified information across different cities of India. Some of the items included in this are: for Housing, Education and Job. Users are able to view all the classified information of different type.

In Scrape your Need, till now we are providing services like housing, job and Education. With future perspective we will try to inculcate the services like Automotive, Household, Financial etc.

Right now, we have bounded our application Scrape Your Need to main five metropolitan cities of India (Delhi, Hyderabad, Bangalore, Pune, Kochi etc.). In future we can expand its scope even to the non-metropolitan cities like Ahmadabad, Bhubaneshwar, Jaipur etc.

REFERENCES

[1] https://www.tutorialsteacher.com/python/python-idle

[2] https://www.geeksforgeeks.org/implementing-web-scraping-python-beautiful-soup/

[3] https://www.guru99.com/web-scraping-tools.html

[4] https://www.geeksforgeeks.org/python-tools-world-web-scraping/

[5] https://realpython.com/tutorials/web-scraping/

[6] Kory Kroft and Devin G. Pope, "Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist"

[7] D. Pratiba ; Abhay M.S. ; Akhil Dua ; Giridhar K. Shanbhag ; Neel Bhandari ; Utkarsh Singh, "Web Scrapping and Data Acquisition Using Google Scholar", International Journal of Computer Applications (0975 – 8887) Volume 179 – No.20, February 2018

[8] https://www.geeksforgeeks.org/introduction-to-web-scraping/

[9] Google.com