

Undertaken By:

M ASfand yar

**Syed Nabeel Abbas** 

**Abdul Qadeer** 

Supervised By:

**Madam Sumra Yasir** 

## DEPARTMENT OF COMPUTER SCIENCE QURTUBA UNIVERSITY DERA ISMAIL KHAN

(2017-2021)



A thesis submitted to Qurtuba University of Science & Information Technology (Dera Ismail Khan) as a fulfillment of necessities for the award of degree of Bachelors of Studies in computer science

Ву	
(2017-2021)	

Sir Dr. Bashir

#### **FINAL APPROVAL**

This is certified that we have read the thesis submitted by **Muhammad Asfand yar, Abdul Qadeer Khan & Syed Muhammad Nabeel Abbas** with registration number **12288 & 12 12**. It is our decision that this thesis of adequate standard to certification its acceptance by the Qurtuba University of Science & Information Technology (Dera Ismail Khan) for the master's degree in computer science.

1. External Examiner		
Sir		
2. Internal Examiner		
Sir		
3. Supervisor		
Madam Sumra Yasir		
4. Head of Department		

#### **DEDICATION**

I dedicate this thesis to my dearest Parents because without their support and love it could not possible for me to complete it, to my respected Teachers who helped me out to develop the system using the modern techniques.

And I would like to thank my family & friends all those people who motivated & encouraged me during the time of my BS degree.

**Muhammad Asfand yar** 

#### **DECLARATION**

I hereby declare this report not a copied from any source. It is additional declared that check to software and attended report entirely based on personal effort, under the sincere guidance of my supervisor, Teachers & Seniors. If any part of this report is copied out from any source I will stand by the consequences.

Signature of Student

#### **ACKNOWLEDGMENT**

"And (He desires) that you should complete the prescribed period and that you should glorify Allah for having guided you and that you may give thanks". (Al-Baqarah 2:185)

I am very grateful to Allah that His blessings and mercies helped me to accomplish this task. I am also very thankful to my parents, siblings and family for their continuous support, prayers, and guidance in the completion of my project and report.

I am immensely grateful to my respected and honorable supervisor "Madam Sumra Yasir" who proposed this useful and fruitful idea and then helped us under his kind supervision to complete this project.

At the end, I would like to thank my batch mates, my friends and especially my group members for their help, support, motivation, and encouragement throughout this final year.

**Muhammad Asfand yar** 

#### **PROJECT IN BRIEF**

Project Title Email Scrapper

Organization Qurtuba University of Science & Information Technology

(Dera Ismail Khan)

Undertaken by Muhammad Asfand yar, Abdul Qadeer & Syed M. Nabeel

**Supervised by** Madam Sumra Yasir

Starting Date -May-2021

**Ending Date** -Oct-2021

Framework Used PHP, My SQLi

**Tools Used** HTML 5, CSS 3, SQL Server

System Used Intel Core i5 inside

#### **PREFACE**

This report represents description of investigation, improvement, employment & analysis of "Email Scrapper". The whole effort is accommodated in following chapters & appendix.

**Chapter One**: Defines the overview, aims & importance of the system & all the problem

statements.

**Chapter Two:** Defines the Existing system and its problems, presenting the new system.

Chapter Three: Comprise of complete requirement analysis, scope & objective,

functional/Non-functional requirements of the system. This aim of prerequisite analysis is to gather requirement & data from the stakeholders

and end users for accomplishment of software.

**Chapter Four:** System Design This part is concerned with the Database design &

development. The complete details of all the tables detail discuss in this

chapter.

**Chapter Five:** Testing Test planning of the system discuss in this chapter. It deals with

testing of the system & accomplish this purpose, designed test plans & test

cases to make sure this system is properly working as per standards.

**Chapter Six:** Conclusion This Chapter concerns with the testing, Conclusions & Future

Work. The testing stage represents that all the proposed services &

functionalities are properly implemented.

**Chapter Seven:** References/ links/ book/website This part has user guide. All information

include that helps the end user to understand & use the application

effectively. In this chapter define the user interface is developed very

simple & easily interact with.

#### **ABSTRACT**

EMAIL SCRAPPER is a web application designed primarily for use in the scrapping email for the user. Email Scraper is a free online tool developed to scrape emails from any websites along with the URL. This online tool can be useful to get the target list of email addresses. This system will allow users to increase scope of email scrapping by reducing the time. The system also allows to quickly and easily to get emails for the users can browse and use to copy and paste URLs with just few clicks.

#### **Table of Contents**

#### **Contents**

Chapter # 1	1
Introduction	1
Introduction	2
Background	2
Motivation	2
Feasibility	3
Technical Feasibility	3
Financial Feasibility	3
Resource Feasibility	4
Operational Feasibility	4
Chapter # 2	2
Requirement Analysis	2
Introduction	6
Product Scope	6
Document Convention	6
Overall Description	7
Product Perspective	7
Product Functionality	7
User and Characteristics	8
Operating Environment	9
Design and implementation constraints	10
User Documentation	11

	Assumptions and Dependencies	11
	Specific Requirements	11
	External Interface Requirements	11
	Functional Requirement	12
	Behavioral Requirements	12
Chapter	# 3	<i>6</i>
System I	Design	6
	System Design	23
	Use Case Diagram	23
	Class Diagram	25
	System Sequence Diagrams	27
	ER Diagram	33
Chapter	# 4	23
Database	e Design	23
4.1	Database Design	35
Chapter	# 5	35
Quality A	Assurance	35
	Introduction of Testing	37
	Testing Levels	37
	Test Cases	38
Chapter :	# 6	37
Conclusi	ion & Future Work	37
	Conclusion	41
	FuSture Work	41
Chapter	# 7	<i>1</i> 1

User Manual	41
Introduction	43
Hardware/Software Requirements for the System	43
Hardware Interfaces	43
Software Interfaces	43
Major Features of the System	43
Installation Guide	43
Operating Manual	43
Home Page	44
Registration Page	44
Login Page	45
Add Foods	46
Show Posted Foods	46
References	47

#### **List of Tables**

Table 1. 1 Financial Feasibility	3
Table 2. 1 Formatting Conventions	6 9
<u>List of Figures</u>	
Figure 7. 1 Home Page	44
Figure 7. 2 Database Page.	45
Figure 7. 3 Email Collector Page	45
Figure 7. 4 Email data collector page	46

## Chapter # 1

### Introduction

#### Introduction

An email parser or scraper is a program designed to extract email addresses from web pages. Such programs can usually extract email addresses from web pages and upload the results to a CSV file format, for example, Excel. For collecting email addresses from the web, professional scrapers usually parse data from social networks (LinkedIn, Facebook, etc.) or forums. If a company needs to find the email addresses of legal entities, it collects the required information from these firms' corporate sites

Email scrapers let you automate the process of collecting the data. Their main advantage is that they do it incredibly quickly. One can find a hundred addresses in a couple of minutes. What's more, the program can save information, process it, and provide it in a graphical form.

#### **Background**

This Case study looks at the problem of setting up a Email scrapper. In existing system there are few problems:

- For taken any E-mail you should be go to personally and ask them to provide their E-mail
- Its very expensive and time consuming.
- You can forgot e-mail or record will be missed.

#### Motivation

Email Extractor is a powerful email extraction extension for Chrome. Extension automatically fetches valid email IDs from the web page, you can copy paste particular email ids you need or export all of them to a text or CSV file.

The new automation tool can discovers all email addresses for a specific domain name. Example: if you want to discover email IDs of people working at Email Hunter, you can step the tool to search for any email address ending with hunter.io or emailhunter.io, this job can be done for any website.

you can add a 5 seconds delay when visiting web pages automatically, so you make sure all pages are loaded and all email ID's are collected, even when JavaScript are delaying the display of the emails ID's.

#### **Feasibility**

The Feasibility of "EMAIL SCRAPPER System" can be ascertained on the following assessments:

#### **Technical Feasibility**

Technical feasibility is based on the assessment of resources and technology. The system requirements are given as follows:

- Bootstrap, HTML, CSS, and JavaScript: Bootstrap Framework will be used for developing responsive website.
- JQUERY: jQuery is a fast, small, and feature-rich JavaScript library will be used for HTML document traversal and manipulation, event handling, animations, and Ajax.
- C#: Server-side coding will be done using C#.

#### **Financial Feasibility**

Possibility of economic aspects e.g. valuing & basis of availability and cost of tools and other assets required for the project are organized below in table 1.1:

Table 1. 1 Financial Feasibility

Resource Name	Price	Source
[Hardware or Software]	[PKR]	[Website URL/Dealer's address]
Sublime Text	Free	https://www.sublimetext.com/download
XAMP	Free	https://www.apachefriends.org/download.html
Laptop	Free	Locally available
SQLi	Free	https://locallhost.me/phpmyadmin
Sponsors/Funding	Self	
Source		

#### **Resource Feasibility**

Resource feasibility is based on the assessment of the hardware and software resources which are required to accomplish to Online Email Scrapping

- All the mentioned table 1.1 software are easily available on the given websites.
- Laptops are already purchased.

#### **Operational Feasibility**

Operational feasibility is based on assessment of cost of hardware and software which are required to accomplish to Online Food Order System.

- The Project is based on Real-World implementation, will be tested on:
- Location: Dera Ismail Khan
- Product Domain: Service
- Localhost: We will test our project using Xampp localhost.

## CHAPTER # 2 EXISTING SYSTEM

#### **Existing System**

It is also possible to use many Email scrappers such as Hunter.io, but they require extensive technical knowledge and have a steep learning curve. Neither are they web-based, so you have to extension on your browser, and then you can use Email scrapping. "Hunter.io" is an email scraping platform with a chrome extension. You can search the email address of a company via domain name. Alternatively, search for a particular professional with email. It starts from \$49/month with 1000 lookups to \$399/month with 50,000 lookups. [https://www.octoparse.com/blog/best-email-scraping-tools-for-sales-prospecting-in-2019]

This system is easy-to-use interface, it is an excellent choice for a first-time self-starter. It can quickly pick up the email addresses on the page. You can also use their search page to find contact information. However, the cost of leads is higher than in others

#### **Problem of Statement**

Let's say there are more much email scrappers that they are working with good and efficient way of scrapping emails and provide more much information which is very use full for the users. In any case, the question is how to extract the information or the Email from the website or the entire Emails of website? How about having something that can download Emails from the internet to your local hard drive?

While, the above mention Email scrapper is very expensive in cost as they mention that [It starts from \$49/month with 1000 lookups to \$399/month with 50,000 lookups.]. One more thing that this tool work totally online and can't save the file into your hard drive.

Amazing, isn't it? Because if you want to do it manually, then probably it will take a hours to complete, it might take a day.

By doing it manually, you may need to go through all pages of a site, copy each Email address. Once you copy the Email address then you need to download it and save it as a proper name. A complete business website has hundreds of web pages.

#### **Proposed system**

Email scrapper has following main features which are given below.

- It is free of cost
- It is freely available
- Its work both online as well as offline
- The data is stored in CSV file, which we can access any time
- Its maintain records in database
- It is easy in use
- You can share your scrapped data easily

# Chapter # 3 Requirement Analysis

#### Introduction

This chapter delivers brief information regarding the Email Scrapper System project's requirement analysis phase. It describes the product-scop, behavior, overall-description, functional requirements, and other requirements of the related project.

#### **Product Scope**

Email scraper online tool works just as well as using Internet search engines but is faster, saving you from searching websites for emails .After you enter targeted website url, my tool uses a script to scan all web pages of your targeted website, remember it only scan the internal links of the website. Then it continues to collect all email addresses from the webpages where it founds and saves them to an excel file .Are you worried to scrape email addresses from web pages, and don't have time to scrape them manually or don't want to spend money? This tool has a powerful multi-threaded email scraper that can grab email addresses from webpages.

#### **Document Convention**

This document uses the following conventions throughout the whole document same convention will use in this document. To choose character sequence some rules are defined which will be used are functions, variables, and other data types.

#### **Formatting Convention**

This part provides us overview of the text format which follows some consistent way. In table defined text font in Times New Roman style and heading will be bold and different sizes.

Table 2. 1 Formatting Conventions

Name	Conventions	
Font	Times New Roman 12pt	
Heading1	Bold, 36pt Size	
Heading2	Bold, 14pt Size	
Heading3	Bold, 12pt Size	
Line Space	1.5 Space	
Justified	Yes	

#### **Naming Convention**

The naming conventions are defined below. This support the reading to identify and understanddissimilar arranging rules applied in the document.

- A term necessitating explanation is defined in "()"
- Upper case letters used to define abbreviations

#### **Overall Description**

Overview of overall system will be detailed discussed in this section. Explanation of the system and its background that how the End-User collaborate with the systems arrangements and function the essential process of it. It will explain how the user uses the features of it. It will also explain what process is attainable for the next. Inference and duress will be discussed and presented at the end.

#### **Product Perspective**

Email Scrapper is a web application, the main aim of this application is to facilitate the user to copy the URLs of emails from different websites and collect in one place with effective and efficient way

#### **Product Functionality**

Product Functionality is the any aspect of product, such as a software application or computing devices, can do for a user.

#### **System**

- The system will show the progress bar showing the percentage that how much Email is copied.
- The system will authenticate the URL and validate it before scraping.
- The system will also tell the user that which task is currently being performed.
- The system will tell the user that how many sites are copied till now

#### User

- The User will add Input URL
- User will get alert if input URL is invalid.
- User will get notifications that the order is placed successfully.

#### **Operating Environment**

This system is web based there will be a need to provide PC server hardware connected to the internet. It is supported on any browser like Google chrome, Mozilla fire fox, Microsoft edge, internet explorer having JavaScript.

Our operating system, hardware and software requirements are mentioned in table 2.2.

Table 2. 2 Operating Environment

Required	Specification
Browser	Chrome, Firefox, Edge, Safari, Opera etc.
Database	SQL
External systems	Cloud Service for database
Hardware platform	Computer, laptops,
IDE	Sublime 3
Operating system	Windows 7 or higher (recommended)
Programming Languages	PHP

#### **Design and implementation constraints**

This part the constraints for planning and employment are given below which this system is accomplished.

#### **Domain Constraints**

Constraints on the system from the domain of operation.

- We will use Sublime Text.
- We will use PHP for server side.

• We will Use HTML 5,CSS 3 for client-side validation.

#### **Language Constraints**

PHP would be used for server side.. HTML5, CSS3 for user interface.

#### **Implementation Constraints**

To implement our system in PHP, Sublime would be used for implementation stage and web hosting servers would be used for final testing stage.

#### **Time Constraints**

Time constraints are documented in a similar manner to business rules and technical requirements. This system is our final year project therefore we must complete within one year.

#### **User Documentation**

The user guide would remain provided to each type of user depending on their role in the system,

To help them understand and guide the user of system.

#### **Assumptions and Dependencies**

Existence of an internet service is required. User must have computer, laptop or smartphone. Availability of server is twenty-four hours.

#### **Specific Requirements**

In this section we will talk in detail about the quality aspect of this application and will be focusing on providing a detail overview of the application modules.

#### **External Interface Requirements**

In this part of the chapter point by point depiction of all the essential requirements of this system like inputs etc., will be given. Also, a brief overview of the tools and models used in order to design the interfaces.

#### User Interface

The system provides a main interface to the Website Copier users.

User

It Provide an interface to the users where user will be able to see a text box input and system will ask user to input a webpage url

#### **Hardware Interface**

There is no hardware interface for this system because it does not require any hardware. It will accessible online. So, users just need to have laptop, mobile and internet connection.

#### **Software Interfaces**

Interface is a contract between the system and the environment. This system software interface discuss are given below:

- MY SQL Server will be used for data storage or database creation.
- Browsers would be used to access the online system.

#### **Functional Requirement**

Functional requirements describe system services. They depend on the type of software, expected users and the type of system where the software is used.

Some basic functionalities of Email Scrapper are:

- The system will check first for a url if It is a valid url.
- System will check if the url is have a webpage or not.

- If the url is have a webpage it will scrape its html content.
- From html content, it will look for all urls for the same domain and put into an array.
- It will then loop through all urls that we just saved into the array.
- Once we get all links, it will then go through each url and copy the content.
- It will create a directory in the same folder, where the script is present or either we can specify the directory path in the class object.
- It will generate directory with name of domain we are trying to copy and with some random numbers.
- Since, we get all contents.

# Chapter # 4 Database Design

#### **Data Base Design**

The database that is used to design the web application is MySQL Server. MySQL Server is used to create tables and run queries. In this application development, we have used MySQL Server to store Website URL, Website Path and Date. Hence, we have identified three columns to achieve desired functionality.

#### Data table: holds details of websites copied

When a user enters the targeted website url with the application, the application inserts the details of the website into the data table. Similarly, when a the page loads, the application queries the database to retrieve the details of the total website copied

```
DB.php
 <?php
class DB {
    private $query;
    private $queryRun;
    public function __construct($host,$username,$password,$dbname){
        $this->con = @mysqLi_connect($host,$username,$password,$dbname);
            if(!$this->con):
                die("Connection error: " . mysqli_connect_error());
    public function insert($table,$fields,$param){
        $this->query = "INSERT INTO ".$table."(".$fields.") VALUES(".$param.")";
        $this->queryRun = mysqli_query($this->con,$this->query);
        return $this->queryRun;
    public function query($query){
        // echo $query;
        $this->queryRun = @mysqLi_query($this->con,$query);
        return $this->fetch_data($this->queryRun);
    private function fetch_data($query){
        $count = 0;
            $data = array();
            while ( $row = mysqli_fetch_object($query)):
             $data[$count] = $row;
                  $count++;
            endwhile;
        return $data;
    public function last inserted id(){
        return $id;
```

Online Food Order System

## Chapter # 6 Conclusion & Future Work

#### **Conclusion**

Email Scrapper stands as a revolutionizing element in the sphere of Scrapping Emails Online. As you can observe, email scrapper has its own unique advantages and limitations. Moreover, it will depend a lot on your respective and specific needs. You should, to start with, identify your needs and study the software in comparison to those needs. The main purpose is to automate the scraping process, offline fetching, viewing. Apart from that, there can be other number of uses.

#### **Future Work**

The following section describes the work that will be implemented with future releases of the software.

- **Website Linking:** we allow other website owners to link it with their website. We create our own website and link this in it as additional work.
- Sale out: We can sale out our project as per owner will
- **Private** If any organization want this project privately we can provide.

## Chapter # 7

### **User Manual**

#### Introduction

This Chapter is to guide the user way to use the system. This chapter will show all the details which will allow user to easily move around this system.

#### Hardware/Software Requirements for the System

Specific requirements of any system are called "System requirements" which include software, hardware, operating system, and some other virtual requirements. Email Scrapper also has some requirements and those are given below.

#### **Hardware Interfaces**

Email scrapper is a web system. Compatible with all the popular desktop and mobile browsers i.e. Google Chrome, Mozilla Firefox, Opera, Safari etc. This system is developed in PHP development tool. You need internet connection to access the system.

#### **Software Interfaces**

Email scrapper will run on internet browsers as it is a web application.

#### **Installation Guide**

As it is a web application and it does not require any installation, all user need to do is to open browser and type this system's URL and they can have access features public.

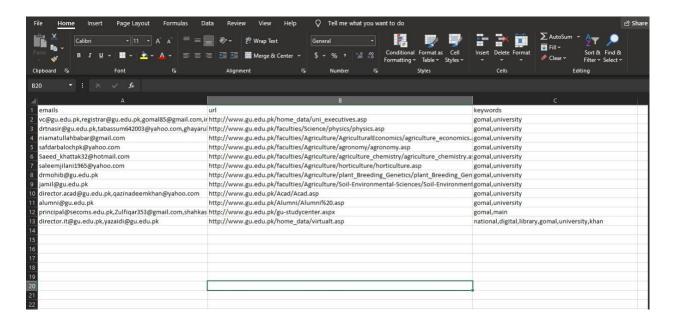
#### **Operating Manual**

It will guide the user to know how different features of the system are used. Following these instructions will guide user to perform different features of this system.

#### **Home Page**

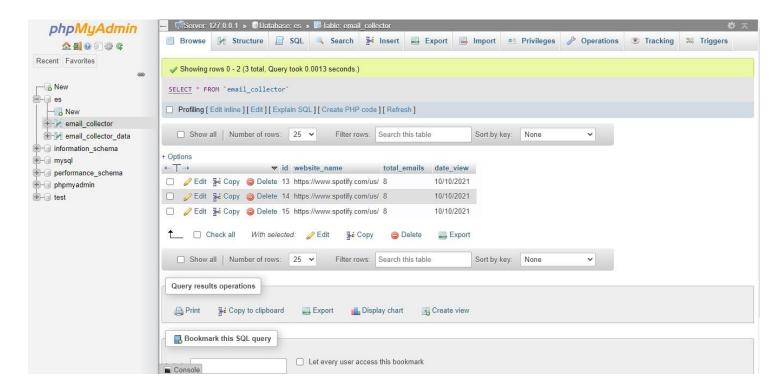
Home page is the front face of any website, it should be attractive and stick to its purpose and so does this system's Home page do. Online Foods Order System Home page is as shown in figure 7.1.

#### **Excel File**



Excel Page

#### **Email Collector**



Email Collector page

