A PROJECT REPORT

ON

"WEB PORTAL ON CONFERENCE ALERT"

Submitted to

UNIVERSITY OF MUMBAI

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER ENGINEERING

BY

Shaikh Uzair Ahd Fayyaz Amina	**13CO59**
Mohammad Hamid Abulkhair Sahana Bano	**13CO46**
Bind Rahul Achehelal Vidya	**13CO20**
Siddiqui Mohd.Sharique Salahuddin Zarina	**11CO45**

UNDER THE GUIDANCE OF
PROF. ANSARI MUKHTAR

DEPARTMENT OF COMPUTER ENGINEERING
Anjuman-I-Islam's Kalsekar Technical Campus
SCHOOL OF ENGINEERING & TECHNOLOGY
Plot No. 2 3, Sector - 16, Near Thana Naka,
Khandagaon, New Panvel - 410206

2017-2018

AFFILIATED TO

UNIVERSITY OF MUMBAI

A PROJECT II REPORT
ON

"WEB PORTAL ON CONFERENCE ALERT"

Submitted to
UNIVERSITY OF MUMBAI

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER ENGINEERING

BY

Shaikh Uzair Ahd Fayyaz Amina	**13CO59**
Mohammad Hamid Abulkhair Sahara Bano	**13CO46**
Bind Rahul Achehelal Vidya	**13CO20**
Siddiqui Mohd.Sharique Salahuddin Zarina	**11CO45**

UNDER THE GUIDANCE OF
PROF. ANSARI MUKHTAR

DEPARTMENT OF COMPUTER ENGINEERING
Anjuman-I-Islam's Kalsekar Technical Campus
SCHOOL OF ENGINEERING & TECHNOLOGY
Plot No. 2 3, Sector - 16, Near Thana Naka,
Khandagaon, New Panvel - 410206

2017-2018
AFFILIATED TO

UNIVERSITY OF MUMBAI

Anjuman-I-Islam's Kalsekar Technical Campus

Department of Computer Engineering
SCHOOL OF ENGINEERING & TECHNOLOGY
Plot No. 2 3, Sector - 16, Near Thana Naka,
Khandagaon, New Panvel - 410206

CERTIFICATE

This is certify that the project entitled

WEB PORTAL ON CONFERENCE ALERT

submitted by

Shaikh Uzair Ahd Fayyaz Amina	**13CO59**
Mohammad Hamid Abulkhair Sahana Bano	**13CO46**
Bind Rahul Achehelal Vidya	**13CO20**
Siddiqui Mohd.Sharique Salahuddin Zarina	**11CO45**

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Engineering) at ***Anjuman-I-Islam's Kalsekar Technical Campus, Navi Mumbai*** under the University of MUMBAI. This work is done during year 2017-2018, under our guidance.

Date: / /

Prof. Ansari Mukhtar
Project Supervisor

Prof. Kalpana Bodke
Project Coordinator

Prof. Tabrez Khan
HOD, Computer Department

DR. ABDUL RAZAK HONNUTAGI
Director

External Examiner

Acknowledgements

We would like to take the opportunity to express my sincere thanks to my guide **Prof. ANSARI MUKHTAR**, Assistant Professor, Department of Computer Engineering, AIKTC, School of Engineering, Panvel for his invaluable support and guidance throughout my project research work. Without his kind guidance & support this was not possible.

We are grateful to him/her for his timely feedback which helped me track and schedule the process effectively. His/her time, ideas and encouragement that he gave is help me to complete my project efficiently.

We would like to express deepest appreciation towards **DR. ABDUL RAZAK HONNUTAGI**, Director, AIKTC, Navi Mumbai, **Prof. TABREZ KHAN**, Head of Department of Computer Engineering and **Prof. KALPANA BODKE**, Project Coordinator whose invaluable guidance supported us in completing this project.

At last we must express our sincere heartfelt gratitude to all the staff members of Computer Engineering Department who helped me directly or indirectly during this course of work.

Shaikh Uzair Ahd Fayyaz Amina(13CO59)

Mohammad Hamid Abulkhair Sahana Bano(13CO46)

Bind Rahul Achehelal Vidya(13CO20)

Siddiqui Mohd.Sharique Salahuddin Zarina(11CO45)

Project II Approval for Bachelor of Engineering

This project entitled ***Ẅeb Portal on Conference Alert¨*** by ***Shaikh Uzair Ahd Fayyaz Amina (13CO59), Mohammad Hamid Abulkhair Sahana Bank (13CO46), Bind Rahul Achehelal Vidya (13CO20), Siddiqui Mohd.Sharique Salahuddin Zarina (11CO45)***is approved for the degree of ***Bachelor of Engineering in Department of Computer Engineering.***

Examiners

1. .....
2. .....

Supervisors

1. .....
2. .....

Chairman

.....

Declaration

We declare that this written submission represents my ideas in my own words and where others ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Shaikh Uzair Ahd Fayyaz Amina
Roll Number:13CO59

Mohammad Hamid Abulkhair Sahara Bano
Roll Number:13CO46

Bind Rahul Achehelal Vidya
Roll Number:13CO20

Siddiqui Mohd.Sharique Salahuddin Zarina
Roll Number:11CO45

ABSTRACT

We are highlighting on the topic of Conference Alert, which is very fundamental approach for the people. Those who are interested in conference alert. For such types of people at the beginning has to visit several sites to collect the information about the conference, it requires a lots of time wasting to achieve the proper information. We are developing a web-portal on conference alert through the scraping from different web sites. Where all the users will get proper information about the conference without the wasting time and it will also notify to the user on which date and time the conference will be going too held on. The library that we are going to use to extract the data from different websites by using Beautiful Soup. It is an incredible tool for pulling out the information from a web pages. You can use it to extract tables,lists, paragraph and you can also put filters to extract information from web pages. Python idioms and few simple methods are provided by Beautiful Soup for navigating, searching, and modifying a parse tree. It is a toolkit for dissecting and extracting a document or information what you need. It required less code to develop an application.

Keywords: Web Scraping, Data analysis, Notification, Beautiful Soup

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A conference is generally understood as a meeting of several people to discuss a particular topic. A conference differs from the others in terms of knowledge and purpose, the term can be used to cover the general concept. It is a gathering of delegates representing several groups. At a conference, innovative ideas are thrown about and new information is exchanged among experts. This technology can help people, organizations who wants to organize any conference or want to join the conference it can also help to the college who wants to published the paper. Scraping is the act of extracting data or information from websites, with or without the consent of the website owner. Web scraping is a computer software technique of extracting information from web sites. This technique mostly focuses on the transformation of unstructured data (HTML format) on the web into structured data (database or spreadsheet) .Scraping can be done manually, or automated. In most cases, it's the latter because of its efficiency. Scraping of content or prices is mostly done with malicious intents, and there are several techniques by which this is done.

1.1 Purpose

The purpose of this project is for the user who not get exact information about the conference on which date and time. The different platform and technologies are for developing this web portal. We recommended different point on the basis of the conference alert to be held on the given date and time. We will also scrap the data from another websites and will display it on our portal which will be beneficial for the users who search the information on different sites.

1.2 Project Scope

Once the user register on our web portal so all upcoming events and conference information will be notified to the users via message, e-mail before a week.If users gives response to then Notification so the users will get reminder before a day through message or e-mail.

1.3 Project Goals and Objectives

1.3.1 Goals

Our main goal us to provide scarped data from the different conference websites.This scraped data is stored on webhostapp and on phpmyadmin. After scraped data is stored online we provide this data to the Android Application.

1.3.2 Objectives

The aim of this project we provide a web sites or web portal for the user to get the correct information about the conference,paper publication and event.After scraping the data, different list of the conferences and will generate the report.This report of the different conference will reported to the Android Application.This application will display the data of different list of websites related to the conferences.

1.4 Organization of Report

Chapter 1 gives a brief introduction about our project.

Chapter 2 describes the literature review of the existing papers and the description about the application.

Chapter 3 talks about the project planning and different roles and capability of the team member.Also talks about the budget of the project.

Chapter 4 describe the brief description of the srs and the other requirement of the projects

Chapter 5 shows the system design,functional requirement and different diagram of the projects.

Chapter 6 shows the implementation of the different conference websites and coding.

Chapter 7 shows the different testings performed and the problems faced. It also shows snapshots of the current working application.

Chapter 8 is the closure to the book and tries to conclude the work in the project and also mentions the future scope as to where it would be used. Chapter

Chapter 9 is a step by step guide about using the final product

Chapter 2

Literature Survey

2.1 The use of web scraping in computer parts and assembly price comparison

If originally computers were used only as a tool to perform some calculations, nowadays computers has a lot of functions to help people finish their tasks in almost every aspect of human life.As a lot of various functions computers have, they also need different specifications for each computer so they can do their tasks according to their functionality. Therefore this application was build with a purpose to recommend a solution to its users in assembling computes suited to their needs. This application also has a price comparison feature based on data sources retrieved from five computer shops so the users can save the costs of purchasing PC parts and assembling the computer easier. This comparison feature is based on a basic consumer's principal which are basically they wanting to buy items not only with the lowest price but also expect the best quality as possible. The research starts with the deployment of questionnaires to some respondents who had bought computer parts or assembled a computer online. This questionnaire is made to assure that all features which previously has been specified by the author is appropriate to user needs.Then, in order to obtain required data from five computer shops, the author use Pentaho Software as a tool to do web scraping and web grabbing method. These methods allow the application to obtain data from those five computer shops. The result of this research is a web-based application built in PHP and javascript with MySQL as its database.

2.1.1 Advantages of Paper

a. This technology help the user to finish their task very easily.

b. The user extract the data in sequential manner.

c. This is best technology for the user for scarping.

2.1.2 Disadvantages of Paper

a. The technology use for data is might cause concern.

b. This is very time consuming

c. This technology use in paper is very costly

2.1.3 How to overcome the problems mentioned in Paper

In our project we have used a a python library which gives in-built scraping tools such as urlib2 for fetching the url of the websites and Beautiful Soup is used for scrap the data from that url which has been fetched by urlib2. This gives very reliable and easy accessable feature to the user.

2.2 Evaluate a Personalized Multi Agent System through Social Networks: Web Scraping

Many new applications have been recently developed to satisfy users special needs on the web.In this context, we are interested in personalized systems and particularly in Personalized Multi-Agent Systems (PMAS) characterized by collective and intelligent resolution in a distributed and parallel environment. This work assesses personalization, the most important characteristic of interface in multi-agent systems. As a few studies dealt with the personalization assessment in a multi-agent system, we try, in this work, to address this issue by focusing on web scraping and crawling social networks.In fact,we propose a new assessment tool that exploits data from user's web navigation in order to improve the delivered personalization, which makes the evaluation process more valuable.

2.2.1 Advantages of Paper

a. The best application recently used is web.

b. The web application is very personalized multi-agent system.

c. This application is characterized in distributed environment.

2.2.2 Disadvantages of Paper

a. The main issue is focusing on web scraping.

b. The assessment tool exploits the data from the web application.

c. The big disadvantage is data security for data from the web.

2.2.3 How to overcome the problems mentioned in Paper

In our project we have used a a python library which gives in-built scraping tools such as urlib2 for fetching the url of the websites and Beautiful Soup is used for scrap the data from that url which has been fetched by urlib2. This gives very reliable and easy accessiable feature to the user.

2.3 Framework for Data Scraping and Semantization

Most of the enormous amount of information from the internet is available just like web pages made for a human reader. They don't have any common interface for accessing, searching or browsing the data. Hence, it's hard to extract the semantic data from the web, categorize them and keep them updated. For this purpose we have designed and implemented a system called Agent Mat. This system is designed for efficient extraction of large amount of data from the web pages. Agent Mat processing is based on an XML-based language describing the given extraction task in a declarative way. The task description consists of system components, which connected together are able to perform the desired functionality on a general web page. Thanks to this scraping system the raw contents from the irregularly updated and unstructured web pages can be kept categorized and accessed together with the semantic metadata. In our pilot implementation we have built the MediaPub system, which extracts the information from various webpages, does automatic categorizing and checks for duplicities.

2.3.1 Advantages of Paper

a. This system is designed for efficient extraction of data.

b. The system component tool connect together to perform desired functionality.

c. We have built MediaPub system which extract the data from various web pages.

2.3.2 Disadvantages of Paper

a. Hard to extract the data from semantic web pages.

b. We don't have any common interface for browsing or searching the data.

c. Difficulty to updated or categorize data.

2.3.3 How to overcome the problems mentioned in Paper

In our project we have used a a python library which gives in-built scraping tools such as urllib2 for fetching the url of the websites and Beautiful Soup is used for scraping the data from url.

2.4 A dive into Web Scraper world

This paper talks about the World of Web Scraper, Web scraping is related to web indexing,whose task is to index information on the web with the help of a bot or web crawler. Here the legal aspect, both positive and negative sides are taken into view. Some cases regarding the legal issues are also taken into account. The Web Scraper's designing principles and methods are contrasted, it tells how a working Scraper is designed. The implementation is divided into three parts: the Web Crawler to fetch the desired links, the data extractor to fetch the data from the links and storing that data into a csv file. The Python language is used for the implementation. On combining all these with the good knowledge of libraries and working experience, we can have a fully-fledged Scraper. Due to a vast community and library support for Python and the beauty of coding style of python language, it is most suitable for Scraping data from Websites.

2.4.1 Advantages of Paper

a. The python is best for scarping the data.

b. It is most suitable for scarping the data from the websites.

c. The bot is helpful for whose task is to index information from the web.

2.4.2 Disadvantages of Paper

a. The web scarping method are contrasted.

b. Not use of this technology is hard to do scarpped the data.

c. Network traffic traces unable to scarpped the data.

2.4.3 How to overcome the problems mentioned in Paper

We will use urlib2 for fetching the url of the websites and Beautiful Soup is used for scraping the data from url.

2.5 Technical Review

2.5.1 Scrapping

Scraping is very fundamental approach in the web system .it is the act of extracting data or information from websites, with or without the consent of the website owner In our project we will scrap the data or information from different websites which is related to conference or paper publications and data will be scraped by Keywords.

2.5.2 Libraries required for Scraping

Urllib2: It is a Python module which can be used for fetching URLs. It defines functions and classes to help with URL actions (basic and digest authentication, redirections, cookies, etc.)In our project we will use Urllib2 to fetch the URL from different website after it will be scraped by Beautiful Soup to the particular information from that URL which has been scraped by the Urllib2.

Beautiful Soup: It is an incredible tool for pulling out information from a web pages.You can use it to extract tables, lists, paragraph and you can also put filters to extract information from web pages .It does not fetch the URL it only scrap the information therefore we will use both Beautiful Soup as well as Urllib2.

2.5.3 Advantages of Technology

a. It is a open use technology.

b. It is best programming technology.

c. It is easy for scraping and fetching.

2.5.4 Reasons to use this Technology

a. We use this technology for fetching different conference websites.

b. We also use this technology for scraping.

c. This technology helps to scrapped the structured data.

Chapter 3

Project Planning

3.1 Members and Capabilities

Table 3.1: Table of Capabilities

SR. No	Name of Member	Capabilities
1	Mohammad Hamid	Python Programming
2	Shaikh Uzair	Documentation
3	Bind Rahul	Presentation
4	Siddiqui Mohd.Sharique	GUI Design

3.2 Roles and Responsibilities

Table 3.2: Table of Responsibilities

SR. No	Name of Member	Role	Responsibilities
1	Mohammad Hamid	Team Leader	Fetching the data from url
2	Shaikh Uzair	Team Member	Documentation
3	Bind Rahul	Team Member	Scarping the structured data
4	Siddiqui Mohd.Sharique	Team Member	GUI Designing

3.3 Assumptions and Constraints

3.3.1 Assumption

The assumption of our project is to assume a data that based on the user knowledge,user experience and useful information is available on hand.We assume that the data we provide is purely true because this purely data is manage and stored online and must be secure from the unauthorized user.

3.3.2 Constraints

In our project,we make schedule for a project to complete on time based on different constraints that required in our project.We may also include the scope of the project

and the cost of the project that required for completing the project.Different quality attributes in projects and resources required in project.No risk tolerance is present in our project.

3.4 Project Management Approach

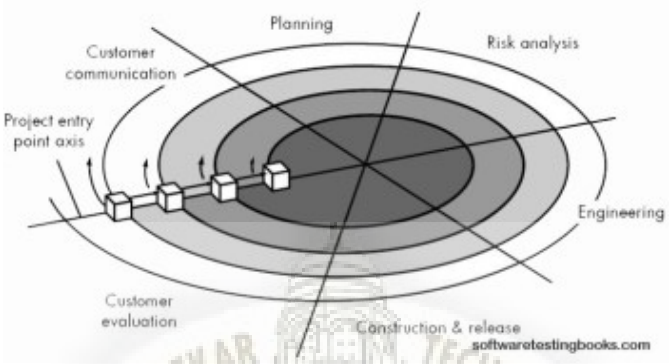

Figure 3.1: Model of our project

In our project we had used spiral model for implementing all the phases successfully.This model involves strategies, which is a combination of incremental and prototype models. This model is suitable for planning and implementing to achieve the goal of the project.It maintains a systematic step wise approach.

These are the different phases involved in our project:

3.4.1 Planning

In any project planning phases are most important phase whenever we are going to make any project.So we need to gather proper information related to our project so therefore we had searched different websites which are related to conferences to understand the structure of the websites to scarp.

3.4.2 Risk Analysis

1) Identify the Proper data.
2) Fetching the data as per user request.
3) Identify the Proper Structure for fetching the data.

3.4.3 Engineering Phase

Testing are also important for any system so before implementation of the project first we have to also test the cases that we are going to implement in our project.we

have used Beautiful soup and request library.once we will integrate these two libraries only parsing part will be remaining that we will get from the website's structure such as Html tag which is used in website's to built that is about to scrap.once our fetching part will be done then we have to check that we are getting the data from the website's which we have targeted based on the website's tag such as html tag.Here for testing purpose we have targeted a website's that the related to the conference. when we are implement the testing part so we are successfully getting the data from the website's that we have targeted to scrap the data such as conference date,title, location.

3.4.4 Evaluation

User's involvement takes place in this Evaluation phase.If users wants any specific data such as computer department so there will be displayed only the data that are related to the computer department.

3.5 Ground Rules for the Project

After using our web portal on conference alert our ground rule is that user does not have to go to many websites in search of what conference he wants that is related to conference . Instead he can subscribe to our newsletter in order to get full information in a single place so in our system user don't need to go to different website's to get the information.

3.6 Project Budget

1) Beautiful Soup: **Free Open Source**

2) Request Library: **Free Open Source**

3) Date-Util: **Free Open Source**

3.7 Project Timeline

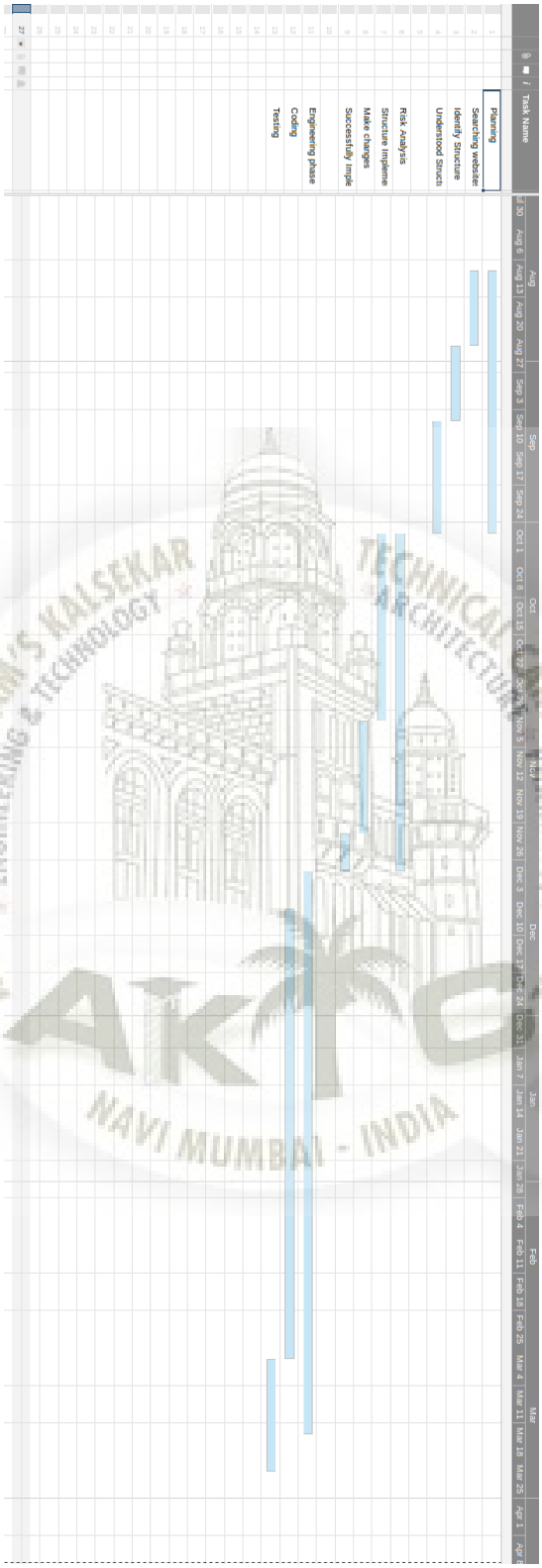

Figure 3.2: Gantt Chart

Chapter 4

Software Requirements Specification

4.1 Overall Description

This Software Requirement Specification is the requirement work product that formally specifies the web portal on conference alert.The objectives of this document therefore is to formally describe the system's high level requirements including functional requirement,non-functional requirement business rules and constraints.

4.1.1 Product Perspective

The various system tool that have been used in developing the back-end and other tools of the project are being discussed in this section The back-end is implemented using MySQL which is used to design the database. MySQL is the world second most widely used open source relational database management.The SQL phrase stands for structured query.And PHP is a server side scripting language designed for web development but also used as a general purpose programmming language.PHP code is interpreted by a web server with PHP processor module which generates the resulting webpages.

4.1.2 Product Features

The system will provide all the data related to the conference to the user.Depending upon the user's role,he/she will be access the data related to the conference after giving a keyword. Managing the database by converting them into json file.This made easily to see the different conference data just providing keyword(For example:keyword=Computer science).Other data related to the conference will see on the android applications.

4.1.3 User Classes and Characteristics

Educational Level: At least graduate and should be comfortable with English language.
Technical Expertise: Should be a high or middle level employee of the organization comfortable with using general purpose applications on a computer.

4.1.4 Operating Environment

We use the Linux Operating Environment for running the Python software.We use minimum 250GB HardDisk, and we use version of the operating 18.0.We use different software like Atom

4.1.5 Design and Implementation Constraints

Hardware Requirement:

1)Minimum 500GB space of Hard-Disk.
2)Minimum 100MB space of memory.

Software and Technologies:

1)MySQL:MySQL is the most popular Open Source Relational SQL Database Management System. MySQL is one of the best RDBMS being used for developing various web-based software applications.

2)Python:Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language.

4.2 System Features

4.2.1 System Feature

1)Control Presentation.
2)Creation of data.
3)Organization through web applications.

Description and Priority

The requirements for this feature set describe how the system provides and controls presentation, creation, and organization throughout the Web Application.The system's users are provide information and features related to the conference from which all of their communication with the system will take place. The conference is related to the meeting and perform scraping and sorting on the data.

Stimulus/Response Sequences

Stimulus: A user wants to provide a different data of the conference.
Response: The system creates an api or json file and then provide keyword to display conference to the system

Stimulus: A user defines a new term in their personal glossary
Response: The user's personal glossary is updated, and places links to other people's definition of the term

Stimulus: A user wants to organize the various information related conference that are currently looking for it.
Response: The user's workspace allows the user to organize the information related to conference he/she is currently looking at.

Functional Requirements

User Interface:
a)The software provides good graphical interface for the user any administrator can operate on the system, performing the required task such as create, update, viewing the details of the book.
b)Allow user to view the quick reports.
c)Verification and searching facility based on different criteria.

Hardware Interface
Operating system: Linux
Hard disk: 40GB
RAM:256MB
Processor:Pentium(R)Dual-Core CPU

Software Interface:
Python language
MySQL
Atom Editors

4.3 External Interface Requirements

4.3.1 User Interfaces

The Web Server must provide a user interface that will acessible through any internet browser the major ones being Google Chrome and internet Explorer 12.

4.3.2 Hardware Interfaces

We don't required any hardware interface in our project.So we required only software interface in our project.

4.3.3 Software Interfaces

1)MySQL:MySQL is the most popular Open Source Relational SQL Database Management System. MySQL is one of the best RDBMS being used for developing various web-based software applications.

2)Python:Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language.

4.3.4 Communications Interfaces

Connections to the system will be over TCP/IP connections

4.4 Nonfunctional Requirements

4.4.1 Performance Requirements

The system must be interactive and the delays involved must be less.When we connecting to the server the delay is because the data is stored or manage online very safely and securily.The data is reliable to the user to see this data very correctly.

4.4.2 Saftety Requirements

The data that use for implementation which concerned with the possible loss or harmful used of the data.The data is stored online is very secure because these data is access by only authorized user by providing username and password to the webapp.The external policies and safety issue that the product design must be satisfied.

4.4.3 Security Requirements

The server on which the Online Data is stored will have its own security to prevent unauthorized write/delete access. There is no restriction on read access. The use of email by an Author or Reviewer is on the client systems and thus is external to the system. The PC on which the database resides will have its own security. Only the Editor will have physical access to the machine and the program on it.

Chapter 5

System Design

5.1 System Requirements Definition

We have made a system which will scrap the data from different different websites and stored into the database then it will analyze the data as per user requirement based on keyword and display it in one place which has been scrapped the data from different websites .so once system will get online it will scrap the websites if there will be any entry comes into the database .if any new entry or data does not come or any duplicate entry will come so system will automatically skip the data and will not store any data which has been already scrapped and stored into the database.we have made the system in python language and also used some python libraries which are suitable to scrap the data from the websites so this system will be beneficial for those who search the information that are related to conference from different difference websites so for that types of user there are no need to go to the different websites to collect the information .our system will automatically scrap the data and display it at one place.

5.1.1 Functional requirements

1)Fetching:Its play a significant role in our project.We have used request library to fetch the URL of the conference websites then the data is obtained in raw format.It can be further modified into the formatted data by Beautiful Soup.

2)Parsing:Once the fetching process is completed so we have to parsing the data from the websites based on HTML tag,Id.

3)Store:Store function is very essential to store the data.In our project it will store the data into the database.

5.1.2 Use Case Diagram

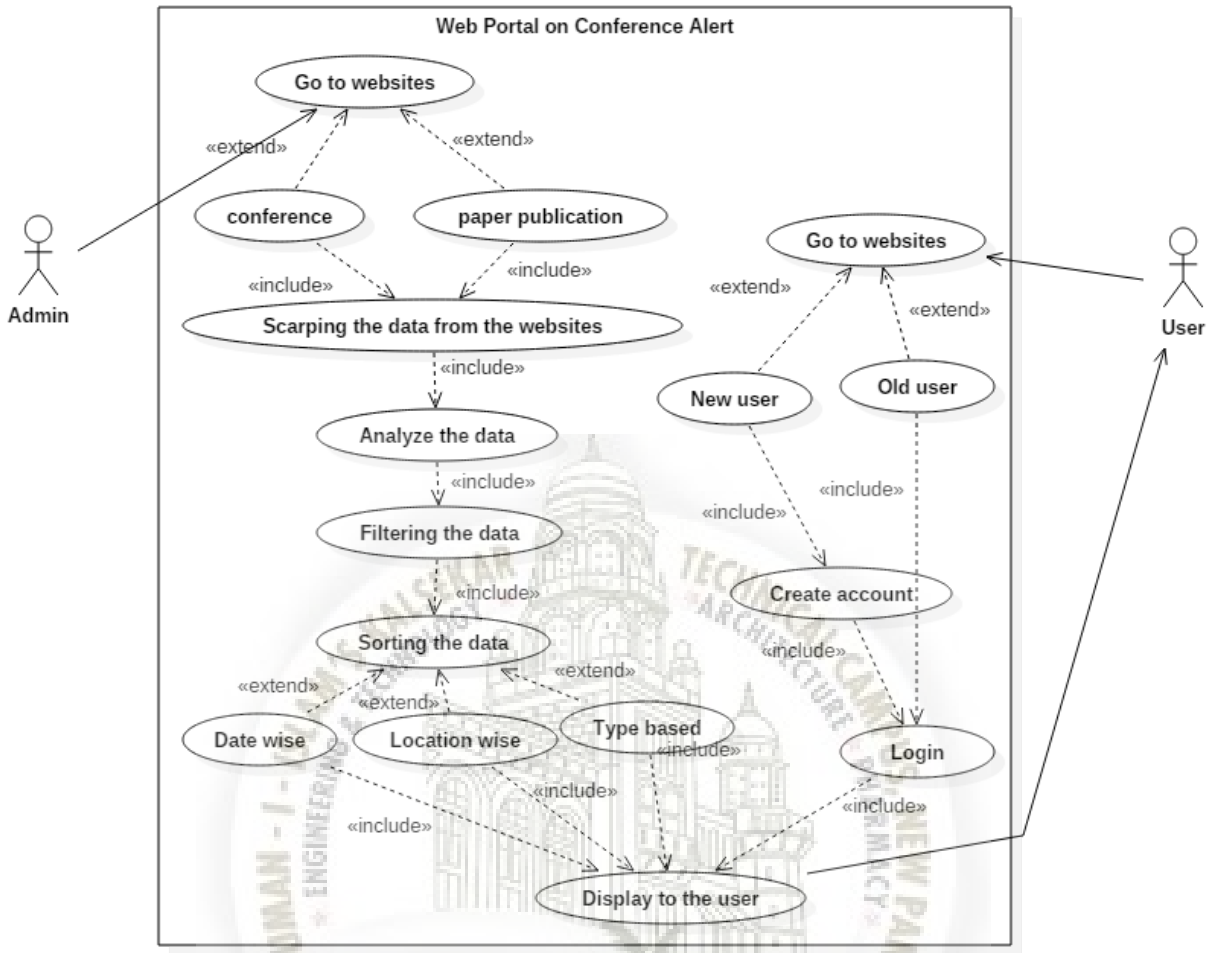

Figure 5.1: Use Case Diagram

There are 2 actor one is data scrapper and other is User.There are some websites which data scrapper is going to scrap which are related to conference, paper publication. Data scrapper will analyze the data which has been scrapped and data is filtering which is unwanted.After filter the data Data scraper has to Sort the data on the basis of Data wise, location wise,Type base then it will display to the Authenticate user.Second Actor is user if there is new user so first user has to register on our portal then login otherwise new user cannot see the content.once user is register and login then user can see the content if the old user so they can see content on the basis of his or her requirement.

5.1.3 Data-flow Diagram

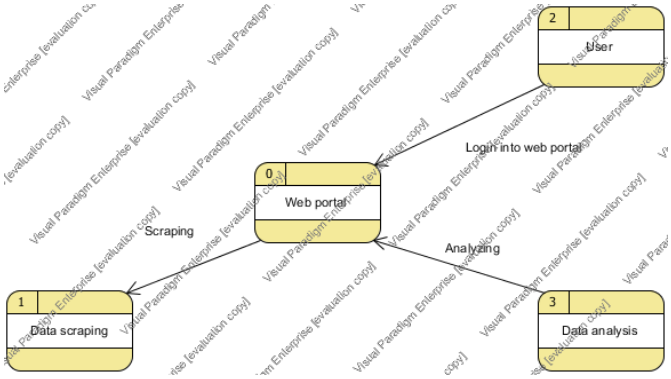

Figure 5.2: Data Flow Diagram

DFD Level 0:- It contains total no.of 4 process in our DFD level 0 diagram.It has web portal,user,data analysis and and data scrapping has the part of our process.Firstly the authenticated user has to login in web portal.Further the scrapping is done from the different websites and the required data is analyzed to achieve the requirement of the user.

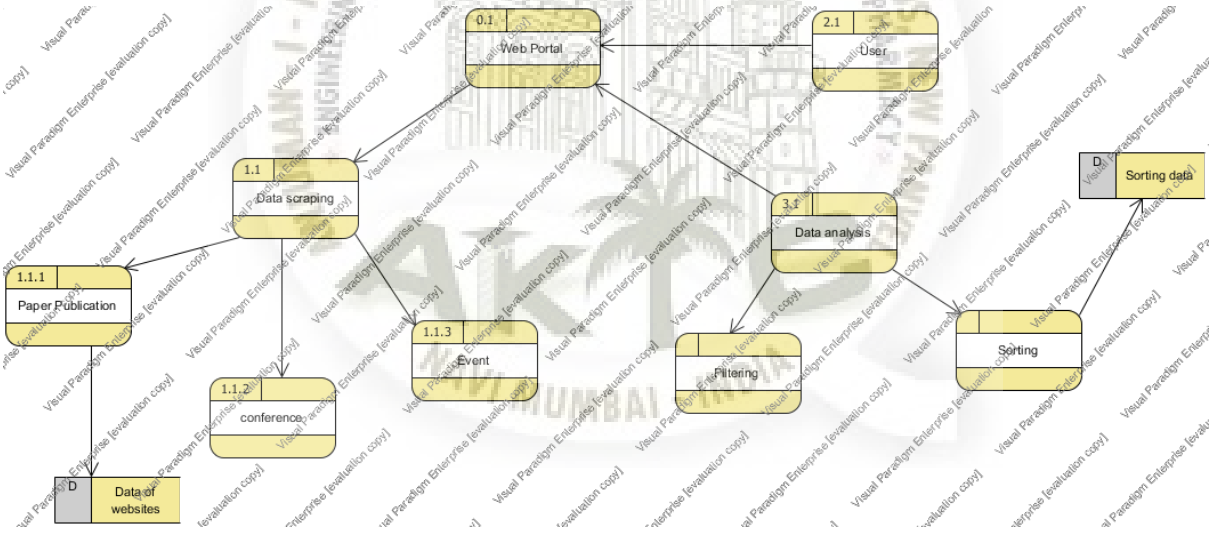

Figure 5.3: Data Flow Diagram

DFD Level 1:- In DFD level 1 the level 0 is enhanced into a greater extend to show the proper clearification and the data flow of the project.In this the above process is enhance such as the data analysis is done on the basis of filtering and sorting the data.After this the required the data is stored in the database.Data scrapping is done on the fundamental approaches which focuses on major process such as paper publication,conferences and events.

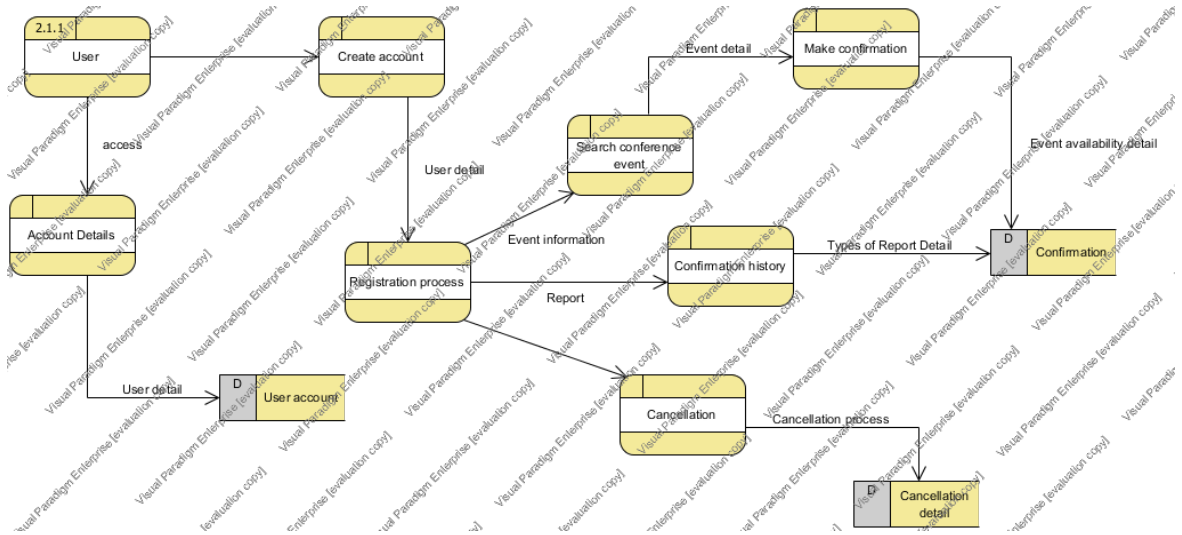

Figure 5.4: Data Flow Diagram

DFD Level 2:- In DFD level 2 the user process is enhanced.It contain two type of user authenticated and unauthenticated. The authenticated user can login and they can access our conference web-site.The unauthenticated user does not have the right access the conferences web-site.It is mandatory for them to go through with the registration process. Along with this they can make confirmation,cancellation and search conference event.All the activity of the user stored in the database.

5.1.4 System requirements (non-functional requirements)

We have made the system in python language and also used some python libraries which aresuitable to scrap the data from the websites so this system will be beneficial for thosewho search the information that are related to conference.

The system must be interactive and the delays involved must be less.

The data is stored online is very secure because these datais access by only authorized user.

The data is stored online is very secure because these datais access by only authorized user.

5.2 System Architecture Design

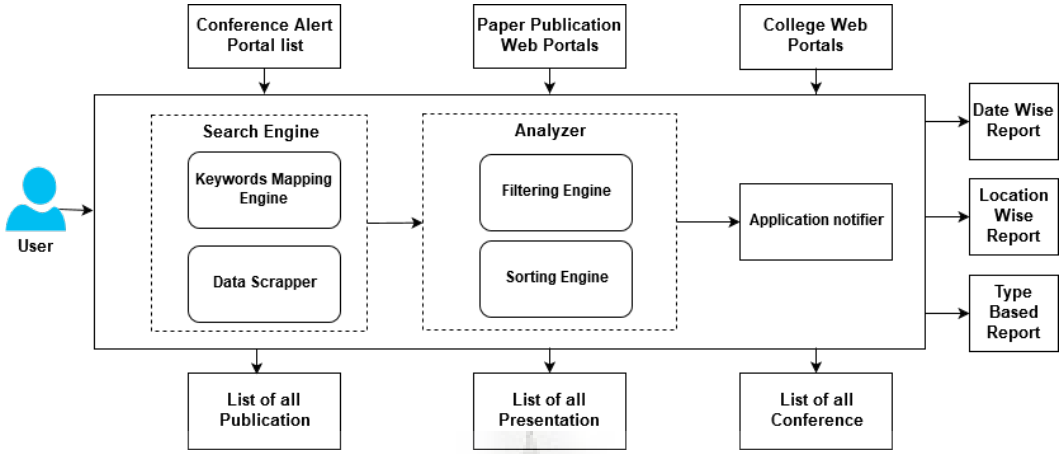

Figure 5.5: System Architecture

5.3 Sub-system Development

In our project we will scrap the data or information from different websites which is related to conference or paper publications and data will be scraped by Keywords.It is an incredible tool for pulling out information from a web pages. You can use it to extract tables, lists, paragraph and you can also put filters to extract information from web pages .It does not fetch the URL it only scrap the information therefore we will use both Beautiful Soup as well as Urllib2.There are also need to analyze the data which has been scraped.In data analyzing there is 2 sub module Filtering engine and Sorting engine.

5.3.1 Search Engine

In our system the search engine will search the information from the different websites based on the keyword. After that data scraper will scraper the data using different python libraries i.e, Beautiful Soup

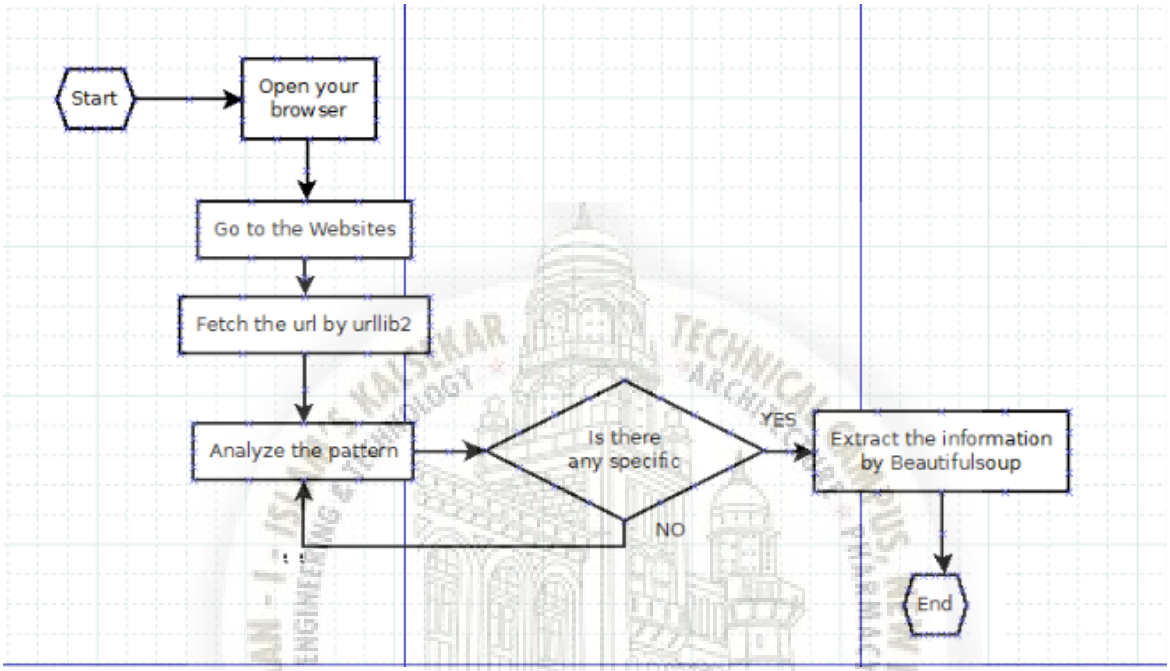

Figure 5.6: Search Engine

5.3.2 Analyzer

After scraping of data, the data will filter in the filtering engine.Filtering of data means filtered the unwanted data.In sorting engine the data will sorted and generate a report based on Datewise,Location-wise and Type-based.Notify to the Users based on user requirement

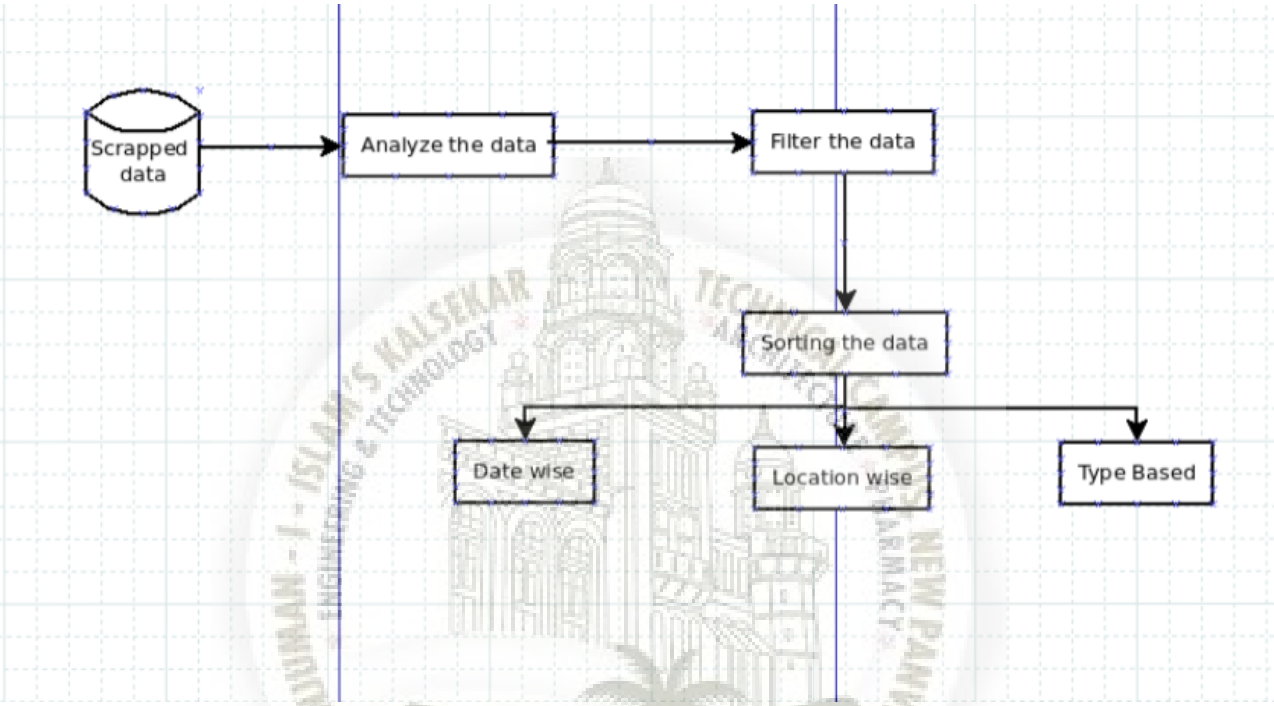

Figure 5.7: Analyzer

5.4 Systems Integration

5.4.1 Sequence Diagram

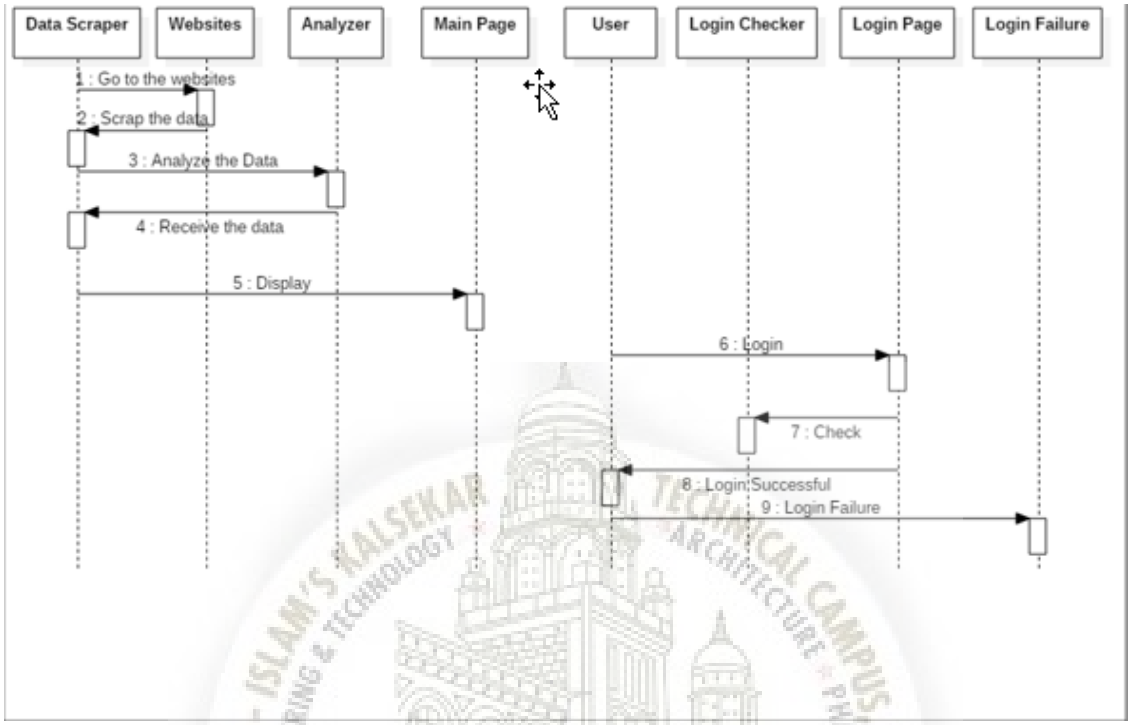

Figure 5.8: Sequence Diagram

There are 9 lifelines in our system in sequence diagram.Data scrapper will go to the websites and fetch the url from different web-sites by urllib2 and also data will scrap by Beautiful soup.Data will be analyzed which has been scarpped and received to the data scrapper and that data will display on main page which will be only visible for the authenticate user.So new user first has to register to our portal and logged in.It will go to the login checker if user is authenticated so it will refer to the main page and display the content which will be related to the user.Otherwise if user is not valid so it go the login failure.

5.4.2 Component Diagram

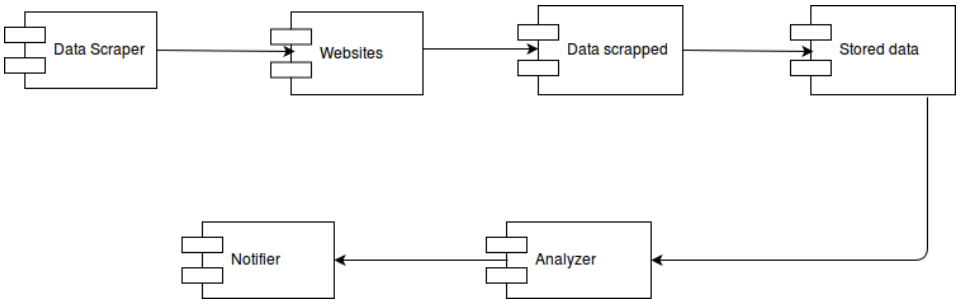

Figure 5.9: Component Diagram

Data scrapper has a major function for the extraction of data.It can be done on various conferences websites.With the help of data scrapper scrapping is done and the scrap data is then stored. Further the data is analysed to get the required result by the data analyzer.And then the analyzed data is notify to the user by departmental wise ,date wise and location wise successfully

5.4.3 Deployment Diagram

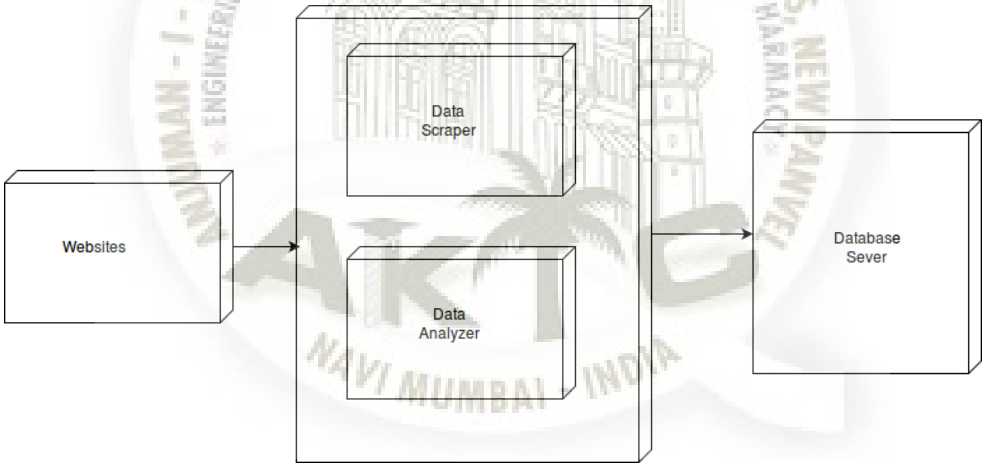

Figure 5.10: Deployment Diagram

From the different various websites the data had been scrapped by the data scrapper.Later on it has been the data scrapped data is further processed by the data analyzed.And the resulted data is then stored in the database server

Chapter 6

Implementation

6.1 Data from Conference Alert

Figure 6.1: Output of Conference Alert

```
from websites import Websites
import dateutil.parser as dp


class ConfAlertCOM(Websites):

    def parse(self):
        if self.error:
            return
        print(f"{self.name}... now scraping!!")
        all_trs = self.soup.find_all('tr')
        self.month = ""
        self.span = ""
```

```
        for row in all_trs[:-2]:
            cells = row.find_all('td')

            if len(cells) == 1:
                try:
                    if cells[0]['id'] == 'eventMonthHeading':
                        self.mon = cells[0].text

                except Exception as e:
                    pass

            if len(cells) == 2:
                try:
                    self.date = f"{cells[0].text} {self.mon}"
                    self.date = dp.parse(self.date, fuzzy=True)
                    self.date = self.date.strftime("%Y-%m-%d")

                    self.spans = cells[1].find_all('span')

                    self.title = self.spans[0].text.strip()
                    self.location = self.spans[1].text.strip()
                    self.link_id = cells[1].find('a')['href'].split('=')[1]
                    self.link = (
                        "https://conferencealerts.com/"
                        "show-event?id="
                        f"{self.link_id}"
                    )
                    # print(self.date)
                    self.store()
                    # print(self.link)

                except Exception as e:
                    pass
        print(f"{self.name}.. completely scrapped and new data stored")


if __name__ == "__main__":
    ca = ConfAlertCOM("https://conferencealerts.com/"
                      "advancedSearch?"
                      "searchCountry=100_India&"
                      "advancedSearchTerm=",'confcom')

    ca.fetch()
    ca.parse()
```

6.2 Data from Conference Alert in

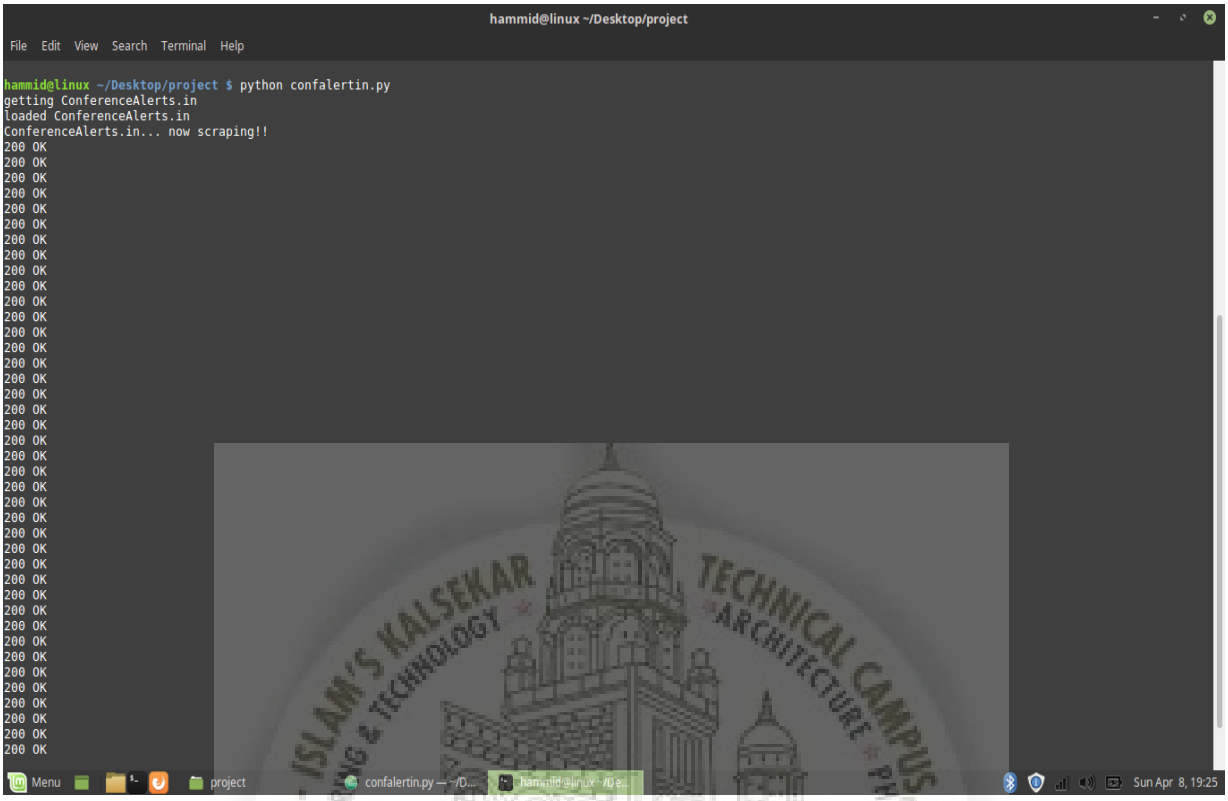

Figure 6.2: Output of Conference Alert in

```python
from websites import Websites
import dateutil.parser as dp


class ConfAlertIN(Websites):

    def parse(self):
        if self.error:
            return
        print(f"{self.name}... now scraping!!")
        all_trs = self.soup.find_all('tr')
        for row in all_trs:

            cells = row.find_all('td')
            if len(cells) == 3:
                try:
                    self.date = cells[0].text.strip()
                    self.date = dp.parse(self.date, fuzzy=True)
                    self.date = self.date.strftime("%Y-%m-%d")

                    self.title = cells[1].text.strip()
                    self.location = cells[2].text.strip()
                    self.link_id = cells[1].a['href'].split('=')[1]
                    self.link = (
                        "https://www.conferencealerts.in/"
                        "Event-Detail.php?EV-id="
                        f"{self.link_id}"
                    )
                    # print(self.date)
```

```
                    self.store()

                except Exception as e:
                    # print(e)
                    pass
        print(f"{self.name}.. completely scrapped and new data stored")


if __name__ == "__main__":
    ca = ConfAlertIN('https://www.conferencealerts.in/','confin')
    ca.fetch()
    ca.parse()
```

6.3 Data from All Conference

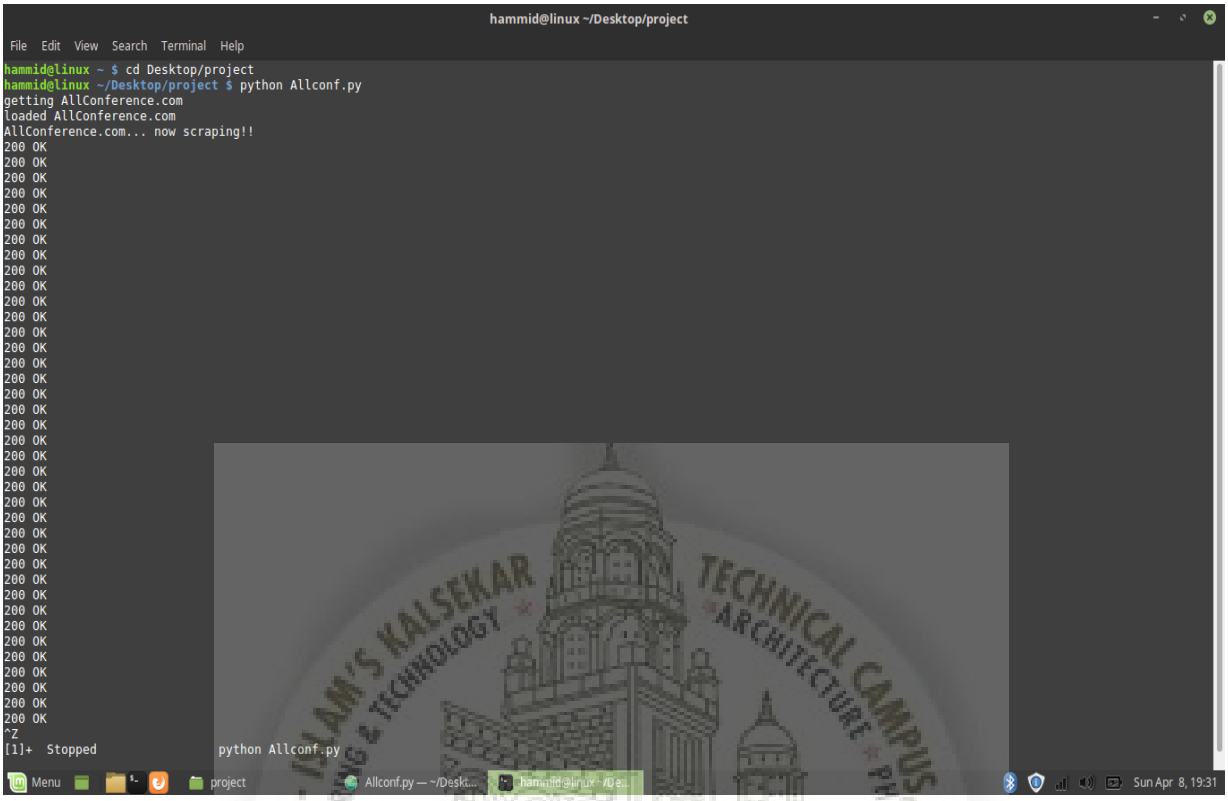

Figure 6.3: Output of All Conference

```
  from websites import Websites
import dateutil.parser as dp


class Allconf(Websites):

    def parse(self):
        if self.error:
            return
        print(f"{self.name}... now scraping!!")
        all_trs = self.soup.find_all('tr')
        rows = all_trs[10].find_all('tr')
        for row in  rows:
            try:
                tds = row.find_all('td')
                if len(tds) == 1:
                    # print(tds)
                    date = tds[0].find('p')
                    if(date):
                        val = date.text.split()
                        self.mon = (val[0]+' '+val[1])


                if len(tds) == 2:
                    self.date = f"{self.mon}"
                    self.date = dp.parse(self.date, fuzzy=True)
                    self.date = self.date.strftime("%Y-%m-%d")

                    self.title = tds[1].find('a').text.strip()
```

```
                    spans = tds[1].find_all('span')
                    self.location = spans[0].text.strip()+spans[1].text.strip()
                    self.link_id = tds[1].find('a')['href'].split('=')[1]
                    self.link = ("www.allconferencealert.com/"
                    "event_detail.php?"
                    f"ev_id={self.link_id}"
                    )

                    self.store()
            except Exception as e:
                raise e

        print(f"{self.name}.. completely scrapped and new data stored")


if __name__ == "__main__":
    ca = Allconf('https://www.allconferencealert.com/india.php','allconf')


    ca.fetch()
    ca.parse()
```

6.4 Data from World Conference

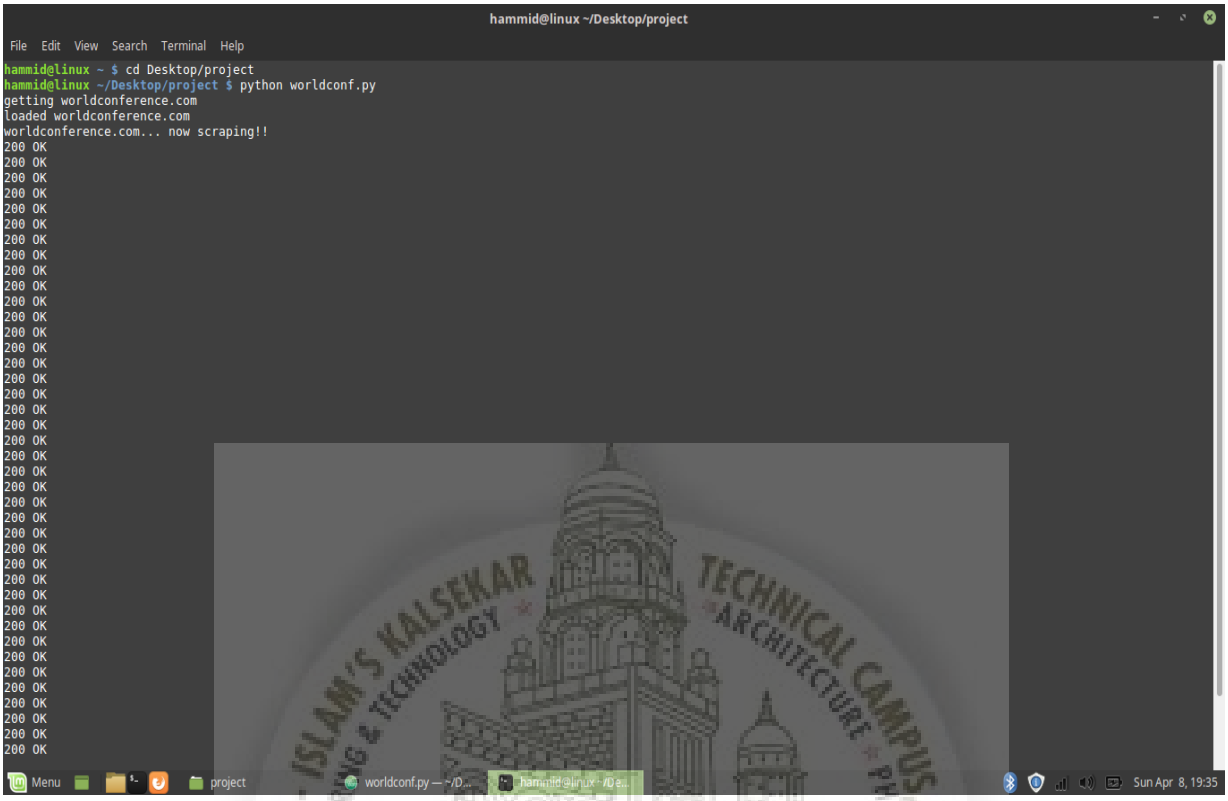

Figure 6.4: Output of World Conference

```
  from websites import Websites
import dateutil.parser as dp


class WorldConf(Websites):

    def parse(self):
        if self.error:
            return
        print(f"{self.name}... now scraping!!")

        trs = self.soup.find_all('tr')
        for row in trs[9:]:
            try:
                tds = row.find_all('td')
                spans = tds[0].find_all('span')
                self.date = spans[0].text+' '+spans[1].text+' '+spans[2].text
                self.date = dp.parse(self.date, fuzzy=True)
                self.date = self.date.strftime("%Y-%m-%d")

                self.title = tds[1].find('a').text.strip()
                self.link_id = tds[1].find('a')['href'].split('&')[0].split('=')
                    [1]
                self.link = ("https://www.worldconferencealerts.com/"
                    "ConferenceDetail.php?EVENT="
                    f"{self.link_id}")
                self.location = tds[1].find('span').text.strip()

                # print(self.date)
```

```
                # print(self.title)
                # print(self.location)
                # print(self.link)
                self.store()

            except Exception as e:
                # raise e
                pass


        print(f"{self.name}.. completely scrapped and new data stored")


if __name__ == "__main__":
    wc = WorldConf('https://www.worldconferencealerts.com/India.php','worldconf')
    wc.fetch()
    wc.parse()
```

6.5 Fetching the Data from Four Websites

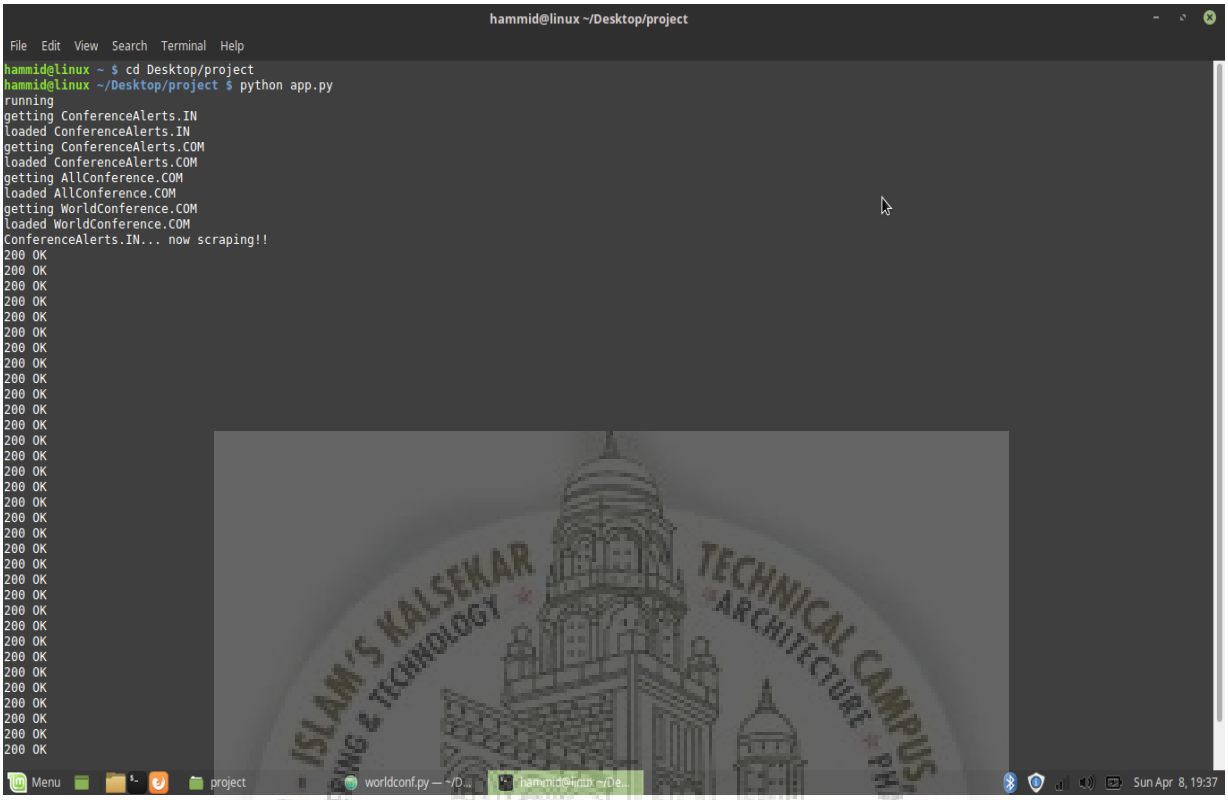

Figure 6.5: Fetching Data

```
from time import sleep
from confalertin import ConfAlertIN
from confalertcom import ConfAlertCOM
from Allconf import Allconf
from worldconf import WorldConf


class App:
    def __init__(self):
        self.websites = [ConfAlertIN('https://www.conferencealerts.in/','
            ConferenceAlerts.IN'),
                         ConfAlertCOM("https://conferencealerts.com/"
                                      "advancedSearch?"
                                      "searchCountry=100_India&"
                                      "advancedSearchTerm=", 'ConferenceAlerts.COM'
                                          ),
                         Allconf('https://www.allconferencealert.com/india.php',
                             'AllConference.COM'),
                         WorldConf('https://www.worldconferencealerts.com/India.
                             php','WorldConference.COM')]
        self.is_active = True


    def fetchWebsites(self):
        for i in self.websites:
            i.fetch()

    def parseWebsites(self):
```

```
25          for i in self.websites:
26              i.parse()
27
28      def run(self):
29          while self.is_active:
30              sleep(5)
31              print("running")
32              self.fetchWebsites()
33              self.parseWebsites()
34
35
36  if __name__ == "__main__":
37      app = App()
38      app.run()
```

6.6 Stored Database on Server

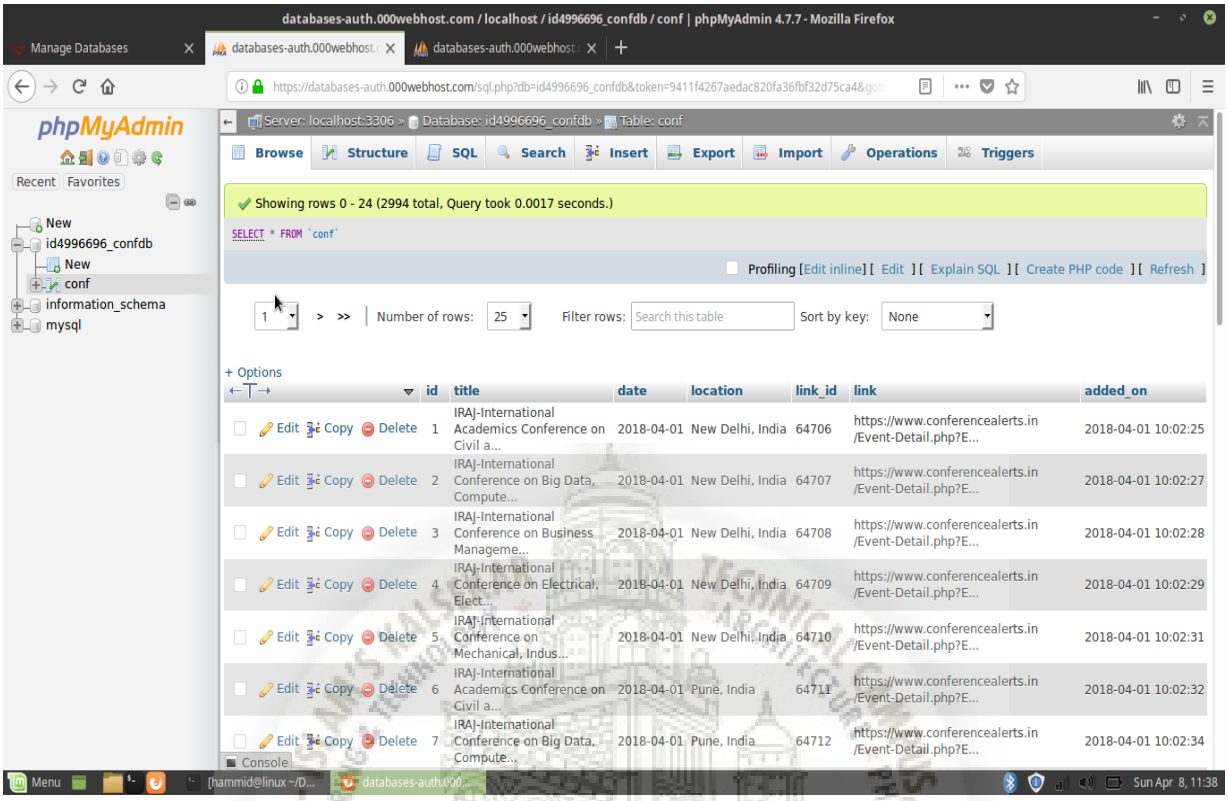

Figure 6.6: Database on Server

```php
  <?php
if ($_GET['keyword']){
  $keyword = $_GET['keyword'];
}else{
  $keyword = "";
}


// $servername = "localhost";
// $username = "root";
// $password = "root";
// $dbname = "confdb";



$servername = "localhost";
$username = "id4996696_hammid";
$password = "12345678";
$dbname = "id4996696_confdb";


// Create connection
$conn = new mysqli($servername, $username, $password, $dbname);
// Check connection
if ($conn->connect_error) {
    die("Connection failed: " . $conn->connect_error);
}

$sql = "SELECT * FROM conf where title like '%$keyword%'";
```

```
30
31
32 $result  = $conn->query($sql);
33
34
35 $rows = array();
36 while($r = $result->fetch_assoc()) {
37   $rows[] = $r;
38 }
39 header('Content-type: application/json; charset=utf-8');
40 echo json_encode($rows);
41
42 $conn->close();
43
44 ?>
```

6.7 Manage Database using API

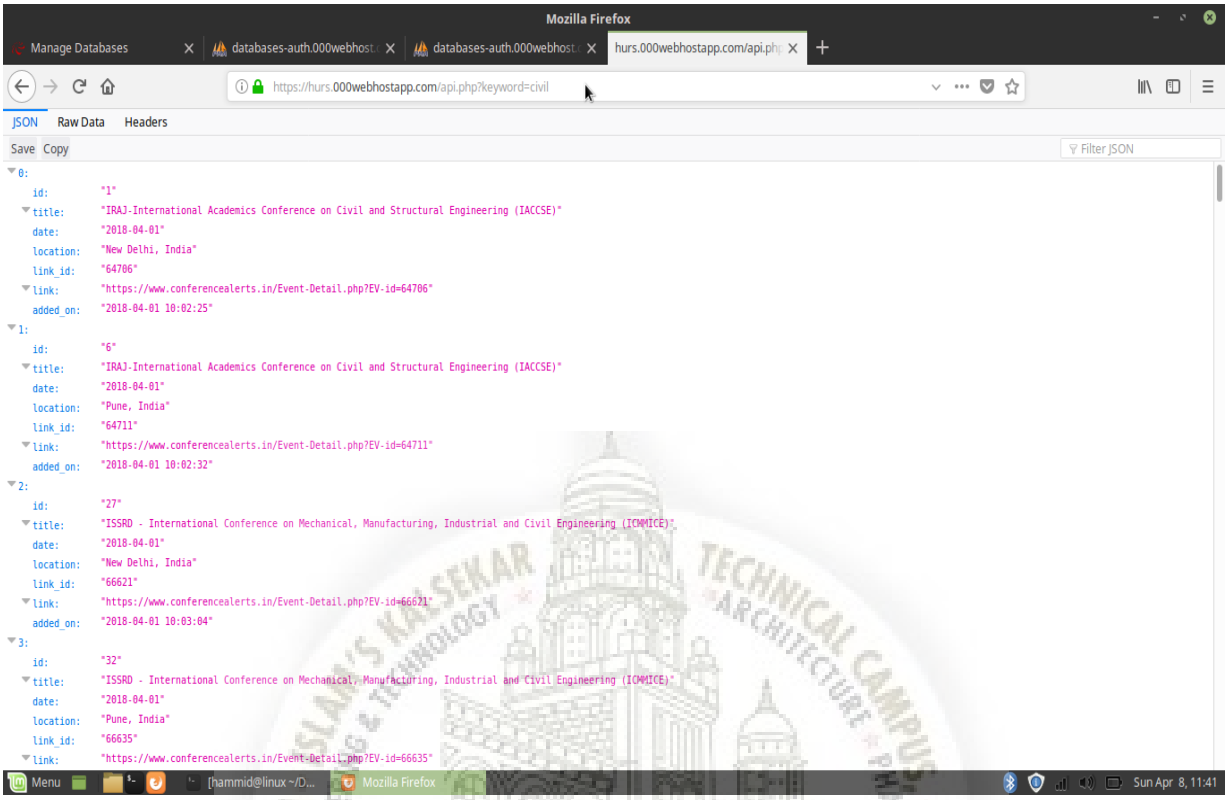

Figure 6.7: Manage Data Using Api

```
<?php

if ($_SERVER["REQUEST_METHOD"] == "POST") {
// if(TRUE){

    $title = addslashes($_POST['title']);
    $date = addslashes($_POST['date']);
    $location = addslashes($_POST['location']);
    $link_id = addslashes($_POST['link_id']);
    $link = addslashes($_POST['link']);




$servername = "localhost";
$username = "id4996696_hammid";
$password = "12345678";
$dbname = "id4996696_confdb";

// $servername = "localhost";
// $username = "root";
// $password = "root";
// $dbname = "confdb";

// Create connection
$conn = new mysqli($servername, $username, $password, $dbname);
// Check connection
```

```
if ($conn->connect_error) {
    die("Connection failed: " . $conn->connect_error);
}

$sql = "SELECT id FROM conf where link = '$link'";

$result  = $conn->query($sql);

if($result->num_rows > 0){

}else{
  $sql = "INSERT INTO conf(title, date, location, link_id, link)

  VALUES ('$title', '$date', '$location', '$link_id', '$link')";


   if ($conn->query($sql) === TRUE) {
       echo "New record created successfully";
   } else {
       echo "Error: " . $sql . "<br>" . $conn->error;
   }
}




$conn->close();
}
?>
```

Chapter 7

System Testing

First system will check the parsing function if that is implemented successfully so it will go the fetch function and take the data from the websites based on the structure that we have mentioned in the fetch function.If the fetch function implemented successfully then it will go to the store function and store the data into the database.

7.1 Test Cases and Test Results

Test ID	Test Case Title	Test Condition	System Behavior	Expected Result
T01	Testing Library	Is it working?	Loaded websites after fetching	Successfully
T02	Test function1	Parsing websites	completely scraped the data	Sucessfully
T03	Test Function2	Store Data	Stored into the database	Sucessfully

7.2 Test Case

Title: Scraping the data from the websites successfully.
Description: Before implementation part testing are also important for any system so before implementation of the project first we have to also test the cases that we are going to implement in our project. our project is first scrap the data from websites and stored into the database so to the website's to scrap we have used Beautiful soup and request library.Once we will integrate these two libraries only parsing part will be remaining that we will get from the website's structure such as Html tag which is used in website's to built that is about to scrap.once our fetching part will be done then we have to check that we are getting

the data from the website's which we have targeted based on the website's tag such as html tag.
Here for testing purpose we have targeted a website's that the related to the conference. when we are implement the testing part so we are successfully getting the data from the website's that we have targeted to scrap the data such as conference date,title, location

Precondition: There is no authentication are required to the users.

Assumption: a supported browser is being used.

Test Steps:

1. Implementing the fetching function.
2. Implementing the parsing function.
3. Getting the data from the websites.
4. Implementing the store function
5. Stored the data into the database.

Expected Result: To get the data from the websites based on the websites structure.

Actual Result: We are successfully getting the data from the website's that we have targeted to scrap the data such as conference date,title, location.

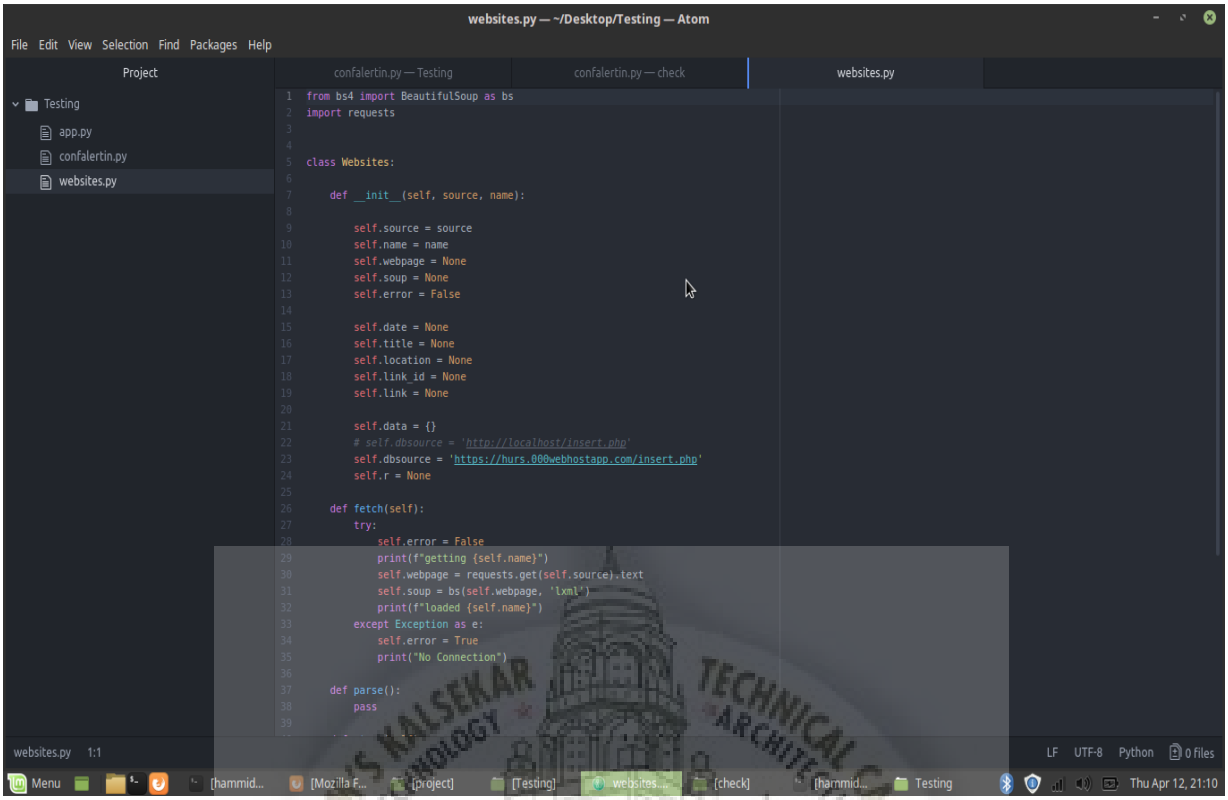

Figure 7.1: Library Tool

In the first figure using of the fetching function for fetch or scarp the data from the different websites that are related to conference.

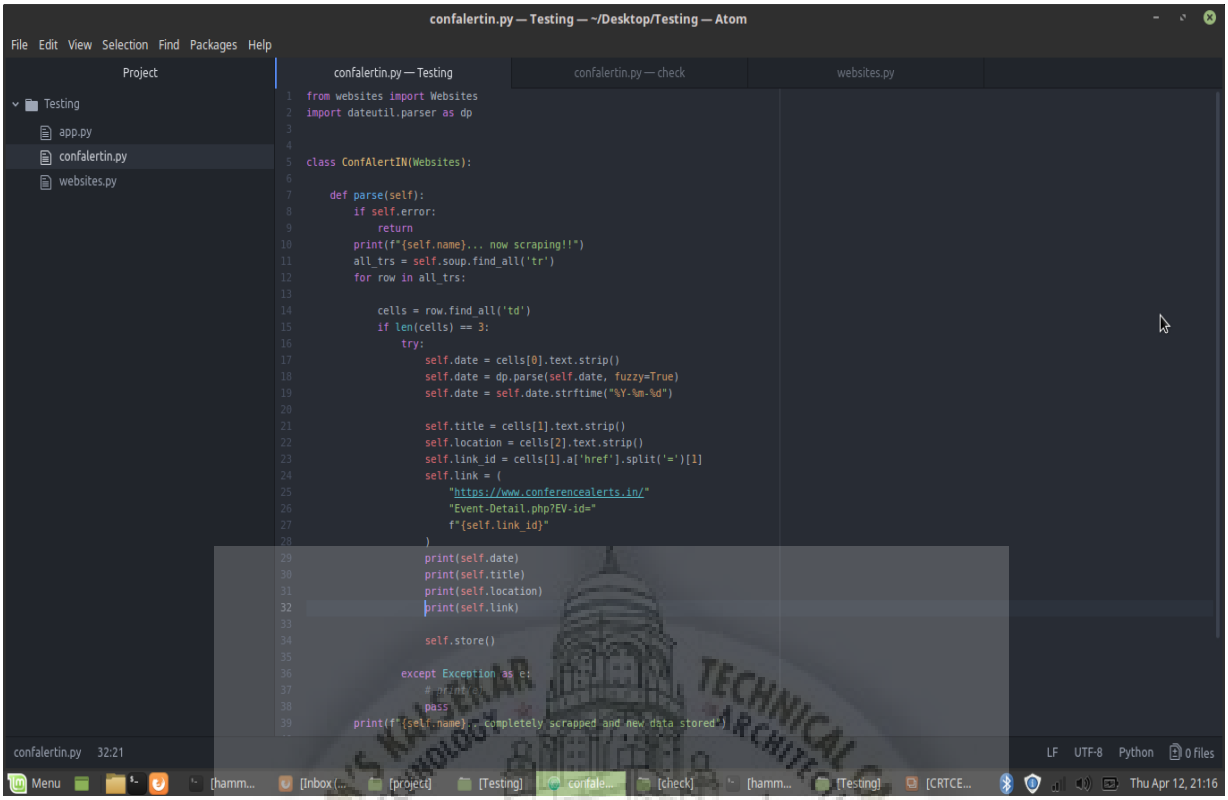

Figure 7.2: Testing 1

In the second figure using of the parsing function for parse the structured data based on the date,time and locations.

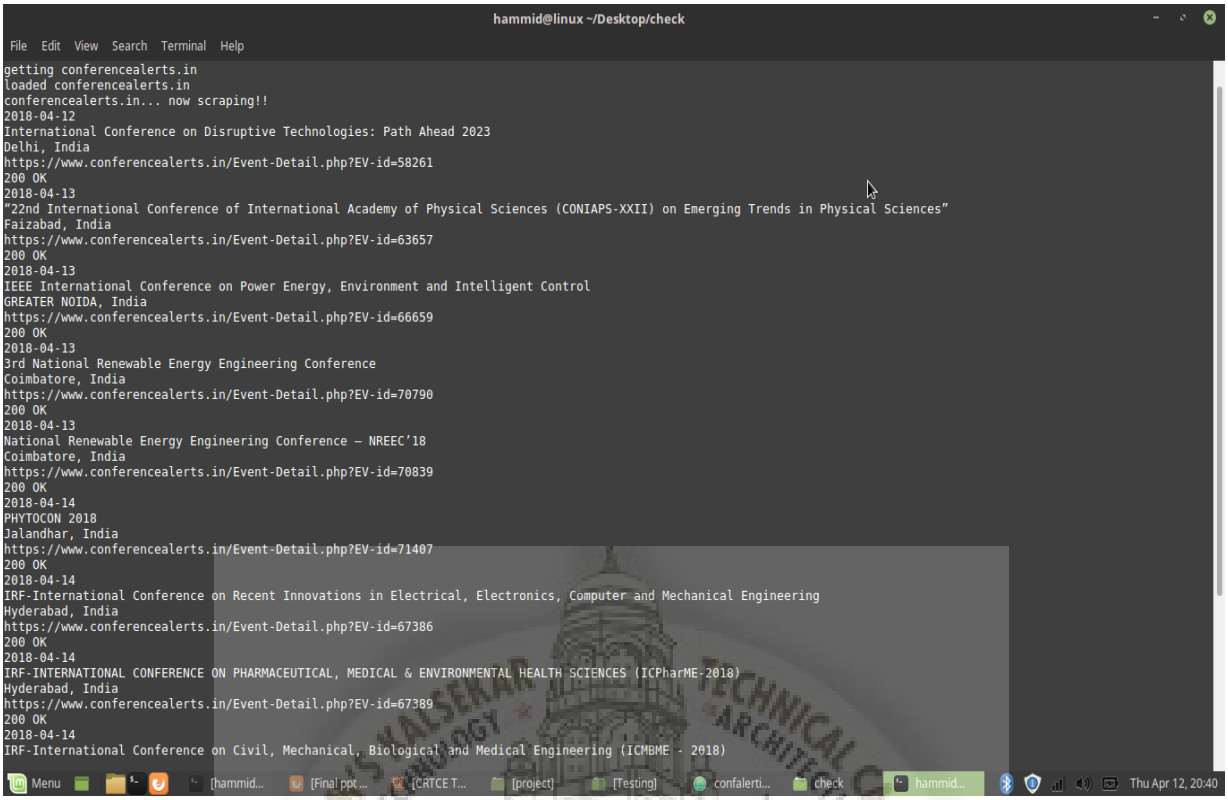

Figure 7.3: Testing 2

In the last figure scraping is completely done and getting the data from different websites related to the conference.

Chapter 8

Screenshots of Project

8.1 Front Page

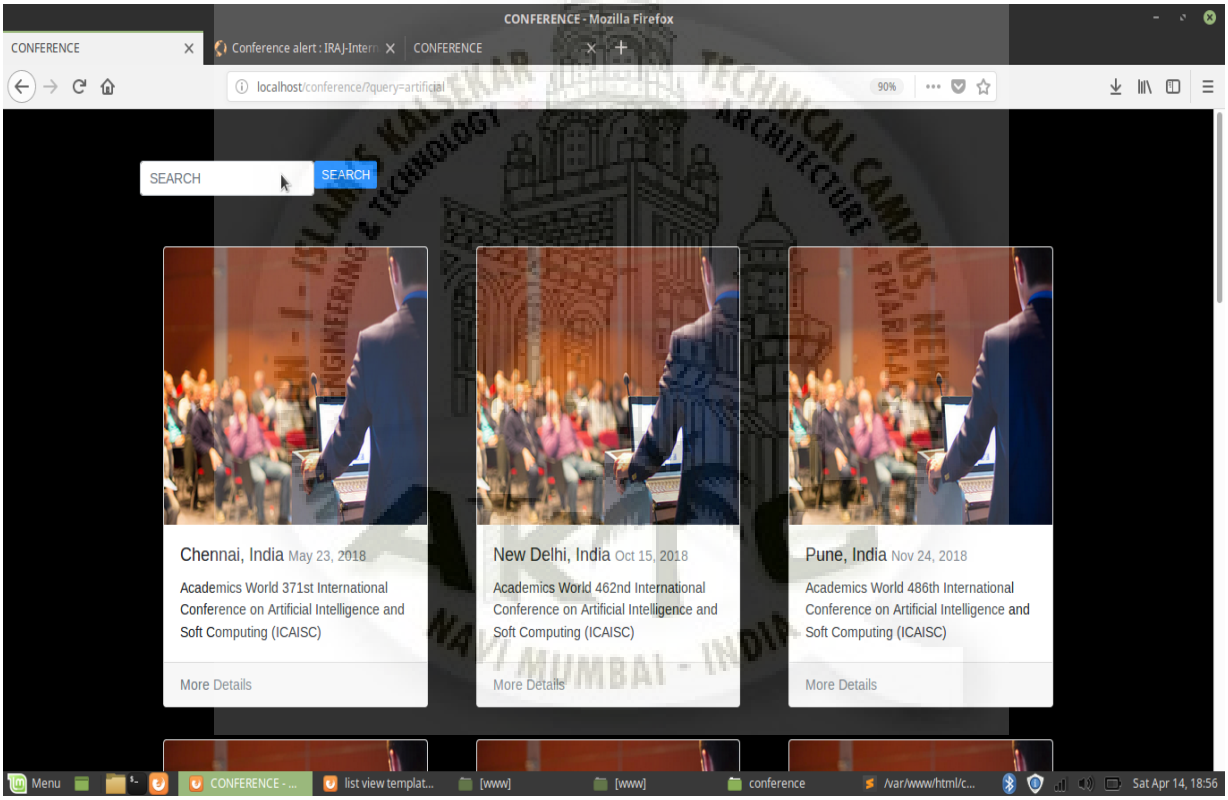

Figure 8.1: Home Page

8.2 Managing Database Online

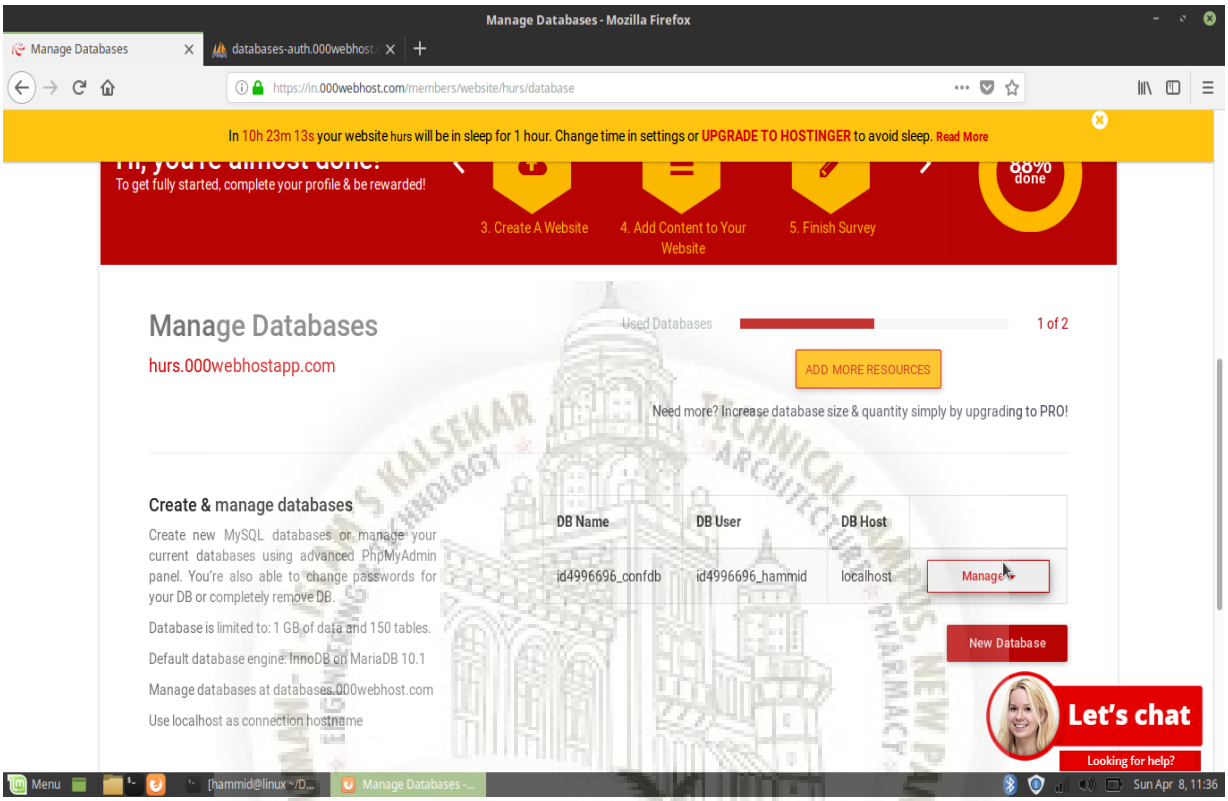

Figure 8.2: Manage Database

Managing the Database online through web-host app.

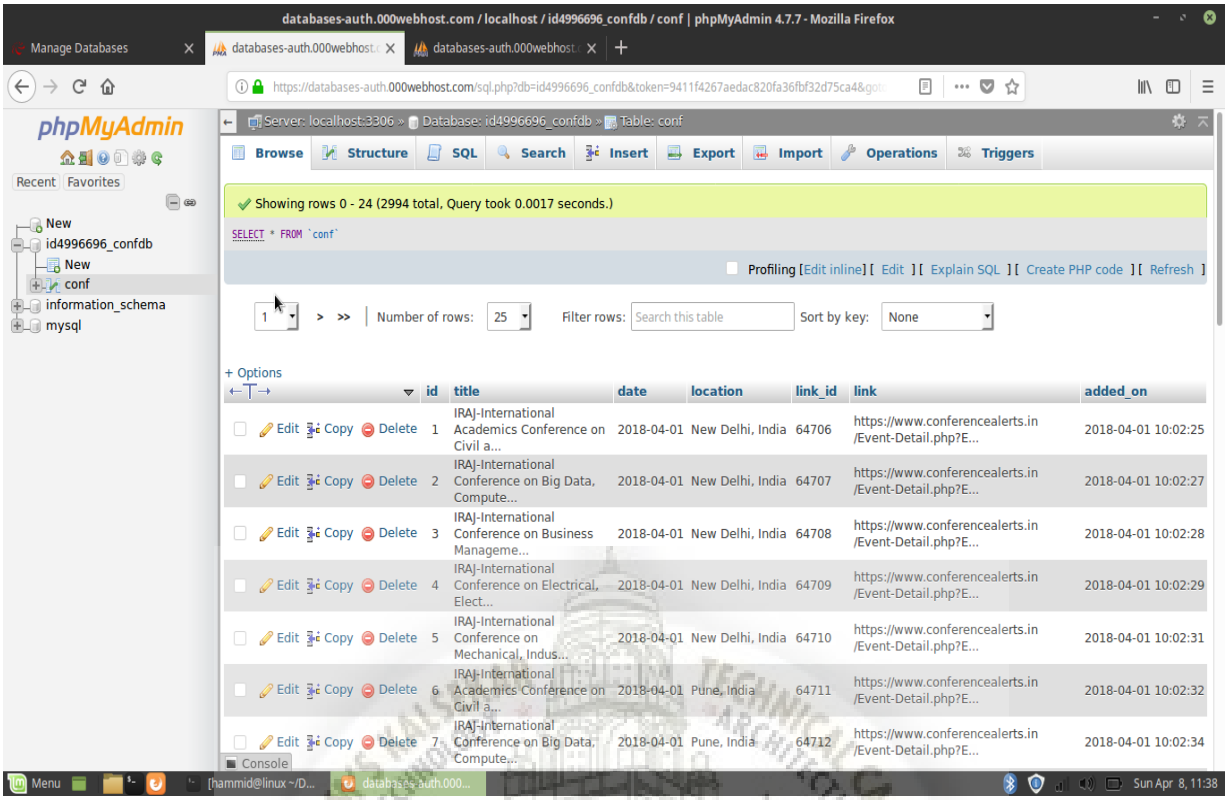

Figure 8.3: Stored Database

Stored the Database on the local-host server

8.3 Display Result Through Keyword

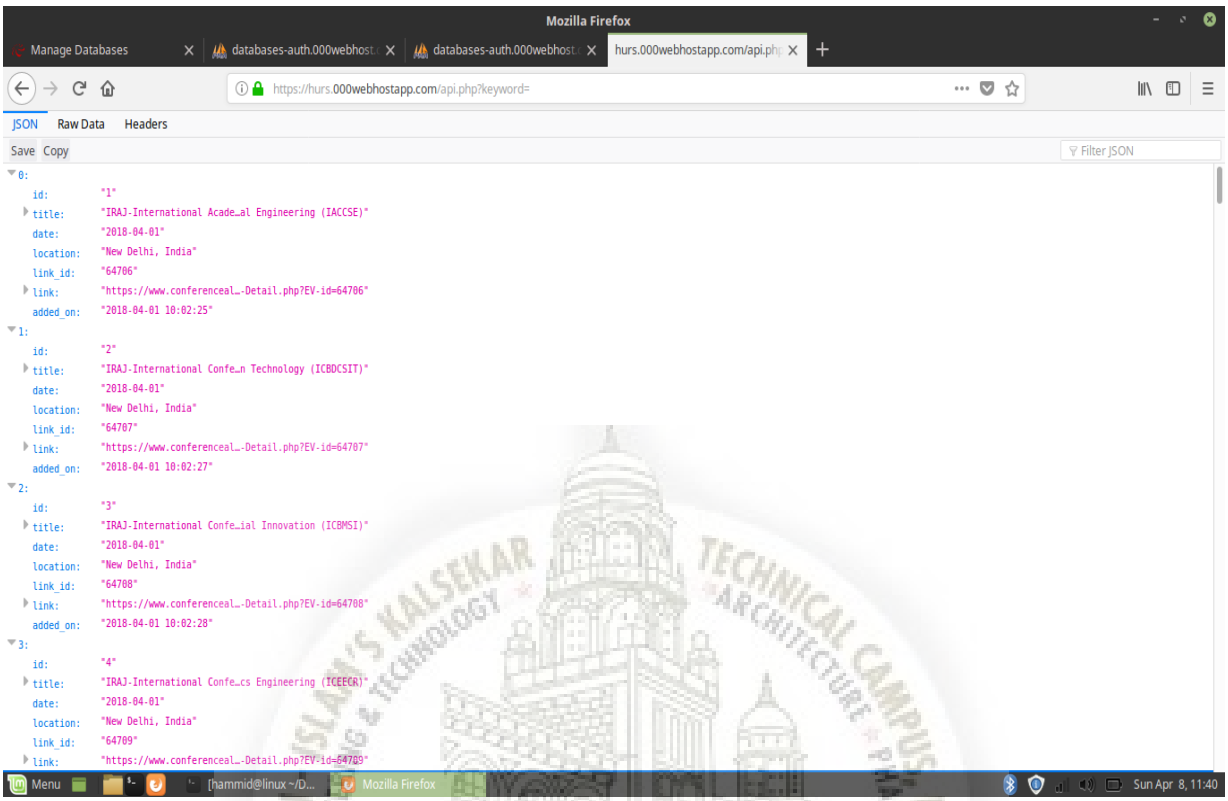

Figure 8.4: Result

Result is displayed online and provide different keyword to display different data related to the different field.

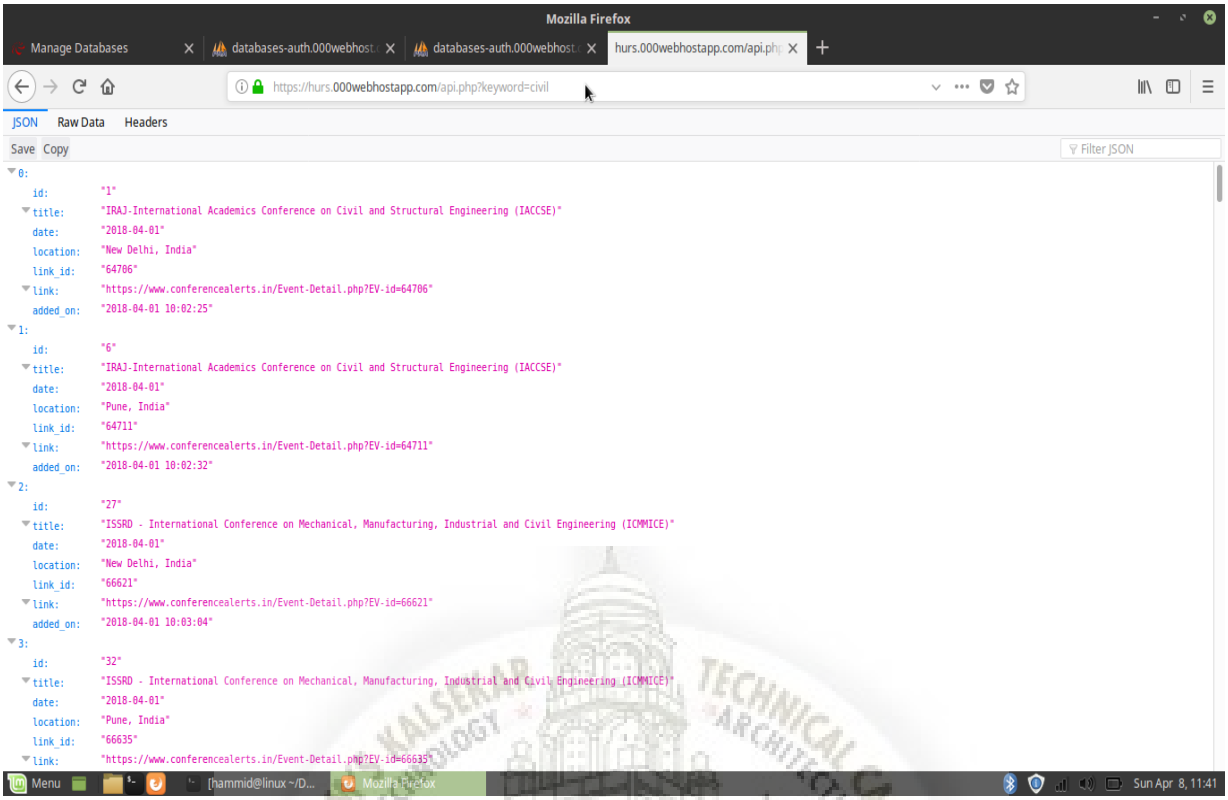

Figure 8.5: Keyword(Artificial Intelligence)

To display the data of the artificial intelligence just provide a keyword to the link

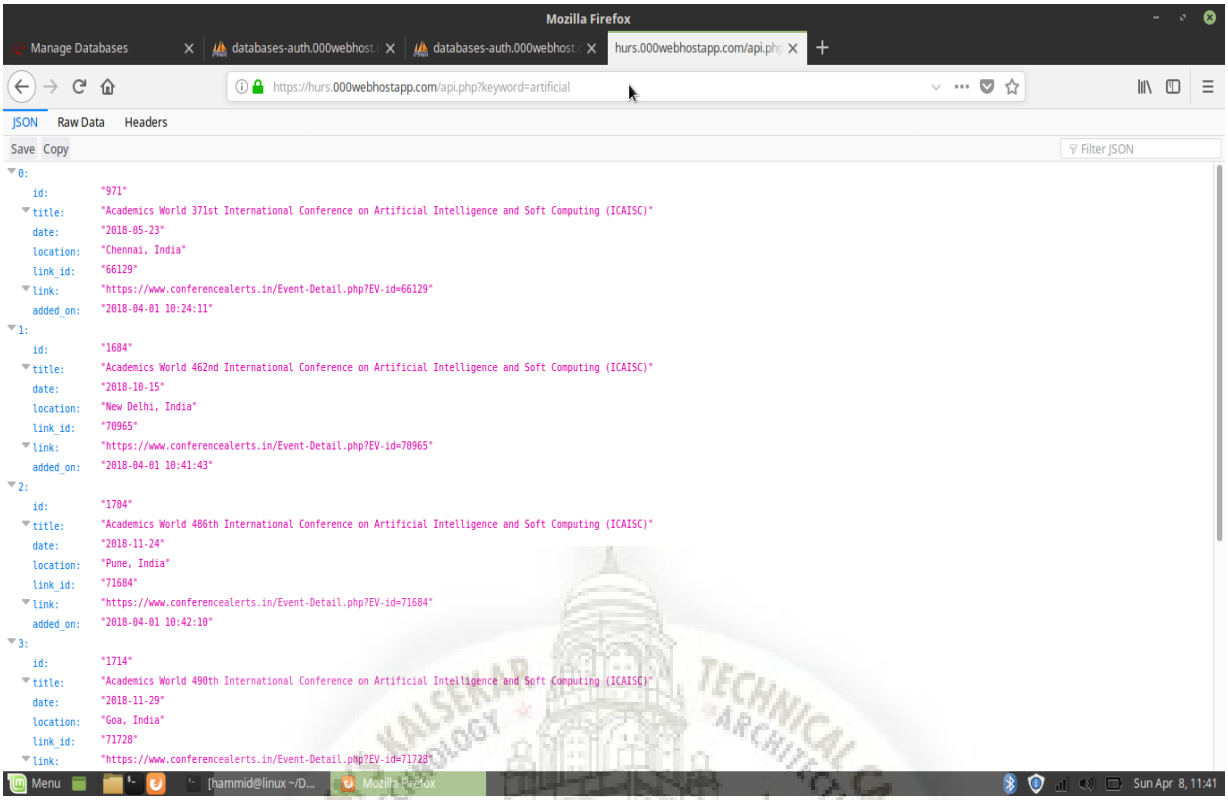

Figure 8.6: Keyword(Civil)

To display the data of the Civil Engineering just provide a keyword to the link

Chapter 9

Conclusion and Future Scope

9.1 Conclusion

Conference Alerts helps in promoting conferences for Academic and scientific studies.The absolute need to attain an international conference at India could be met by using conference alerts to search for the meeting on appropriate areas of interest to present once innovative research articles.

9.2 Future Scope

Once the user register on our web portal so all upcoming events and conference information will be notified to the users via message, e-mail before a week.If users gives response to the notification so the users will get reminder before a day through message or e-mail.We have visited lots of websites which is related to the conference, paper publication some websites notify to he user's via email, message notification but some websites allow only email notification to the users but there is no websites are available which allow the users via desktop notification

References

[1] *Guide to Web Scraping with PHP,*; M. Turland,Muskeeters.me, LLC, 2010

[2] *An Introduction to DataMining,*; D.T. Larose dan C.D. Larose John Wiley Sons Inc, 2014

[3] *Free Canadianclimate data scraping tool,*; Bonifacio, T. E. Barchyn, C. H. Hugenholtz dan S. W. Juebzke 2014.

[4] *https://www.analyticsvidhya.com/beginner-guide-web-scraping-beautiful-souppython/.*;M. Bakaev dan T. Avdeenko,Application in Web Scarping," vol. 4, 2014.

Achievements

1. Publications
   (a) *Web portal on Conference Alert*; Shaikh Uzair Ahmad, Mohammad Hamid, Bind Rahul, Siddiqui Mohd.Sharique, IJISRT, January,2018 (`http://www.ijisrt.com`)
2. Conferences
   (a) *Web portal on Conference Alert*; Shaikh Uzair , ICTCE,February,2018 (Venue :Thakur College of Engineering, Kandivali)
3. Project Competitions
   (a) *Web Portal on Conference Alert*; Shaikh Uzair, Paper Presentation,February,2018(Venue :Thakur College of Engineering,Kandivali)