A project report on

DATA SCRAPPING IN REAL ESTATE HOUSING INDUSTRY

Submitted in partial fulfilment of the requirement

for the award of the degree of

MASTER OF COMPUTER APPLICATIONS

Of

Visvesvaraya Technological University Belgaum, Karnataka

By

JYOTHSNA S 1CR18MCA65

Under the guidance of

Internal Guide	External Guide
Dr. A. Abdul Rasheed MCA Department, Institute of Technology,	Mr. Himanshu Mangaraj Professor, Real Time Signal Technologies, CMR Bangalore. Bangalore.

CMR INSTITUTE OF TECHNOLOGY

132, IT Park Road, Kundalahalli, Bangalore-560037

2019-2020

A project report on

DATA SCRAPPING IN REAL ESTATE HOUSING INDUSTRY

Submitted in partial fulfilment of the requirement

for the award of the degree of

MASTER OF COMPUTER APPLICATIONS

Of

Visvesvaraya Technological University Belgaum, Karnataka

By

JYOTHSNA S
1CR18MCA65

Under the guidance of

Internal Guide

Dr. A. Abdul Rasheed
MCA Department,
Institute of Technology,

External Guide

Mr. Himanshu Mangaraj Professor,
Real Time Signal Technologies, CMR
Bangalore. Bangalore.

CMR INSTITUTE OF TECHNOLOGY

132, IT Park Road, Kundalahalli, Bangalore-560037

2019-2020

CMR INSTITUTE OF TECHNOLOGY

Department of Master of Computer Applications

Bangalore - 560 037

CERTIFICATE

This is to certify that the project work entitled

DATA SCRAPPING IN REAL ESTATE TRANSFORMATION

Submitted in partial fulfilment of the requirement
for the award of the degree of
Master of Computer Applications of the
Visvesvaraya Technological University, Belgaum, Karnataka
bonafide work carried out by

JYOTHSNA S
1CR18MCA65

during the academic year 2019-20120.

Signature of the Guide	**Signature of the HOD**	**Signature of the Principal**
Dr. A. Abdul Rasheed	**Ms.Gomathi.T**	**Dr. Sanjay Jain**
Professor, MCA	HOD, MCA	PRINCIPAL, CMRIT
	External Viva	

Name of the Examiners — Signature with date

1.
2.

DECLARATION

I, **JYOTHSNA S,** student of 6th MCA, **CMR Institution of Technology**, bearing the USN **1CR18MCA65**, hereby declare that the project entitled **"DATA SCRAPPING IN REAL ESTATE TRANSFORMATION"** has been carried out by me under the supervision of External Guide **Mr. Himanshu Mangaraj**, Project Manager**,** and Internal Guide **Dr. A. Abdul Rasheed**, **Professor**, **Dept. of Master of Computer Applications** and submitted in the partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications by the **Visvesvaraya Technological University** during the academic year 2019-2020.The reports has not been submitted to any other University or Institute for the award of any degree or certificate.

Place: Bangalore JYOTHSNA S

Date: (1CR18MCA65)

ACKNOWLEDGEMENT

I would like to thank all those who are involved in this endeavour for their kind cooperation for its successful completion. At the outset, I wish to express my sincere gratitude to all those people who have helped me to complete this project in an efficient manner.

I offer my special thanks to my external project guide Mr. Himanshu Mangaraj Project Manager, Real Time Signals Technologies Pvt. Ltd., Bangalore, and to my Internal Project guide Dr. A. Abdul Rasheed, Department of MCA, CMRIT, Bangalore without whose help and support throughout this project would not have been this success.

I am thankful to Dr. SANJAY JAIN, Principal, CMRIT, Bangalore for his kind support in all respect during my study. I would like to thank Mr. Himanshu Mangaraj, Project Manager, Real Time Signals Technologies Pvt. Ltd., Bangalore, who gave opportunity to do this project at an extreme organization Most of all and more than ever, I would like to thanks my family members for their warmness, support, encouragement, kindness and patience. I am really thankful to all my friends who always advised and motivated me throughout the course.

JYOTHSNA S
(1CR18MCA65)

Real Time Signals Technologies

CERTIFICATE OF COMPLETION

We hereby conform that **Ms. Jyothsna. S** of your collage **CMR Institute Of Technology** with USN: **1CR18MCA65** has successfully completed the Project at **Real Time Signals Technologies Pvt. Ltd.** from January 27-2020 to May 30-2020.

The Project based on Web Development with **"Data Scrapping In Real Estate Transform Housing Industry"** under the guidance of **Mr. Himanshu**, Project Guide, **Real Time Signals Technologies Pvt. Ltd.**

Real Time Signals Technologies Pvt. Ltd.

Kind Regards,

Project Manager

REAL TIME SIGNALS TECHNOLOGIES PVT.LTD

Krishna Grand, Over Marthahalli Bridge, Bengaluru, Karnataka 560037

Ph: 080-42008777, 9686939421, Email info@realtimesig.com

Website www.realtimesig.com

TABLE OF CONTENTS

1.INTRODUCTION

Web Scraping is a procedure wherein a ton of data is isolated from destinations. The data which is removed using web scratching is saved to a close by archive in your PC or to a database in table (spreadsheet) structure. Web scraping can likewise be called as web reaping or web information extraction. It is a procedure of separating information from different information distribution centres and assets.

Web scratching is the absolute first thing that you're doing once you got to gather legion various forms of data. Every business needs the newest possible data to know all the trends, possible opportunities, what's going on in the market, etc. So here are some samples of how web scraping is used:

Market Research: In the field of market researching, web scraping finds its use in gathering valuable data about the market, competitors, target audience and its habits, data about social networks and the ways to use it for the advantage, etc.;

Price Scraping: companies scrape prices in order that they could know what's the value of an equivalent products provided by the opposite companies, retail sites, etc. so they could put the best possible price in the market and attract more customers;

SEO: Every person who has worked with online business knows that SEO matters. Web scraping can help to collect key data to enhance SEO and have a far better engagement also on rank within the top of search engines;

Sales Intelligence: deals insight in straightforward terms is the wide scope of innovations that help sales reps discover, screen, and comprehend data on possibilities' and existing customers' day by day business. Web scraping helps to collect useful data for this purpose and make necessary changes;

These are just a couple among the humongous applications where web scraping can profit business and help it to develop. Since there is a ton of interest for information for statistical surveying, value knowledge or contender investigation, and so forth the interest for mechanizing the way toward scraping the information has additionally developed. This is the spot web scratching turns into a basic factor.Web scraping is the mechanized procedure of scraping the information from your preferred web in a configuration.

Why web scraping has become so basic is a direct result of a lot of components. Right off the bat, the information that you access on the Internet isn't accessible for download. Be that as it may, you need it downloaded and in an alternate organization. Thus, you need an approach to download the information from numerous pages of a site or from various sites. Along these lines, you need web scraping.

Web scraping is likewise required in light of the fact that you have no ideal opportunity to worry over how to download, duplicate, spare the information that you see on a website page. What you need is a simple, robotized method of scraping whatever information that you see on the site page and thus web scraping! What web scraping does so well separate from giving you the information that you need is that it spares you many worker hours that you will in any case need on the off chance that you attempt to physically get the information.

On occasion, there is no API from the source site and consequently, web scraping is the best way to remove the information.

1.1 PROJECT DESCRIPTION:

WHY WEB SCRAPING IS USED?

Web scratching is the arrangement of thusly slithering and data downloading from goals and emptying unstructured or around sifted through information into a created structure. Web scraping is a procedure of computerizing the extraction of information in an effective and quick manner. With the help of web scraping, you can separate information from any site, paying little heed to how enormous the information may be, on your PC.

Suppose, information is huge for your internet business organization. You can see the information on your rival's site. The inquiry is by what method will you download it in a usable configuration? The vast majority would be prepared to just reorder it physically. Be that as it may, it isn't possible to do it for enormous sites with hundreds of pages. This is where web scraping comes into play.

Also, sites may have information that you can't reorder. Web scratching can assist you with extricating any sort of information that you need. That is insufficient. Suppose, you reorder a few information yet how to change over or spare it in an arrangement of your decision? Web scratching deals with that as well. At the point when you remove web information with the assistance of a web scratching as well, you would have the option to spare the information in an organization, for example, CSV. You would then have the option to recover, examinations, and utilize the information the manner in which you need. Thus, web scratching disentangles the way toward separating information, speeds it up via mechanizing it, and makes simple access to the rejected information by giving it in the ideal configuration. In basic terms, web scratching spares you the trouble of physically downloading or duplicating any information and computerizes the whole procedure.

GETTING READY:

Before we find the opportunity to mix building up the scratching contraptions, first we have to set up our progress condition. The fundamentals we require are as indicated by the going with:

1) An Integrated improvement condition (IDE) for making the code and directing endeavors. PHP is the programming language we are using, for executing our code.

2) MySQL the database for managing our scratched data.

3) PhpMyAdmin to coordinate relationship of databases. PHP, MySQL, and phpMyAdmin can in like manner be presented uninhibitedly. In any case, we will present the XAMPP pack, which joins those, nearby by further programming, for instance, the Apache Server, which will come satisfying later in case in case you develop your scrubber further. In the wake of presenting these instruments, we'll change the basic structure settings and the test that everything is working totally.

SETUP:

Presently, we take a look at the best approach to set up our improvement condition, by playing out the accompanying advances:

1. During this previously set of steps, we'll introduce our improvement condition, Zend Eclipse PDT.

Select Zend Eclipse PDT download choice for your working framework, as appeared in the screen capture,and spare the ZIP document to your PC.

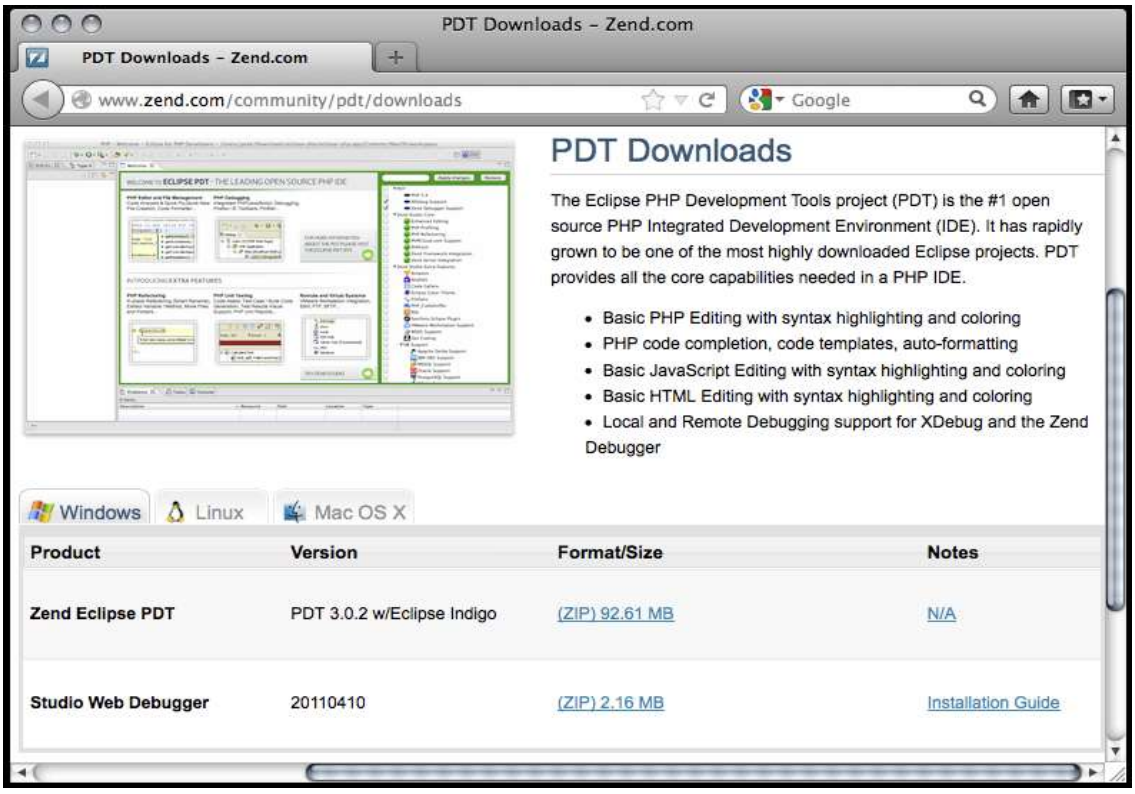


Figure 1

2. When the document has been downloaded, unfasten the substance. The subsequent catalog, overshadow php, is that the shroud program organizer. Simplified this into the C:\Program Files registry on your PC.
3. Next, we'll present XAMPP, which wires PHP, MySQL, phpMyAdmin, and Apache.

Visit the subsequent URL and you can download the most cutting-edge type of XAMPP, clinging to the foundation bearings on the online page http://www.apachefriends.org/en/xampp-windows.html, as showed up inside the going with screen catch:

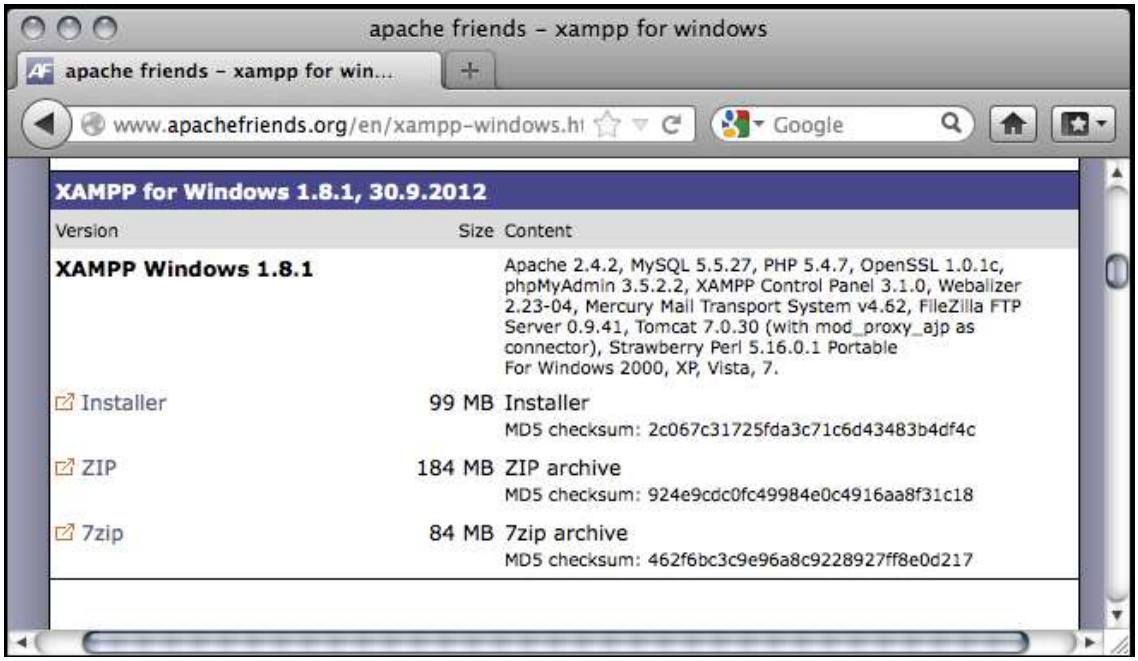


Figure 2

4. Upon fruitful establishment, start XAMPP just because and select the accompanying segments to introduce:

5. ● XAMPP – XAMPP Desktop Icon
6. ● Server – MySQL, Apache
7. ● Program Languages – PHP
8. ● Tools – phpMyAdminSave in the default destination.

Snap on Install and the picked projects will introduce.

Double tap on the XAMPP work area symbol to dispatch the XAMPP control board.

In te XAMPP control board start Apache and MySQL by playing out the following arrangement of steps.

Snap on the Start button for Apache. Snap on the start button for MySQL

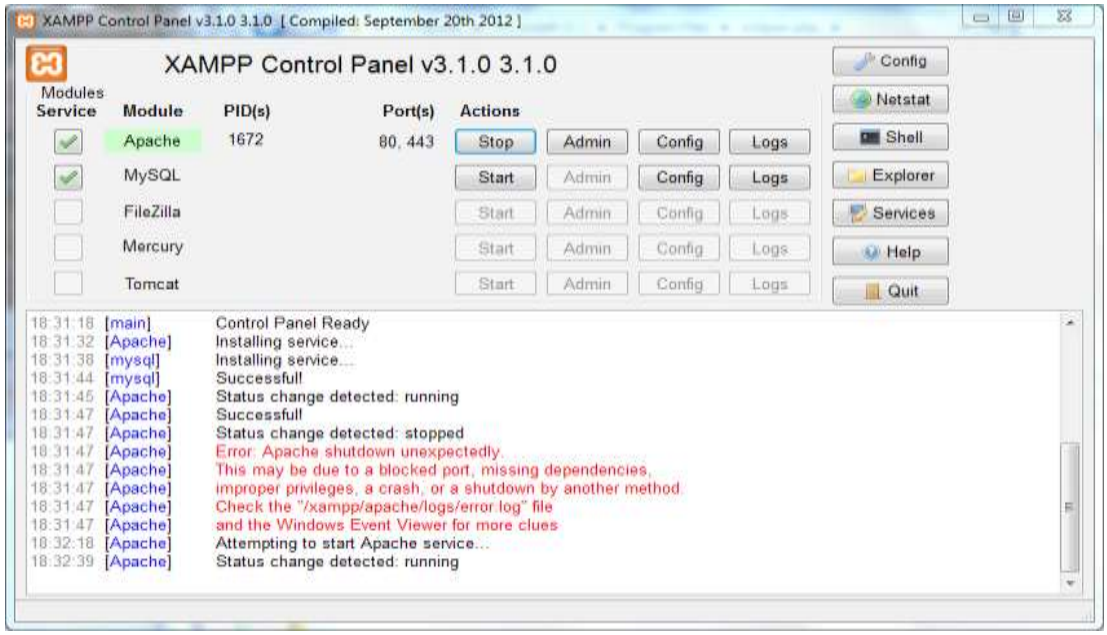


Figure 3

With the important programming and devices presented, we'd want to line our PHP way factor.

Investigate to Start | Control Panel | System and Security | System.

In the left menu bar click on Advanced system settings.

In the System Properties window select the Advanced tab, and snap on the Environment factors... button

In the Environment Variables window there are two records, User elements and System factors. In the System factors list, look down to the line for the Path variable. Select the line and snap on the Edit button.

In the textbox for variable's worth: add to the uttermost furthest reaches of the line the vault where PHP is presented, C:\xampp\php, and a while later snap on OK, as given in the going with screen catch:

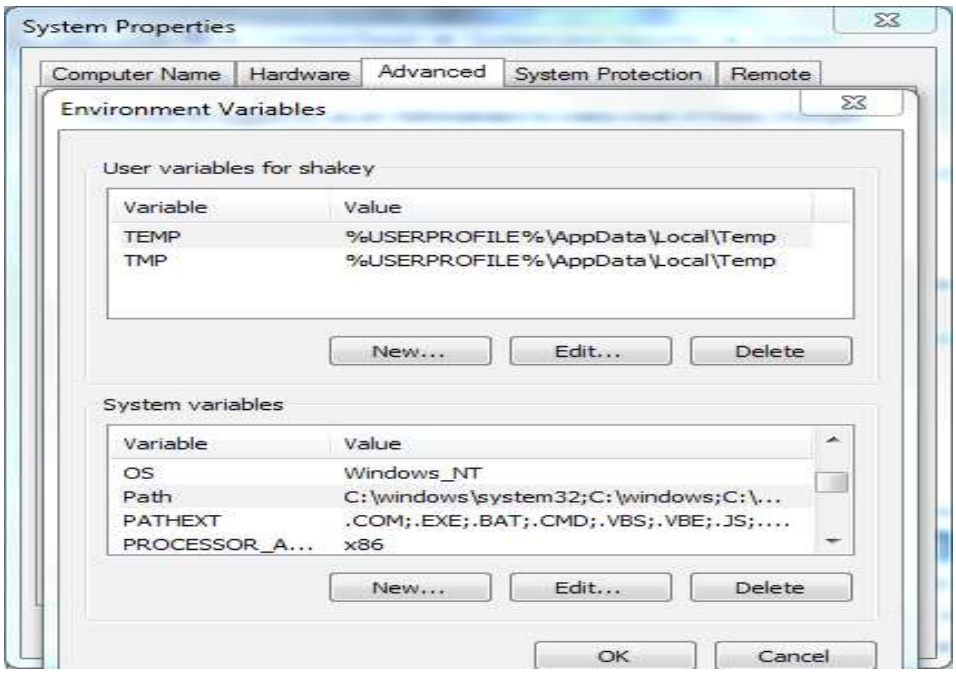


Figure 4

9. Presently, we will have the PHP catalog in our way factors.
10. At long last, we'd prefer to ensure that URL is empowered in PHP. Explore to our XAMPP establishment catalog, at that point into the PHP index and open the record php.ini for altering.
11. Locate the accompanying line and expel the semicolon from the earliest starting point of it extension=php_curl.dll
12. Spare the record and close the word processor.
13. In the XAMPP control board, restart Apache.
14. We would now be able to test whether the establishment is working accurately by opening our internet browser and visiting http://localhost/xampp/status.php or

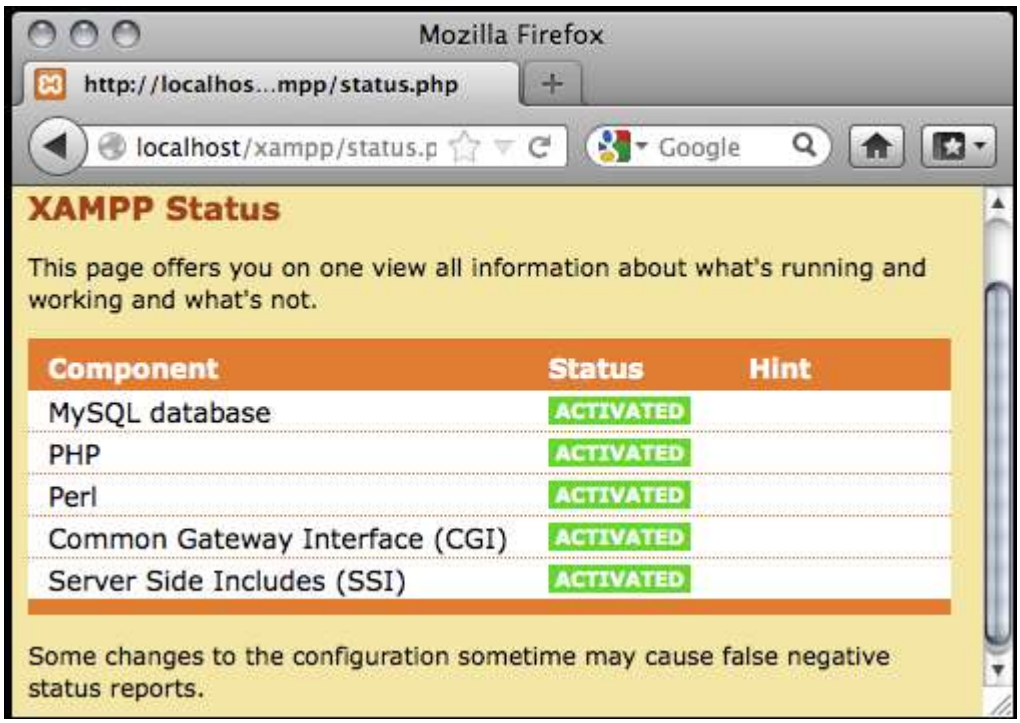


Figure 5

15. A definitive advance is to make a substitution venture in Eclipse and execute our program.
16. We start Eclipse by exploring to the envelope in which we spared it before and double tapping on the shroud php symbol.
17. We are approached to choose our Workspace. Peruse to our xampp registry and afterward explore to htdocs, for instance C:\xampp\htdocs and snap on OK.
18. At the point when the Eclipse begins, Go to File > New > PHP Project. Let the entirety of the settings stay as they are and name our venture as Web Scraping. Snap on Next, and afterward click on Finish.
19. Presently we can compose our first content and execute it. Explore to File | New | PHP File, leave the source organizer as Web Scraping and name the PHP

WORKING:

Let's look at how we performed the previously defined steps in detail:

1. After the product is introduced, the PHP way factor ought to be arrangement. This guarantees we can execute PHP straightforwardly from the guidance by composing php as opposed to composing the total area of our PHP executable record, at whatever point we wish to execute it.

2. Within the following stage we ensure that whether cURL is empowered in PHP. Twist is the library which we'll be utilizing to ask for and download target destinations

3. We at that point watch that everything is introduced accurately by visiting the XAMPP status page.

4. Using a definitive arrangement of steps, we discovered Eclipse, at that point make alittle PHP program which echoes the content Hello world! to the screen and execute it.

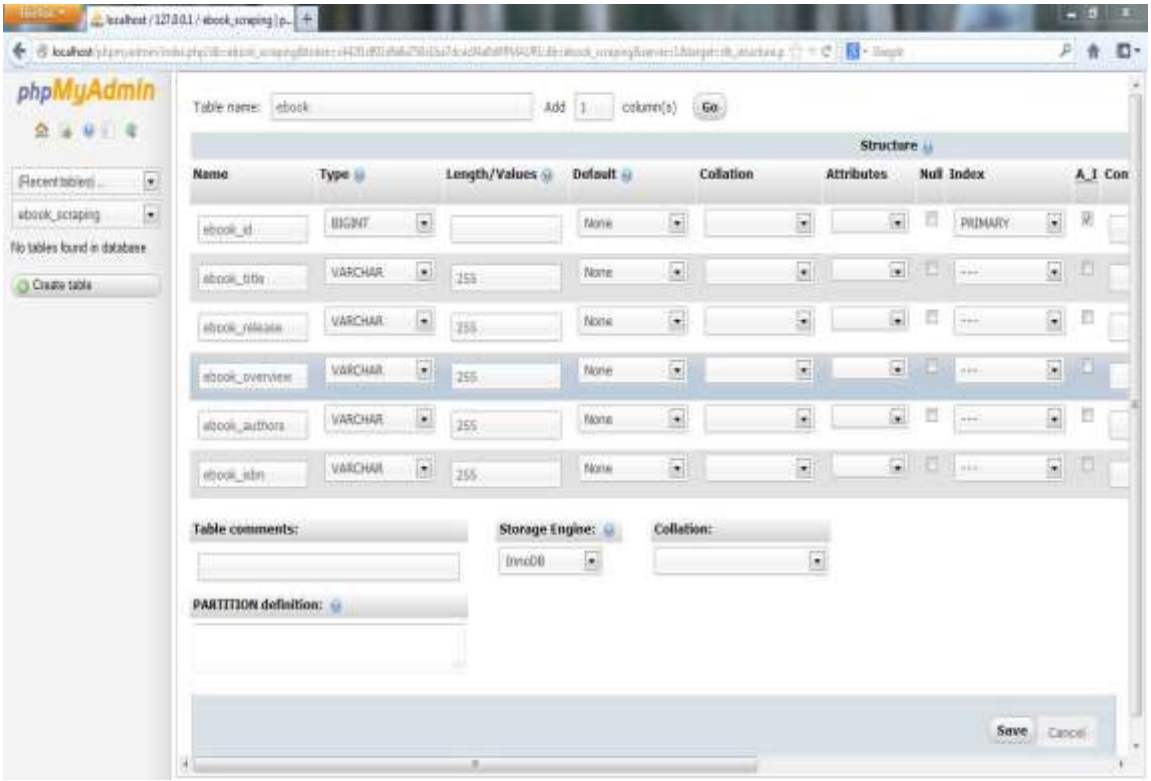


Figure 6

1.2 COMPANY PROFILE

Real-Time Signals Technologies Pvt Ltd Training Institute in Bangalore. Real-Time Signals Technologies Pvt Ltd Training Institute has risen as the pioneer preparing a place for Data Analytics/Data Science/Machine Learning, Embedded Systems, MySQL/Database, Networking/CCNA, System Admin, courses in Bangalore, India. Real-Time Signals Technologies Pvt Ltd Science/Machine Learning, courses in India and furthermore lands 100 percent position certification and situation help for the prepared competitors from Real Time Signals Technologies Pvt Ltd Training Institute. Real-Time Signals Technologies Pvt Ltd Training Institute likewise giving various affirmation courses under the direction of best and experienced resources.

RTS - Real Time Signals Improvement's has connected java and j2ee to give the quality flexibility, openness and straightforwardness of relationship of undertaking basic bundles. We influence utilization of the spring to shape, ejb for our endeavor approach for addressing layersto get to any kind of backend be it mysql, prophet, sap erp, web enterprise ormessage line. These are joined with json, cxf and focus factor for cleanser arranges.Frontends are finished through the most make mvc introduction layers like jsp, jsf,icefaces, myfaces, richfaces, and swaggers 2. We solidly use ajax degrees of progress to offerextraordinary human pc connection and comfort. Our arms on information of sitestreamlining (search engine optimization) strike the correct invitingness among the prerequisites of interoperabilityand convenience.

Our venture growth administrations encompass:

Portable manual arranging mechanical capacity is primary in latest outstandingly forcefulcommercial enterprise recognition. Anyhow, if adaptable advancement tiers ofdevelopment are not made inner an association of key masterminding, you areimprobable to comprehend the overall business capability of the versatile development.Our flexibility directing gathering can portray the compact development guide in youraffiliation/aspect that meets your enterprise destination. Research and developmentadministrations/evidence of concept (%) development our lord and skilled compact tradecollecting let you with doing an unequivocal plausibility exam of adaptable programmingmotion, improvement evaluation and confirmation of idea (%) utilization.

Engineering And Description

Our lord adaptable designers will assist you with building and convenient plan diagramusing becoming flexible traits and versatile stages.Adaptable software improvement our

compact alternate amassing permit you to withcustom convenient programming headway which you require in light of your enterprise.In light of gift instances, should not the great programming answer in your business bethe one that is based on upon your business?

Versatile Purpose Porting

Adaptable specialist let you provide your handy software on various gadgets of identicaldegree or port the bendy packages throughout over different stages like iphone utility,ipad utility, android utility, blackberry application or windows smartphone packages.

Marking And Confinement

Our convenient experts will let you in stamping/white naming your adaptable applications (iphone programs, ipad packages, android applications, blackberry packagesor windows phone packages) for extraordinary clients/directors with control for exceptional topographies.

By way of the usage of our India primarily based seaward flexible software development and seaward transportable checking out limits we assist our customers to develop unexpected things, even as inside and out lessening the value of headway/trying out andtime to exhibit.

Being a seaward software advancement organization we're recollected all round all through the world are a foundation of technical education of various technical programming languages center situated in BTM Layout of Bangalore. It is one of the guidance institutes available in Bangalore for better knowledge and center for perusing a well education about the software and hardware aspects of the current IT fields. With the help of well trained and under industry level profiled trainers guidance, students can acquire the knowledge of any programming language in “Real Time Signal”. The institute provides an opportunity to improve their technical skills as well as communication skills with persistent training and guidance. The institute provides efficient services and infrastructure for students to enhance their skills. The trainers here train the students to become a software expertise, even a non-technical people can opt in this institute in learning of software standards. They offers courses from primary languages (C,C++,Java) to advanced languages like Data Science, Node JS,PLSQL etc. It provides to a chance to the students to accomplish their goals with current industry based path and also offer assistance for students to get placed in well reputed software companies. With the assistance of IT professionals, students can build their technical as well as communication skills to maximize their skills to IT standards. Through practical assistance and guidance students can acquire information very rapidly and people without computer background can accomplish easily with the guidance of Real Time Signal. The institute assures most of the students to get placed in software companies even with an average knowledge of IT standards. The students can make use of the infrastructure of the institution and can get access to labs according to their continent time for practicing and any doubts will sorted out with the assistance quality professionals.

2. LITERATURE SURVEY

2.1 EXISTING SYSTEM:

The existing system does not have web scrapping tool thus it cannot extract data and always a step behind other websites holding numerous data. A particular amount of data is never enough. Data keeps improving and increasing one must behold of that data time to time.

Drawbacks of Existing system:

- Data has to be updated time to time
- Time consuming
- Accurate data not available
- Data extraction is complicated
- Buying data is expensive.

2.2 PROPOSED SYSTEM:

Developing a website with a web scrapping tool is the main aim of the project. The Web scrapping tool in the website helps the website to extract data from other resources and data warehouses through internet. The data extracted can be used for various purposes especially for the growth of website. Isn't it great to get all required data at one place?

Advantages of proposed system:

- **Inexpensive:** Web scrapping administrations offer a fundamental support easily. Rather than buying data this method is much more inexpensive.
- **Easy Implementation: W**hen a proper mechanism is deployed by the web scraping services to remove information, you are guaranteed that you are getting information from a solitary page as well as from the whole area. This implies one needs to contribute just a single time and afterward for endless, a great deal of information can be gathered.
- **Low maintenance:** The database needn't bother with a physical update each day. The web rejecting device assumes responsibility and updates its information on schedule. The upkeep. One perspective that is regularly belittled when new administrations are

introduced, is the support cost. Undertaking financial plan can be spiraled crazy because of the Long-term upkeep costs.

- **Speed:** A web scrapper can cover the work of a man for a week in few seconds/minutes.
- **Accuracy:** Direct goofs in data extraction can cause noteworthy blunders later on. In this way, it is exceptionally important to have an Accurate extraction of an information. The exactness is critical in the sites that manage data about valuing information, deals costs, land numbers or any sort of money related information and so on.,

2.3 FEASIBILITY STUDY:

The common sense study is formed to see whether the endeavors on summit will serve the purpose of the relationship for the measure of work, effort, and along these lines the time spent thusly. The assessment will let the architect foresee the endeavor and thusly the comfort. A practicality investigation of a framework proposition is predictable with its usefulness, which is that the effect on the association, the ability to fulfil their client needs, and hence the viable utilization of assets. Along these lines if a substitution application is proposed it experiences a practicality study to comprehend in the event that it works. During the investigation, the project goes through Technical, Economic, and operational feasibilities.

Technical Feasibility:

Firstly, the system is judged from the specialized perspective. The evaluation is totally upheld the blueprint structure of the framework prerequisites inside the terms of info, yield, projects, and strategies.

Specialized issues rose during the examination:

Does the regular development satisfactory for the proposed one?

Can the structure develop its unforeseen development?

The endeavor must be developed that the important limits and execution are cultivated inside the prerequisites. It's made using the latest advancement. The system has been made using JavaScript the endeavor is really possible for development.

Economic Feasibility:

The creating framework must be fulfilling the desires for cost and advantage. Measures to put forth sure that attempt is focusing on the undertaking will give the best, return at the most timely. One of the factors, which impact the event of a replacement system, is that the cost it might require. The important financial questions rose during preliminary investigation:

1. The value of the equipment and programming
2. The points of interest inside the kind of decreased costs or less over the top slip-ups.

The costs direct a full structure assessment.The system is made as a bit of the endeavour work, there's no manual cost to spend on the proposed structure. All the advantages are presently open, it offers a trace of the structure is monetarily practical for progression.

2.4 TOOLS AND TECHNOLOGIES USED:

Hardware used:

- Processor : Intel Pentive IV and higher versions
- RAM : 512 MB or more
- Cache Memory : 1 MB
- Hard-Disk : 10GB recommended

Software used:

- Clent on Internet : Web Browser, Operating system (any)
- Database : MYSQL
- Frontend : HTML, CSS, Java Script, Bootstrap, JQuery, AJAX
- Back end : PHP , OOPs of PHP

PHP:

Hypertext Preprocessor is a broadly used, powerful programming language that lets in us to increase big internet applications. Other scripting languages that may be used other than personal home page are asp and ruby. However, Hypertext Preprocessor continues to be being carried out the most, and it has no plans of backing down. Php's recognition is attributed to how clean it's far to analyze and use, in contrast to different scripting languages. A personal home page framework is a primary platform that allows us to increase net applications. In other words, it presents shape. By way of the usage of a personal home page framework, you will end up saving hundreds of time, stopping the want to produce repetitive code, and you will be capable of construct applications unexpectedly (rad). With out a Hypertext Preprocessor framework in region, it gets a good deal more tough to provide applications considering you may ought to time and again code lots of personal home page. .You may additionally have to execute the connection among your database and Whatever software you increase from scratch In the meantime, the use of a php framework makes it easierfor you to ensure this connection.

Features of PHP

• Lets in you to build templates to ease web page protection.

• Serve different content to customers primarily based on their browser, ip deal with, date and time, or severa different characteristics.

• Allows connections with databases consisting of mysql.

• Build dialogue forums or internet-based e mail packages.

• Examine and technique xml

HTML:

Hyper text markup language is used to creating the web side both of Static or of dynamic and used to develop the person pleasant web pages.Html is used for developing internet pages.Html is popularly utilized in www. It usesascii characters for each the main text and formatting commands.The principle textual content is data and the entire information is used by the browser to layout the information. A html document is simply a text document, which contains certain statistics you would love to publish..

CSS:

CSS is a manner verbal communication that defines arrangement of html credentials. As an instance, css covers, letters, colorations, boundaries, strains, top, distance across,history photos, sophisticated status and lot of different matters. Html may be used to feature format to web sites. However css offers more alternatives and is more accurate and complex. CSS is supported with the aid of all browsers these days.

- CSS 1 is the level 1 style sheet which is developed by 2 people called Hakon Wium Lie and Bert Bos. It contains elements like font, color, backgrounds, margin, border, padding, text, attributes, images and tablets.

- CSS 2 is the level 2 style sheet which is published in May 1998. The advanced version of CSS 2 is CSS 2.1 which is final revised version of level 2.

- CSS 3 is the level 3 style sheet which is divided into separate modules. It extends features of CSS 2 and CSS 1. It contains rounded corners, 2D/3D transforms, animations, transitions, gradients, and pagination.

- CSS 4 is the level 4 style sheet in which modules are separated and is an advanced version of CSS. It uses all the latest version of CSS and also contains advanced features.

JAVA SCRIPT:

JavaScript is a significant level language which is the significant language where a web engineer ought to learn. The fellow benefactor of the Mozilla Corporation, the Mozilla venture, and the Mozilla Foundation Brendan Eich discovered this JavaScript. It is an entirely adaptable language which gives an assortment of devices. JavaScript serves to effectively learn and utilize. It is upheld by all new internet browsers like Chrome, Mozilla Firefox, Opera, Safari, Internet Explorer, and UC Browser. It runs on the client side which is used to design the project. It tells how pages will behave when there is an event occurred. It is also a lightweight language which is used commonly in the web pages as a part. It is popularly used in the current world to make a web page interesting. It helps to build applications and sites rapidly. JavaScript has first started as a scripting language. It is founded by Netscape which is also called live script. It does not require separate compiler because it contains client implementation. So it is a very powerful language used across the world

BOOTSTRAP:

It is a front-end framework and an open source language which is developed by two popular persons named as Mark Otto and Jacob Thornton. It is released in August 2011 in GitHub and developed for free. Bootstrap has become a popular and no.1 project in June 2014 on GitHub. It helps to create designs easily. It includes HTML and CSS based designs. Bootstrap is faster and easier to develop web applications. We can include JavaScript plugins as optional in bootstrap. It is originally named as Twitter Blueprint. It has JavaScript components in the form of plugins of jQuery which provides elements with additional user interfaces [2]. Bootstrap consists of style sheets which have less series and can make adjustments. It contains much functionality like a grid, typography, tables, images, jumbotron, buttons, pagination, tooltip, glyph icons and much more [2]. Bootstrap has many advantages. List of advantages is explained below.

Advantages of Bootstrap:

- With the help of basic knowledge of HTML and CSS, anybody can easily use bootstrap.
- It has many features like responsive CSS, adjust to phones, tablets, and desktops.
- Bootstrap is compatible with all browsers in which it can perform in any latest versions of Opera, Internet Explorer, Google Chrome, Safari, and Firefox browsers.

- It saves lots of time efforts.
- We can arrange dynamically the layout of web pages with the help of responsive web design.
- Bootstrap contains mobile first approach or mobile first design which can be accessed by default. It is a part of the core framework

JQuery:

JQuery is known to be a JavaScript library. JQuery is made to improve the procedure of HTML DOM tree traversal. It likewise handles the control, additionally called as occasion dealing with, CSS movement, and Ajax. it's free, open-source programming utilizing the lenient MIT License. As of May 2019, jQuery is utilized by 73% of the ten million well known sites. Web investigation demonstrates that it's the preeminent generally conveyed JavaScript library by an outsized edge, having 3 to multiple times more use than the other JavaScript library..

AJAX:

AJAX is an Asynchronous JavaScript and XML. AJAX is a procedure for making quick and dynamic locales. AJAX will make the locales update nonconcurrently. It does as such by trading little measure of information with the server off camera. this proposes it's conceivable to refresh portions of a web page, without reloading the whole page..

Database – MYSQL:

For securing controlling and recouping the store data in a social database we will be used in composed request language Different approvals can be resolved to the related procedures points of view and tables Grants creation and the leading group of database and tables Customers are allowed to get to the data Controls and describe the data in the database Supplement inside various vernaculars will be allowed The Mysql server gives a database management system with querying and connectivity abilities, in addition to the capability to have extremely good statistics shape and integration with many unique systems. It can deal with huge databases reliably and speedy in high-annoying production environments. The Mysql server additionally offers wealthy function together with its connectivity, velocity, and protection that make it appropriate for gaining access to databases.

The mysql server works in a client and server gadget. This machine includes a more than one-threaded square server that helps various backbends first rate client packages and libraries, administrative equipment, and many utility programming interfaces (api)s.

• Mysql is a database management device.

• Mysql databases are relational

3. SOFTWARE AND HARDWARE REQUIREMENTS

3.1 SOFTWARE REQUIREMENT:

This software is designed in a way that every user can access this from any remote place. Hence, it is required that the software be uploaded to a web host apache server like godaddy & bigrocks. The software quire for this software are

1. Apache 2.0 web server with ssl secure certificate
2. Php 5.4.x version
3. Mysql database version 5.3.x

3.2 HARDWARE REQUIREMENT:

For accessing the software, a user will require a computer system with internet connectivity and an updated browser version. These are the following hardware requirement

1. Mozilla Firefox 17.0+, Chrome browser, opera browser
2. Internet connection having minimum 512kbps bandwidth
3. System requirement depends on browser basis

3.3 FUNCTIONAL REQUIREMENTS:

- The structure runs of apache server so it is essential that the server must have apache server interpretation 2.0 immediately available
- We have used PHP for server-side scripting so the current variation of PHP must be available on the server.
- For putting away the information on the site, the MYSQL database is utilized.
- HTML is utilized for making the design of the site application.
- For making the structuring of the site pages, CSS has been utilized.
- JavaScript is a scripting language that has been realized on the system for playing out the whole of the client side server endorsement.

3.4 NON-FUNCTIONAL REQUIREMENTS:

- It should be effectiveness. The resulting performance should be compared in relation to the effort of configuration.
- The administration of the crawler should be done over a graphical user interface.
- System requirements should be as low as possible.
- The crawler should be extensible. It should be possible to add features, plug-ins and other forms of customizations.
- The crawler should be able crawl more than one site at the same time.
- The availability has to be up to 99.6%. Every interaction has to be logged in detail to be able to relate to errors.
- It should be able to back up the configuration files of the crawler.
- **Data discovery**: Technical requirement gathering process is defining what data needs extracting and where can it be found.
- **Data Extraction**: During this process, our intention is to clearly capture what data we want to extract from the target web pages.
- It should provide proper and reliable information about fixtures and results of filtered data.
- The links to the websites should use relevant parameters to get automatically to the correct page.
- If an error is thrown then Notification should be clear and easy to understand

4. SYSTEM DESIGN

4.1 GENERAL:

Setup Engineering deals with the contrasted UML [Unified Modeling language] plots for the use of the endeavor. The arrangement may be a significant planning depiction of a thing that will be amassed. Programming setup may be a method through which the requirements are changed over into a depiction of the item. Arrangement is the place quality is rendered in programming structuring. Design is the best approach to unequivocally make an understanding of customer necessities into a finished thing. This paper is implemented in such a way that the user can register himself onto the cloud and share data amongst several users. The creator of the group is assumed to be the manager of the group. The manager can add members to the group as well as remove users who misbehave. Any member of the group can upload files which can be accessed and downloaded by every member of that group.

4.2 DATA FLOW DIAGRAM

Data Flow Diagram is defined as the graphical description of the data flow which flows among the modeling, process aspects, and information system. It describes the overview of a system. It tells what kind of information is passed on to the input and how the output is represented and how the data is advanced and stored in the system. It does not tell the timing of a system or process time instead it shows information. Data flow diagram contains symbols to represent, they are a rectangle which represents Input/output of this system, a circle represents the process or function, arrow mark represents the flow and two parallel lines represent the database or file. As shown below the diagram tells how the flow of the project is taking place. When the user got success in his registration process, he will get a link to his mail address and he activates for the first time. If the user wants to log in then he can access through his email address and password provided in his registration process. After logging in he can control the site. When the user fails then he will return to the registration page with errors. There are three common levels of DFD namely Level-0, level-1, and level-2.

4.2.1 WEB SCRAWLER

A Web crawler, is in like way called a bug or 8-legged creature bot and regularly abbreviated to crawler, is an Internet bot that deliberately takes a gander at the World Wide Web, for the most part with an authoritative objective of Web referencing (web spidering).

Web records and some various destinations use Web crawling or spidering programming to revive their web substance or chronicles of other zones' web content. Web crawlers copy pages for managing by a web search instrument that records the downloaded pages so customers can glance through more gainfully. Crawlers eat up resources on visited structures and sometimes visit districts without ensuring. Issues of the schedule, weight, and "remarkable inclinations" become perhaps the most tremendous factor when gigantic groupings of pages are gotten to. Structures exist for open goals not wishing to be crawled to make this known to the crawling executive.

For example, including a robots.txt record can request bots to list just bits of a site or nothing in any way at all. The proportion of Internet pages is enormous; even the best crawlers come up short concerning making an immovable record. Thusly, web records fight to give fundamental outline things in the early tremendous stretches of the World Wide Web, before 2000. Today, fitting results are given in a brief second.

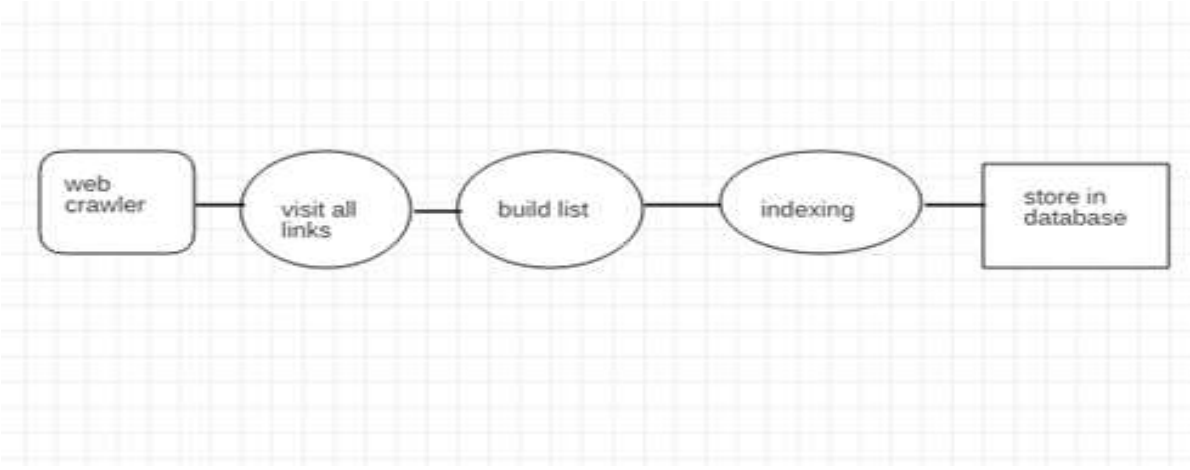


FIGURE 6: Web Scrawler

4.2.2 WEB SCRAPPER

Web scratching, web gathering, or web data extraction is data scratching used for withdrawing data from objectives. The web scratching programming may get to the World Wide Web genuinely using the Hypertext Transfer Protocol or through a web program. While web scratching should be conceivable really by a thing customer, the term by and large surmises motorized strategies wrapped up a bot or web crawler. It is a kind of imitating, where unequivocal data is amassed and copied from the web, when in doubt into a central close to database or spreadsheet, for later recuperation or appraisal.

Web scratching a page wires bringing it and disengaging it from it. Bringing is the downloading of a page (which a program does when you see the page). In this way, web crawling is the central area of web scratching, to bring pages for later arrangement. Once brought, by then extraction can occur. The substance of a page may be parsed, looked, reformatted, its data recreated into a spreadsheet, and so forth. Web scrubbers normally expel something from a page, to use it for another explanation somewhere else. A model is found and copy names and phone numbers, or affiliations, and their URLs, to a blueprint (contact scratching).

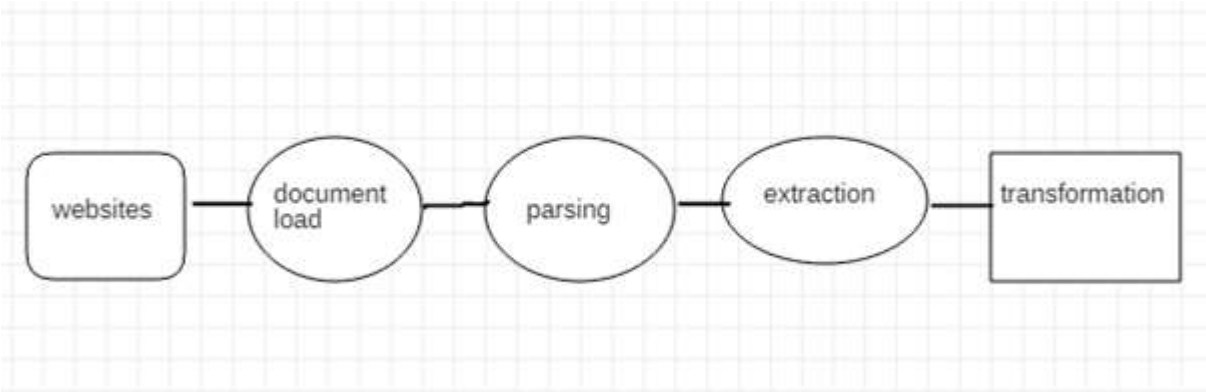


FIGURE 7: Web Scraper

4.3 USE CASE DIAGRAM

Use Case diagrams are the representation of a behavior diagram which describes a set of activities performed by the system. It collaborates with more apparent users. Use case must give valuable and observable outputs to the actors of the system. It is the requirements which are required for the system usage under analysis or design. It explains how the system or environment works so that it will perform services. Main elements of use cases are included and extend relationships, actor, subject, and use case. An actor is a behavioral distributor which is specified by an external entity that acts together with the subject. For example, users, customers, client, student etc. The subject is a distributor which includes component, subsystem, and class. It is not separately specified as it is attached to actor and use case. Below diagram shows the use case representation

The above use case diagram shows how the project works with one or more simultaneous functions. When a user is doing work on some task at the same times other users can also do the same task. There will not be any data loss or merging. Admin can do activities such as share, download, remove, add favorites, and upload than at the same time users also can access the same activities at the same time..

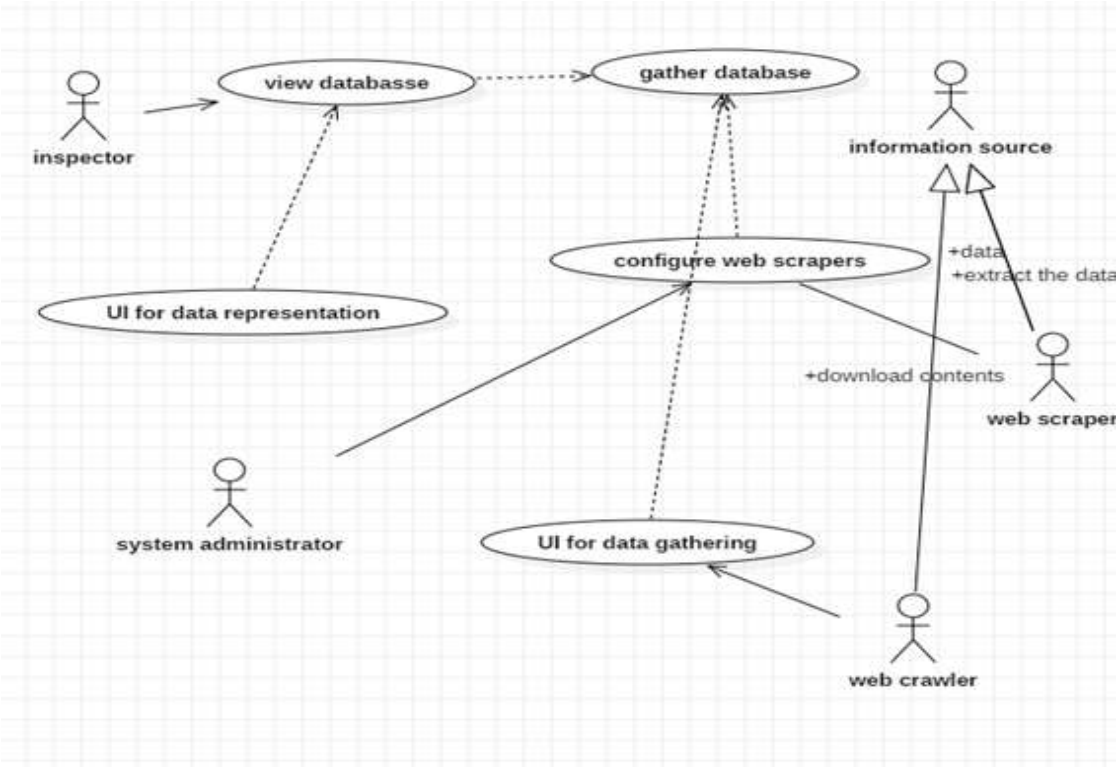


FIGURE 8: Use Case Diagram

4.4 ACTIVITY DIAGRAM:

UML is incredibly helpful for envisioning and reporting programming frameworks, yet the phrasing is regularly to some degree overpowering for someone unacquainted UML. An action graph is really a flowchart that shows exercises performed by a framework. The action outline is another significant graph in UML to clarify the dynamic parts of the framework.

The accompanying outline shows the activity of web scraping which represents that it is a easy service of data acquisition from a website. And DOM tree is used to identify and extract the data. Activity is a particular operation of the framework.

The activity diagram is a step by step graphical representation of actions and activities. The most important symbols are black circle denotes the start of the activity, rounded rectangles which show action, diamond denotes decisions, arrows flow from the starting point to ending point of the activity, bars denotes split and join activities, and an encircled black circle denotes stop or end of an activity. Below diagram shows the activity diagram of the project.

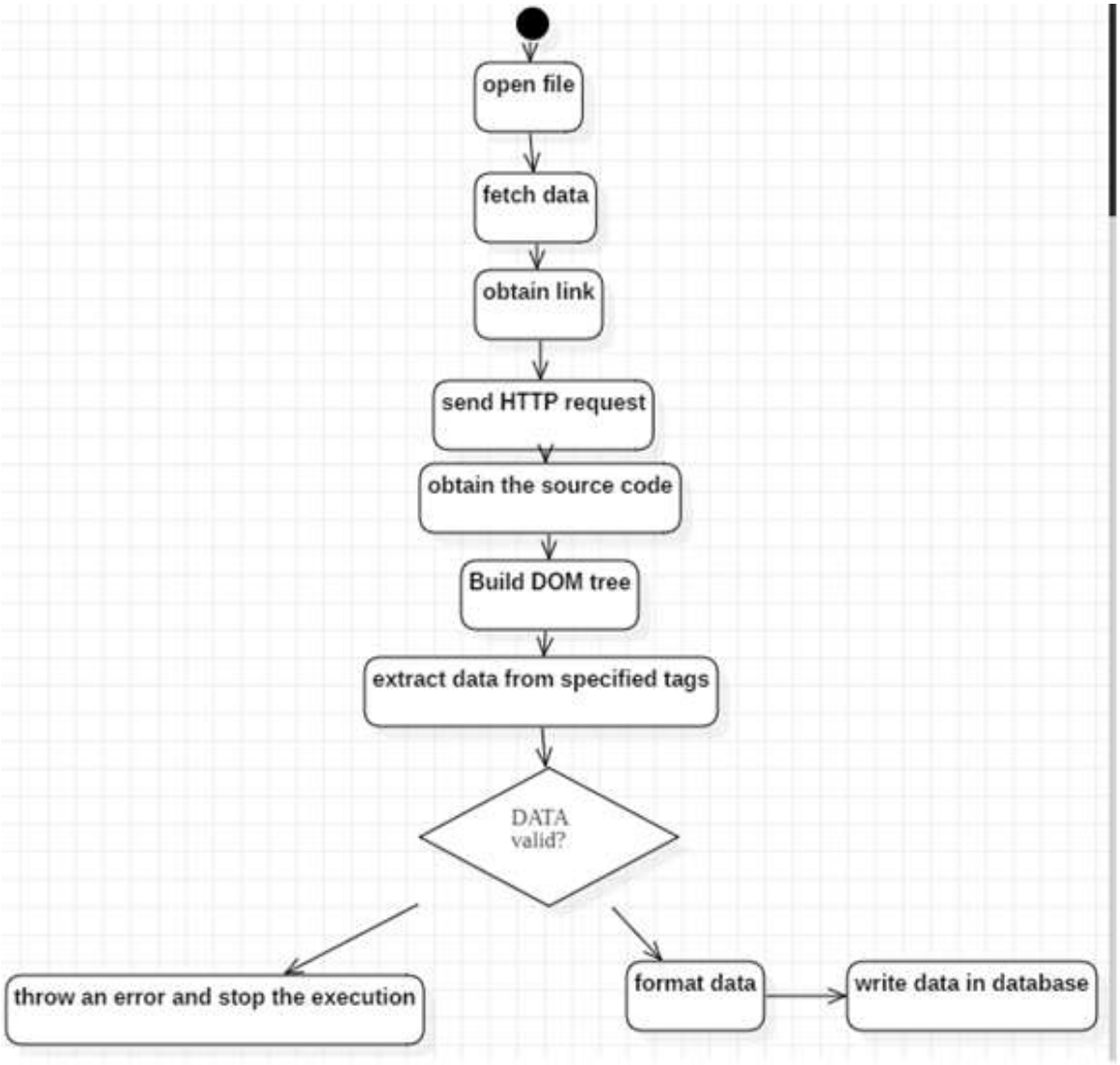


FIGURE 9 : Activity Diagram

4.5 SEQUENCE DIAGRAM:

A sequence diagram is the time sequence of an interaction of object. It defines classes and objects which are involved at an outline of epitome or synopsis. It shows how the messages are interchanged between the objects. A sequence diagram is also called event scenarios or event diagrams. Below figure shows the sequence diagram.

Game plan graphs are usually related to utilize case recognize inside the steady point of view on the structure being taken a shot at. Progression diagrams are also all around familiar with the terms as event outlines or event circumstances.

The diagram shows the sequence of the project. When users register then they are taken to log in page with success message. After logging in to the site users can perform many actions. If the file is uploaded then it returns a success message with the file uploaded. When the users delete then the file or folder has the backup facility. At last, when users' log out, it returns to log in page.

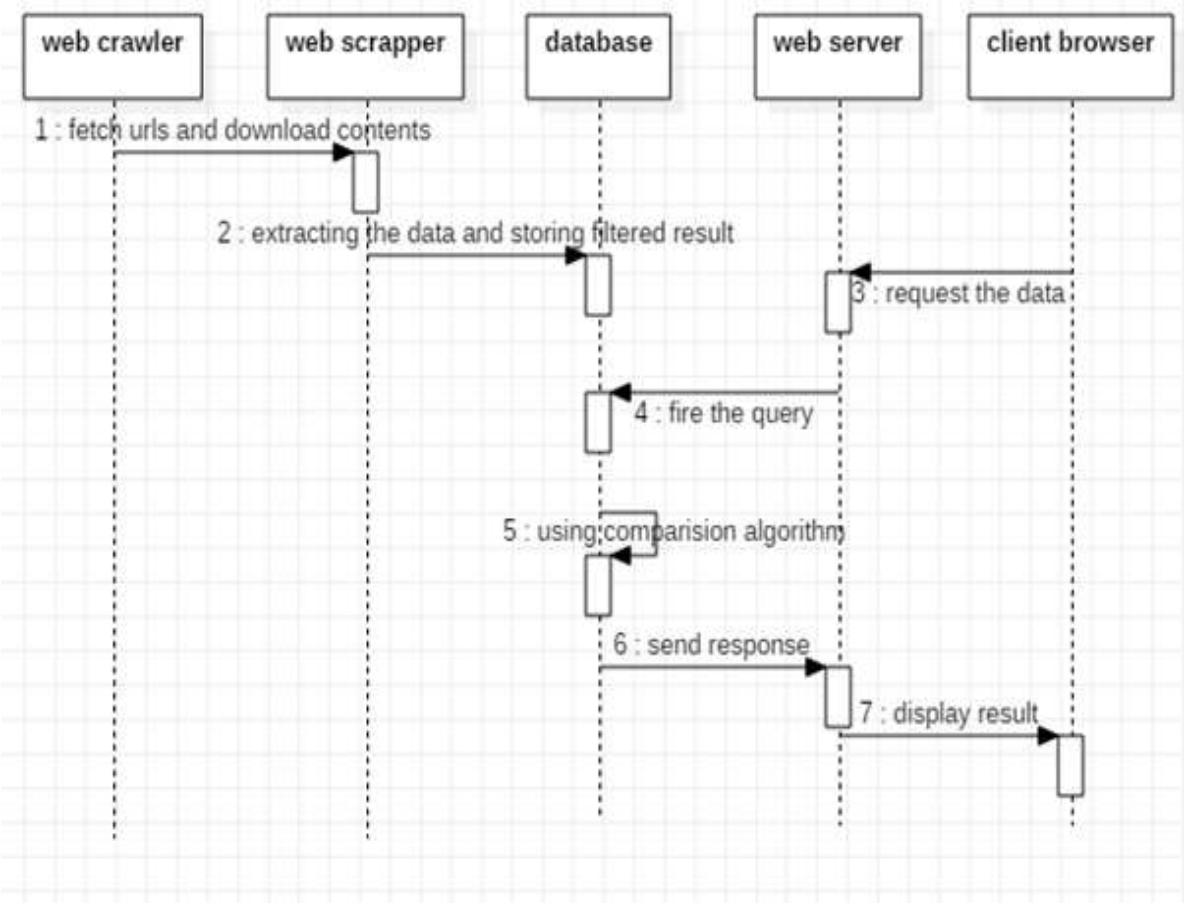


FIGURE 10 : Sequence Diagram

4.6 ENTITY RELATIONSHIP DIAGRAM:

The following diagram has entities like user, website, database, web scraping software as shown representing relationship among them.

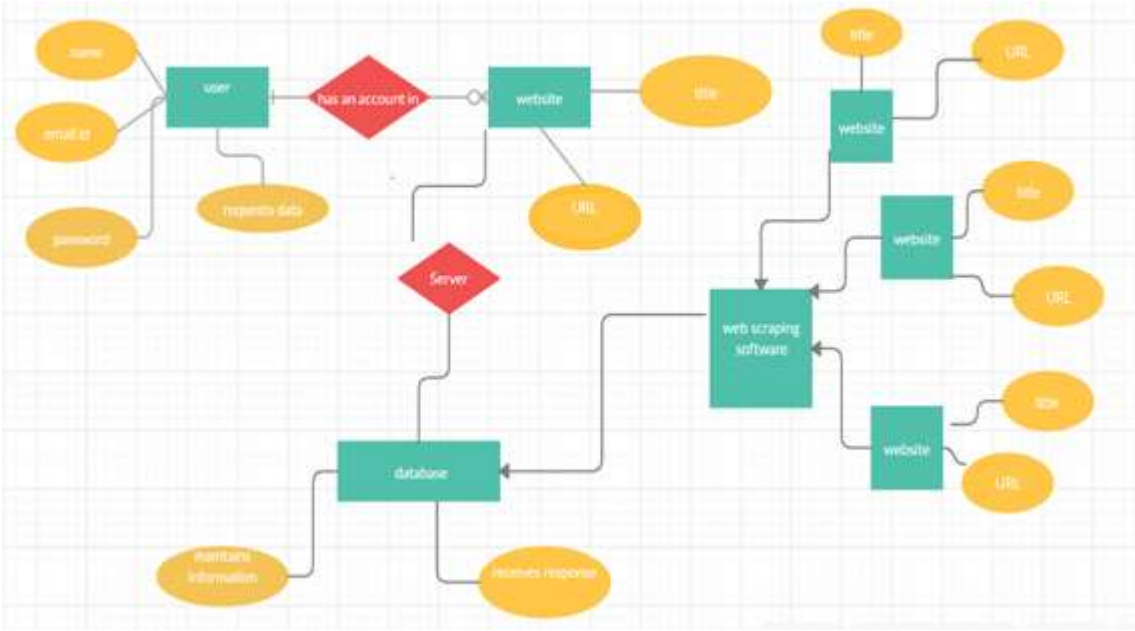


FIGURE 11 E-R Diagram

An entity relationship diagram describes the relationship among the entities which are stored in the database. The main symbols used in ER-diagram are the rectangle which is used for an entity, rhombus which is used for action and circle which is used for attributes.

The below diagram explains how the flow of the project is taking place as the user registers. When the user registers with his basic information like name, e-mail, password, contact information, and country, he will get a confirmation link in the mail, this helps to log in to the site.

5. SCREENSHOTS

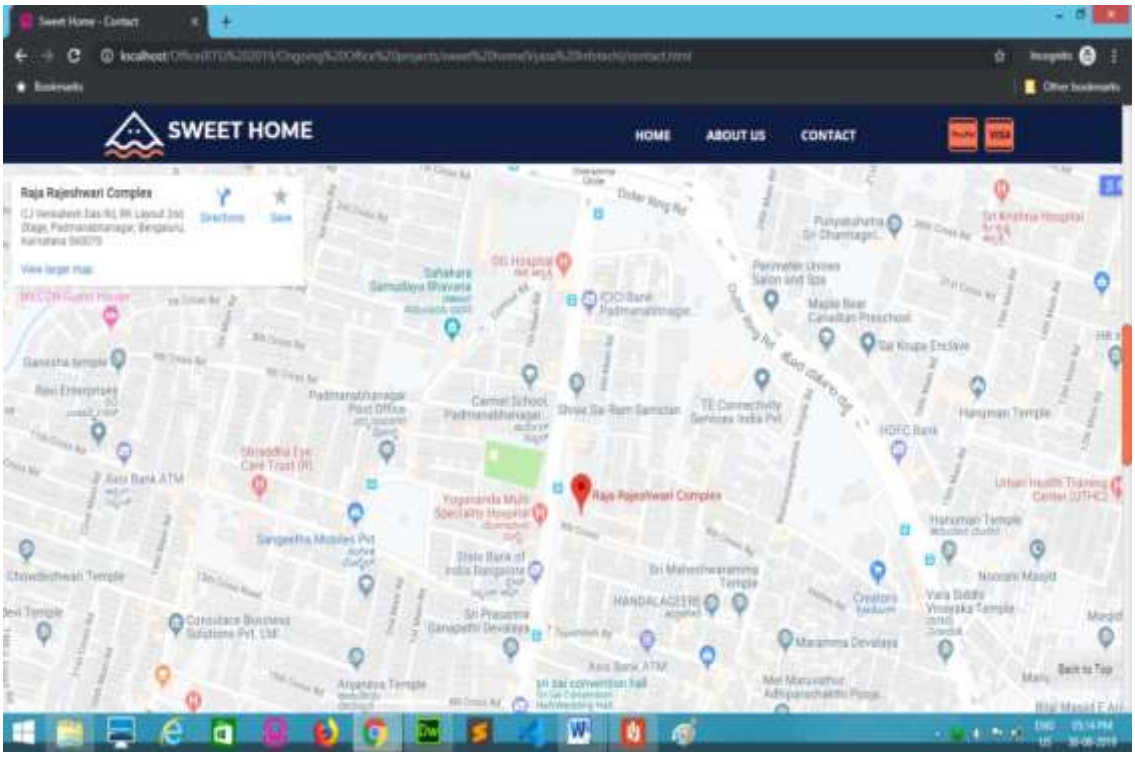


Figure 1 : Displaying Address in google maps

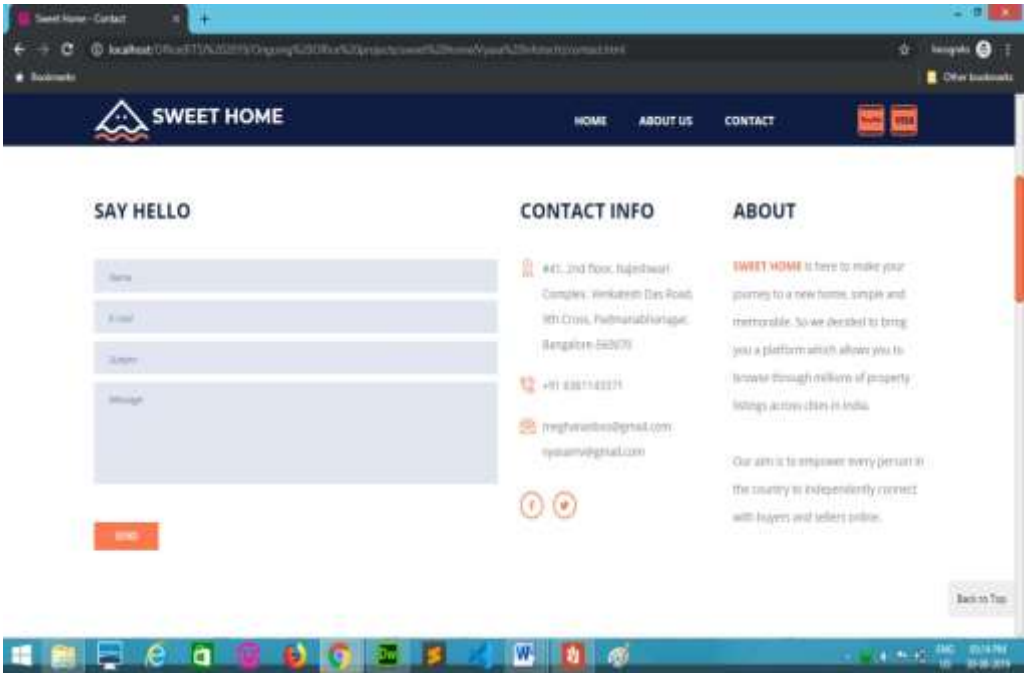


Figure 2 : Page displaying contact information, about the website, feedback form

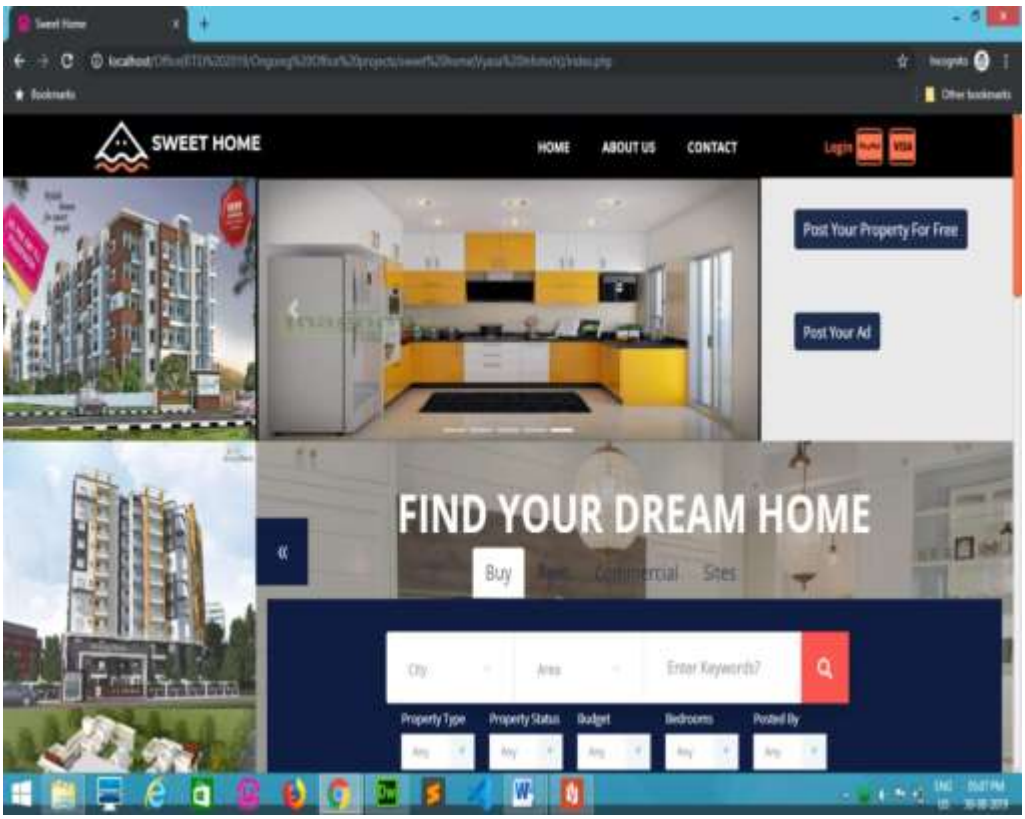


Figure 3 : Home page

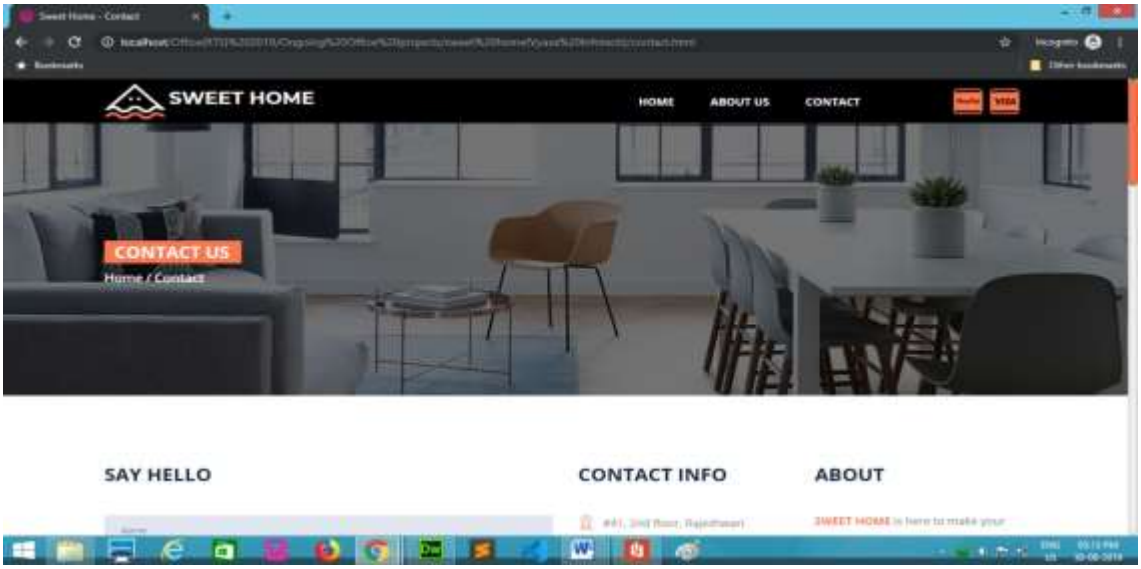


Figure 4 : Contact page

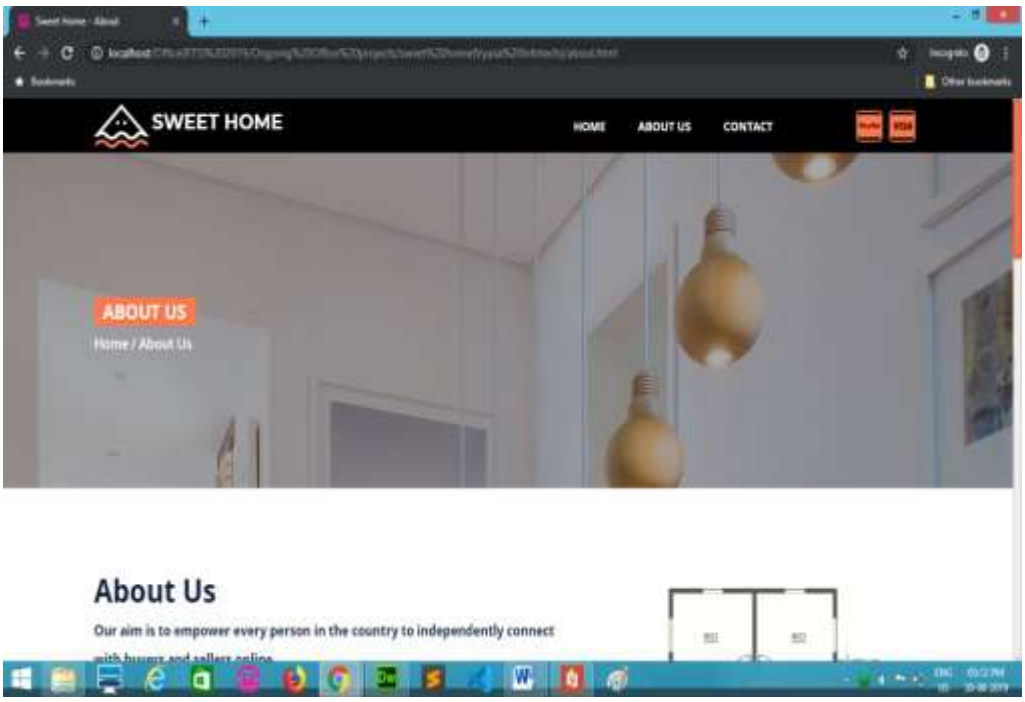


Figure 5 : About section

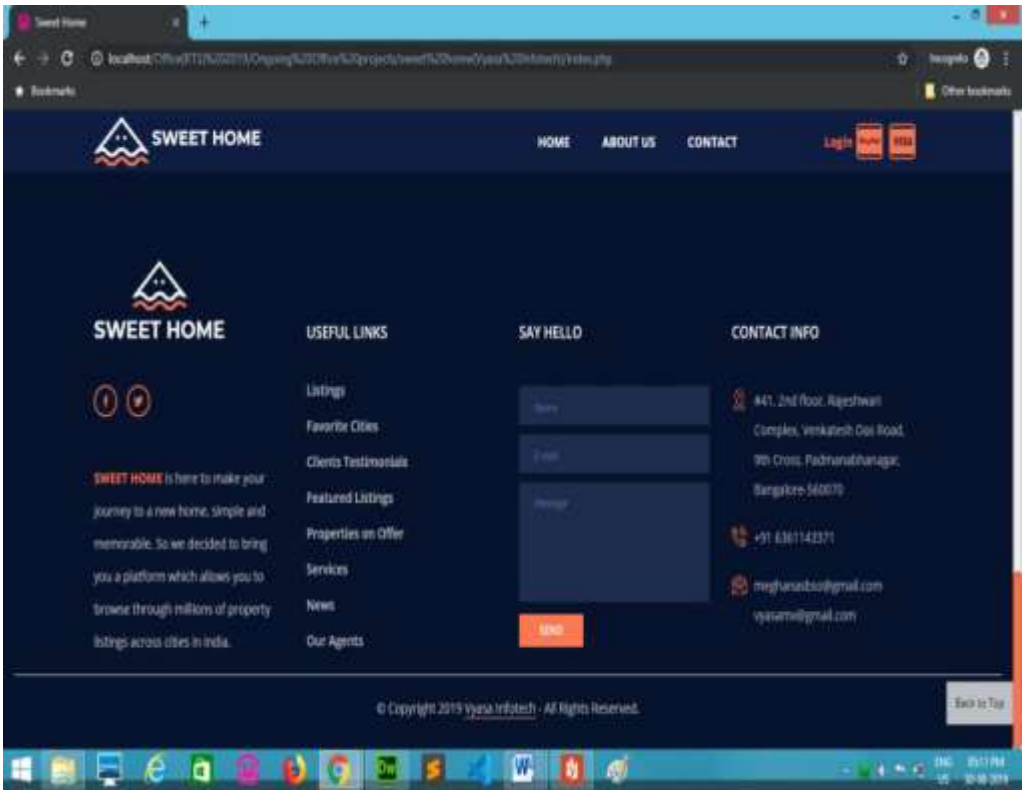


Figure 6 : Footer of the page

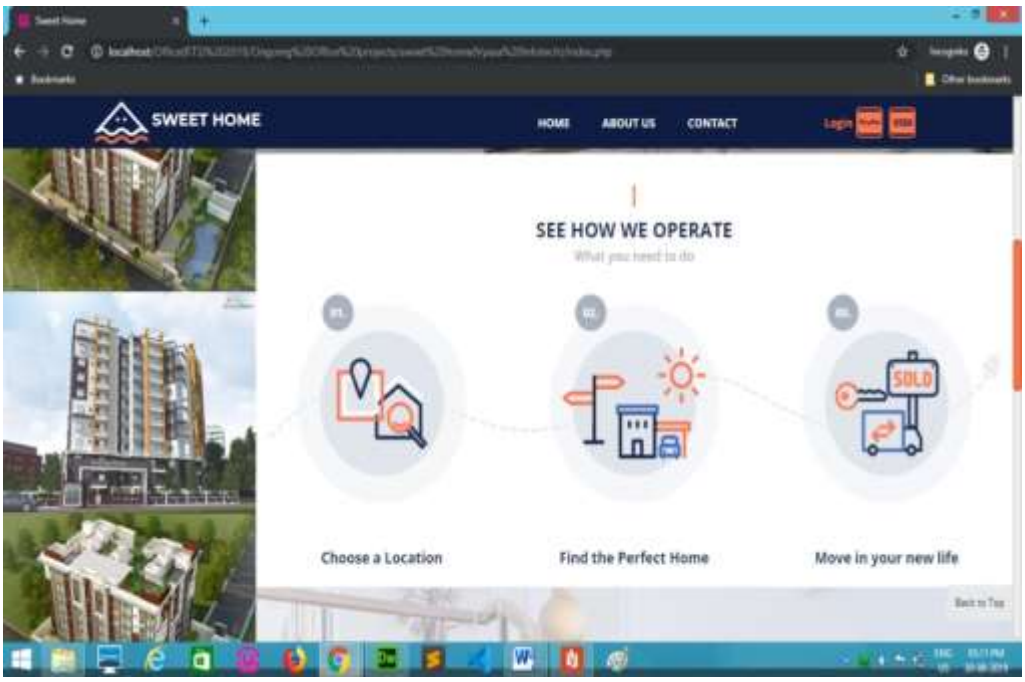


Figure 7 : Page showing the Operational functions of the company

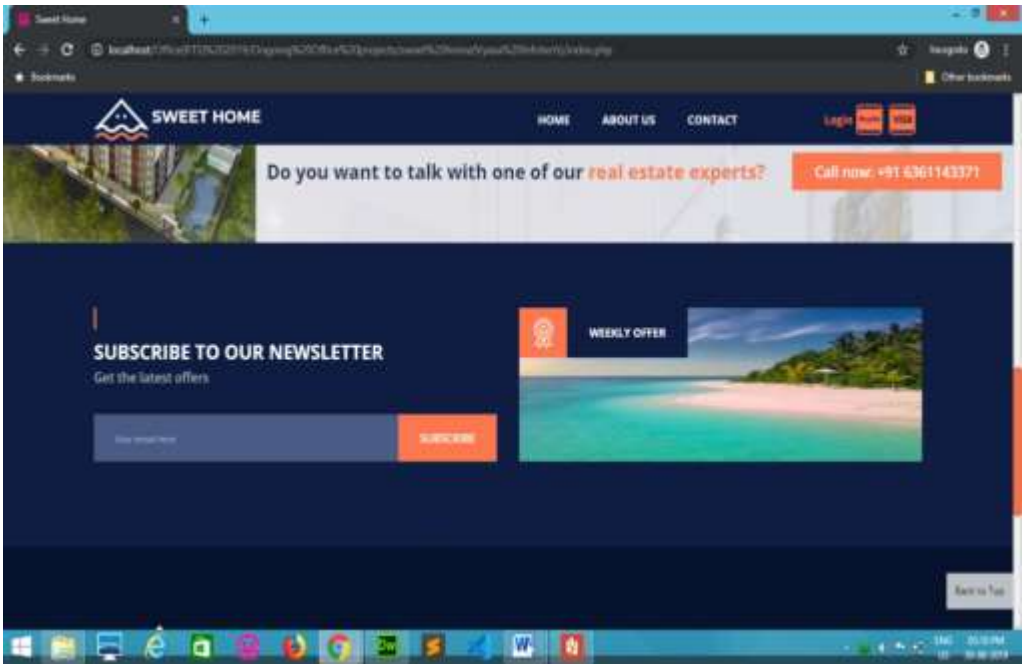


Figure 8 : Page showing form for newsletter and telephonic interaction with experts.

6. SOFTWARE TESTING

Software testing is a progress of executing an application with its errors. It is also defined as verifying or validating the software program or product which meets the technical and employment requirements. The components of software testing include requirements, giving results to all kind of inputs, performing functions with accurate time, used sufficiently, can be run in any environment, and achieving the results. It also provides the consequences of risks in software implementation. It keeps finding the software bugs (defects or errors) and helps to verify whether the software product is suited for use or not. There are several testing methods which use some generalship to choose a test that is practicable or feasible for the specific available resources and time. The regular degrees of testing contain programming testing, unit testing, part interface testing and framework testing and strategies incorporate discovery testing, white box testing, and static or dynamic testing. A portion of the testing techniques and test levels are clarified beneath.

7.1 LEVELS

System testing is a step-by-step process of software testing where entire integrated software is examined and tested. It verifies whether it meets the requirements. Following explains the system testing levels which include integration testing, unit testing, system testing, and acceptance testing.

Unit Testing

Unit testing defines the particular section of code at the functional level which verifies the functionality. It is written by developers because they will work on typing the codes to ensure that whether the function is working as he needs. There may be several tests for one function to hold the branches or corners cases of code. It cannot verify the functionality of the software alone but it checks whether the software is working independently or not from each other. It is the action of software development process which has an application which is synchronized and strategies of detection to reduce time, risk, and costs of software development. It is executed by an engineer or software developer in the development phase of the software growth lifecycle. It consists of metrics analysis, data flow analysis, peer code reviews, static code analysis, code coverage analysis, and other verification of software practices**.**

Unit testing is a development procedure where programmers create tests as they develop software. The unit testing regarding the EMR project has been done to every particular unit to check the functionality of each unit of the modules (doctor, patient and hospital modules) developed.

All the testing in the EMR project is done manually without any automation testing with several other software like C unit. All the tests are performed in parallel to the unit test cases developed which proves the correctness of the project developed.

Validation Testing

Validation testing encapsulates the integration testing and the system testing. This is done to check whether the system or website developed accomplishes the specifications which are a major concern of integration testing. The validation testing is done to check the adaptability of the system developed to its working environment. The IDE used to develop the EMR project is the pycharm which is compatible of incorporating all the components used to develop the EMR project which are python collaborated with machine learning algorithms and django as a web framework. The following tests are included in the Validation testing.

- **Test for Regression :** The python 3X and the pycharm IDE is tested whether the features and the compatibility of the system developed is working on the all the versions of the software used. As the python 3X differs in syntax with its previous versions, the system developed is not compatible for the previous versions but can be used in upgraded versions of python.

- **Test for Recovery:** The system developed is to be deployed on cloud to counter the disaster management in case of any interruption in a number of ways to the software used. Thus, there will not be a lost of data, hence resulting in the disaster management.

- **Security Testing:** The test for checking the security is done perfectly and is succeeded. The unauthorized attempts to access the system is denied. A user(doctor, patient or hospital) should first register with the proper credentials and those credentials are given to login access by the user. The password for doctor and hospitals are highly strong where the password should include lowercase, uppercase letters, a number and a special character as a must.

The system is developed robust as the patient details are highly confidential which cannot be changed by any of the user. The updating of clinical details may be done by the person(using patient credentials) who has a personal hatred and this can result in a loss of life. Thus the users can only view the patient details and treatment details but cannot change their details, personal or clinical details. The Indian patient data fetch is made easy and unique through their Aadhaar ID.

- **Stress Testing:** The stress testing is made to analyze whether the system developed can handle the stress when there is a typical hike in demands of using it. The machine learning algorithm may take time but still the system responds faster. As the database used is

SQLite which doesn't have a separate server and can be utilized by many users at a time and there will be no hung or interruption in the system while used by the users.

- **Performance testing:**

The performance testing is done to check whether the system developed meets the performance requirements. This also includes details like memory management etc which results in the efficient performance. It also calculates the time taken to respond when an input is given.

According to the project developed, the performance testing is successful as the EMR website meets all the performance requirements. The system, when used by many users is not hung or doesn't result in interruption or failure.

- **Usability Testing:** The usability testing is to test the usage of the system by the users for whom the system is developed. The usability testing ensures whether the system developed is useful and worthy to the end user using it. The EMR project is efficiently used by its users and thus is successful.

- **Alpha & beta testing:** The alpha and beta testing are done before providing the software to the end user. Alpha testing is done to check whether the EMR is feasible to release in the market. As the deployment is not done yet, the alpha testing is not done. The system developed is not a ready market product. The alpha release is an initial release where the bugs if any found can be reported to the developing tea

6.4 TEST CASES

SL.no	Description	Expected output	Actual output	Status
Ds_01	To verify that Users should put an details in enquiry section and o empty values should be accepted	Verified that the all the fieds in the enquiry section has been filled	Enquiry section is filled properly	Pass
Ds_02	To verify that in enquiry section name field should accept only alphabets	Name field in enquiry section accepts only alphabets	Veified that name field only accepts only alphabets	Pass
Ds_03	To verify that in enquiry section Email field should accept only pattern of email with @and correct pattern of email	Email id field accepting only desired pattern of Email id	Verified that Email id field accepts only desired pattern	Pass
Ds_04	To verify that in enquiry section contact numberfield should accept only 10digit numbers	Contact field only accepts only 10 digit numbers and should not accept any other characters or special characters	Verified that contact fields accepts only 10 digit numbers	Pass

Ds_05	The search keyword should include the city in which you are looking property for	City field should be selected to buy or rent the property	Verified that if city is not selected it should not process further	Pass
Ds_06	The filter section should be appropriate to searches and desires of customer	Filter should be selected and applied and fetch the results	All filters are selected and according to the filters search results should appear	Pass
Ds_07	The minimum price should be lesser than the maximum price in filter	The price should match the range and minimum price should not be greater than maximum price	If minimum price is greater than maximum price it should show the error	Pass

7. CONCLUSION

So as to remain track of the information with respect to people, items, or organizations, news scratching is kind of helpful. Web scratching is fundamental to the strategy since it permits brisk and productive extraction of information inside the kind of news from various sources. since information is getting included at a lightning pace, web scratching is that the regular reaction! As the present universe of business procedure is driven altogether by information, web scratching is only getting the chance to develop complex. Considering the conditions of the present competition and individual development, a web scrapper must be in use to put up with the world.

8.FUTURE ENHANCEMENTS

The Web is enormous, complex, and ever-advancing. About 90% of all the information inside the world has been produced in the course of the most recent two years. during this tremendous expanse of information, how would I get to the pertinent snippet of data? this is regularly where web scratching dominates. With the ascent in information, irregularities in web scratching are rising.

No more is a web scratching a site of the coders; Actually, organizations currently offer modified scratching instruments to customers which they will use to encourage the data they need. the aftereffect of everybody prepared to slither, scratch, and concentrate, has neither rhyme nor reason misuse of valuable labor. Community scratching could well recuperate this hurt. Here, web scrubbers work in an unexpected way. In the event that one web crawler is found to do a wide scratching, the others are known to scratch information off an API. An expansion of the issue is that text recovery pulls in more consideration than mixed media; and with sites getting progressively unpredictable, this implements restricted scratching limit.

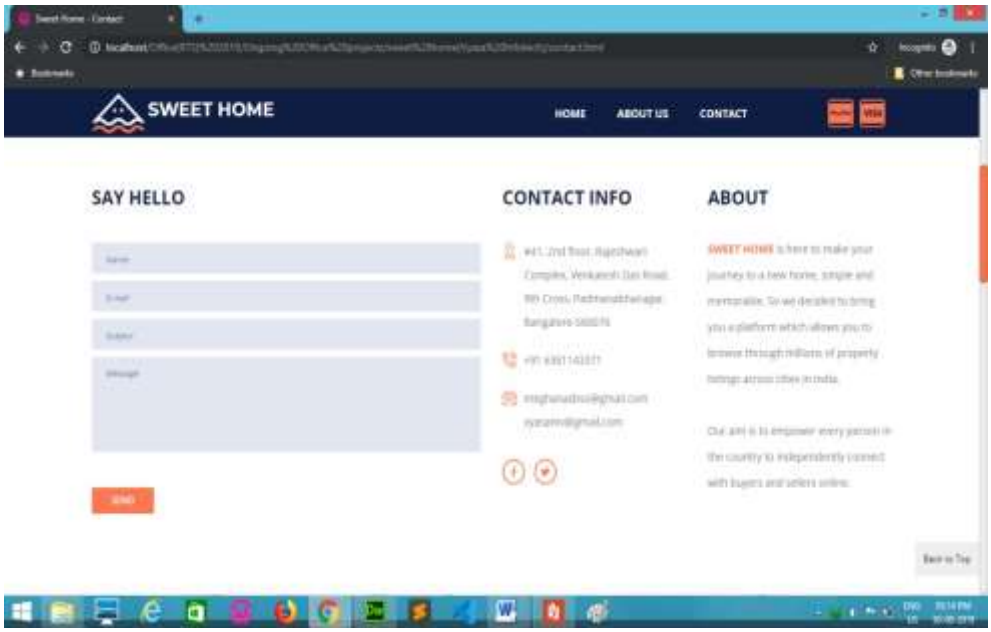
Effectively, the most significant test for web scratching innovation is Privacy concerns. With information unreservedly accessible (the majority of it deliberate, a lot of it automatic), the choice for stricter enactment rings most intense. Unintended clients can undoubtedly focus on an enterprise and money in of the business utilizing web scratching. The contempt with which "don't scratch" arrangements are dealt with and terms of use disregarded, reveals to us that even legitimate limitations aren't sufficient. With Internet and web innovation spreading, gigantic measures of information will be available on the web. Especially with an expanded selection of the versatile web. reliable with one report, by 2020, the measure of versatile web clients will hit 3.8 billion, or around a large portion of the total populace! Since 'large information' is frequently both, organized and unstructured; web scratching instruments will just get more honed and sharper.

9.BIBLIOGRAPHY

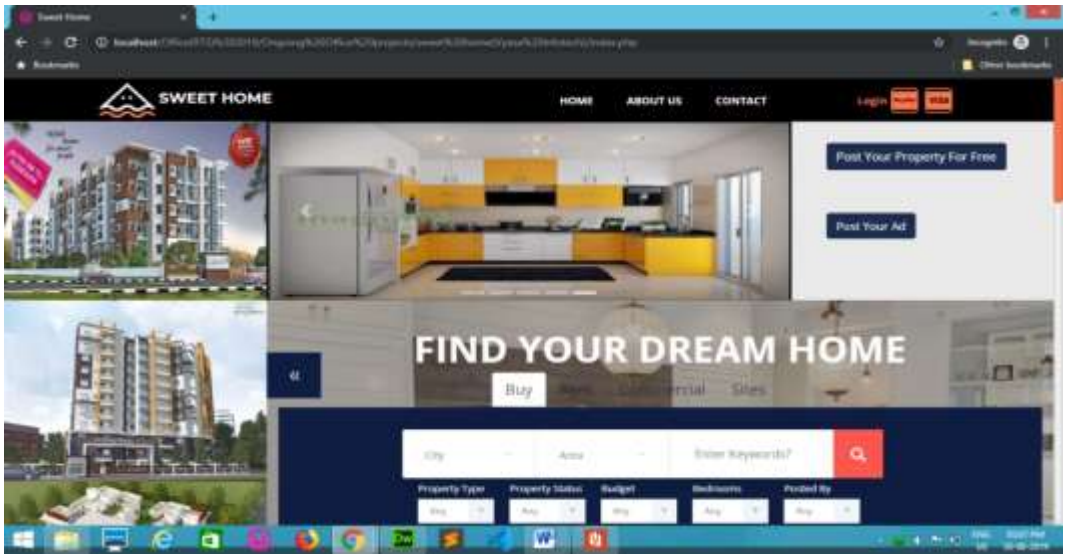
1. Mayank Dhiman Breaking Fraud & Bot Detection Solutions OWASP AppSec Cali' 2018Retrieved February 10, 2018.
2. Neuburger, Jeffrey D (5 December 2014). "QVC Sues Shopping App for Web Scraping That Allegedly Triggered Site Outage". The National Law Review. Proskauer Rose LLP. Retrieved 5 November 2015.
3. scrapingexpert.com/advantages-disadvantages-web-scraping/
4. Web Scraping And Data Acquisition Using Google Scholar by D.Prathiba **Published in** 2018, 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS).
5. Web scraping and mapping urban data to support urban design decisions, ITU volume 15 no.1 by Elif ENSARİ, Bilge KOBAŞ.

9.User Manual

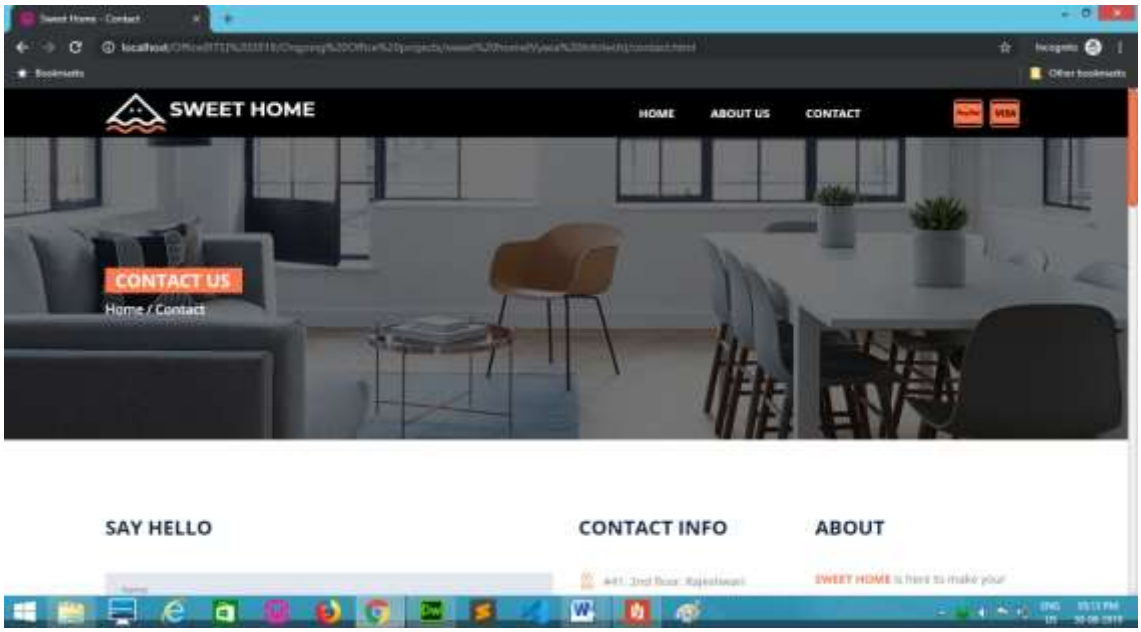
This is the home page of website where in you can see all the options of the website Here u can put your details in enquiry page so that u will be notified with the best rated properties



This is the home page where in you can select whether you want to sell your property or you are looking for a rented property with city u prefer



This is contact us page which will help you in the inconveniences you may get in searching a property



This is About us page where you can get to know about the company and wht company does