1. INTRODUCTION

1.1 Introduction To Project

Online Social Media portals play an influential role in information propagation which is considered as an important source for producers in their advertising campaigns as well as for customers in selecting products and services. In addition, written reviews also help service providers to enhance the quality of their products and services. These reviews thus have become an important factor in success of a business while positive reviews can bring benefits for a company, negative reviews can potentially impact credibility and cause economic losses. The fact that anyone with any identity can leave comments as review, provides a tempting opportunity for spammers to write fake reviews designed to mislead users' opinion. These misleading reviews are then multiplied by the sharing function of social media and propagation over the web. The reviews written to change users' perception of how good a product or a service are considered as spam, and are often written in exchange for money.

As shown in, 20% of the reviews in the Yelp website are actually spam reviews. On the other hand, a considerable amount of literature has been published on the techniques used to identify spam and spammers as well as different type of analysis on this topic. These techniques can be classified into different categories; some using linguistic patterns in text which are mostly based on bigram, and unigram, others are based on behavioral patterns that rely on features extracted from patterns in users' behavior which are mostly meta databased.

Despite this great deal of efforts, many aspects have been missed or remained unsolved. One of them is a classifier that can calculate feature weights that show each feature's level of importance in determining spam reviews. The general concept of our proposed framework is to model a given review dataset as a Heterogeneous Information Network (HIN) and to map the problem of spam detection into a HIN classification problem. In particular, we model review dataset as a HIN in which reviews are connected through different node types (such as features and users). A weighting algorithm is then employed to calculate each feature's importance (orweight). These weights are utilized to calculate the final labels for reviews using both unsupervised and supervised approaches. To evaluate the proposed solution, we used two sample review datasets from Yelp and Amazon websites. Based on our observations, defining two views

for features (review-user and behavioral- linguistic), the classified features as review behavioral have more weights and yield better performance on spotting spam reviews in both semi-supervised and unsupervised approaches. In addition, we demonstrate that using different supervisions such as 1%, 2.5% and 5% or using an unsupervised approach, make no noticeable variation on the performance of our approach. We observed that feature weights can be added or removed for labeling and hence timecomplexity can be scaled for a specific level of accuracy.

As the result of this weighting step, we can use fewer features with more weights to obtain better accuracy with less time complexity. In addition, categorizing features in four major categories (review-behavioral, user-behavioral, reviewlinguistic, user-linguistic), helps us to understand how much each category of features is contributed to spam detection. In summary, our main contributions are as follows:

(i) We propose NetSpam framework that is a novel networkbased approach which models review networks as heterogeneous information networks.

(ii) A new weighting method for spam features is proposed to determine the relative importance of each feature and shows how effective each of features are in identifying spams from normal reviews. Previous works also aimed to address the importance of features mainly in term of obtained accuracy, but not as a build-in function in their framework. As we explain in our unsupervised approach, NetSpam is able to find features importance even without ground truth, and only by relying on metapath definition and based on values calculated for each review.

(iii) NetSpam improves the accuracy compared to the stateof- the art in terms of time complexity, which highly depends to the number of features used to identify a spam review; hence, using features with more weights will resulted in detecting fake reviews easier with less time complexity.

1.2 Existing System:

The results show that NetSpam outperforms the existing methods and among four categories of features; including review-behavioral, use behavioral, review linguistic, user linguistic, the first type of features performs better than the other categories.

Despite this great deal of efforts, many aspects have been missed or remained unsolved. One of them is a classifier that can calculate feature weights that show each feature's level of importance in determining spam reviews. The general concept of our proposed framework is to model a given review dataset as a Heterogeneous Information Network (HIN) and to map the problem of spam detection into a HIN classification problem. In particular, we model review dataset as a HIN in which reviews are connected through different node types. The general concept of our proposed framework is to model a given review dataset as a Heterogeneous Information Network and to map the problem of spam detection into a HIN classification problem. In particular, we model review dataset as in which reviews are connected through different node types. A weighting algorithm is then employed to calculate each feature's importance. These weights are utilized to calculate the final labels for reviews using both unsupervised and supervised approaches.

DISADVANTAGE:

- ➢ This utilizes spam features for modeling review datasets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks.
- ➢ Time Complexity.

1.3 Proposed System

NetSpam is able to find features importance even without ground truth, and only by relying on metapath definition and based on values calculated for each review. NetSpam improves the accuracy compared to the stateof- the art in terms of time complexity, which highly depends to the number of features used to identify a spam review; hence, using features with more weights will resulted in detecting fake reviews easier with less time complexity.

A new **Content Based Algorithm** for spam features is proposed to determine the relative importance of each feature and shows how effective each of features are in identifying spams from normal reviews.

ADVANTAGE:

To identify spam and spammers as well as different type of analysis on this topic. Written reviews also help service providers to enhance the quality of their products and services.

2. REQUIREMENT ANALYSIS

Preliminary Investigation:

The first and foremost strategy for development of a project starts from the thought of designing a mail enabled platform for a small firm in which it is easy and convenient of sending and receiving messages, there is a search engine ,address book and also including some entertaining games. When it is approved by the organization and our project guide the first activity, ie. preliminary investigation begins.

The activity has three parts:

➢ Quest Clarification

➢ Feasibility Study

➢ Request Approval

Request Clarification:

After the approval of the request to the organization and project guide, with an investigation being considered, the project request must be examined to determine precisely what the system requires.

Here our project is basically meant for users within the company whose systems can be interconnected by the Local Area Network (LAN). In today's busy schedule man need everything should be provided in a readymade manner. So taking into consideration of the vastly use of the net in day to day life, the corresponding development of the portal came into existence.

2.1 Feasibility Report:

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resource and time. The different feasibility that have to be analyzed are

- ➢ Operational Feasibility
- ➢ Economic Feasibility
- ➢ Technical Feasibility

Operational Feasibility:

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

Economic Feasibility:

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at anytime. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economically feasible.

Technical Feasibility:

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform Independent environment. Java Server Pages, JavaScript, HTML, SQL server and Web Logic

Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility.

Request Approval:

Not all request projects are desirable or feasible. Some organization receives so many project requests from client users that only few of them are pursued. However, those projects that are both feasible and desirable should be put into schedule. After a project request is approved, it cost, priority, completion time and personnel requirement is estimated and used to determine where to add it to any project list. Truly speaking, the approval of those above factors, development works can be launched.

SYSTEM DESIGN AND DEVELOPMENT

Input Design:

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer- based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant.

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2.Select methods for presenting information.

3.Create document, report, or other formats that contain information produced by the system. The output form of an information system should accomplish one or more of the following objectives.

- ➢ Convey information about past activities, current status or projections of the
- ➢ Future.
- ➢ Signal important events, opportunities, problems, or warnings.
- ➢ Trigger an action.
- ➢ Confirm an action.

2.2 SYSTEM REQUIREMENTS

2.2.1 SOFTWARE REQUIREMENTS:

- ➢Operating system : Windows 7,10,11
- ➢Coding Language : JAVA/J2EE
- ➢Tool : Net beans 7.2.1
- ➢ Database : MYSQL

3. LITERATURE SURVEY

3.1 Payword And Micromint: Two Simple MicroPayment Schemes

The Basic Peppercoin method can be implemented in a variety of ways, to maximize ease of use for the customer in a given situation. While the basic peppercoin method requires that each consumer have digital signature capability, one can easily eliminate this requirement by having a party trusted by the consumer sign payments for him as a proxy, this might be a natural approach in a web services enivornment.

The pepper coin method can also be implemented so that it feels to the consumer as a natural extension of his existing credit-card processing procedure, further increasing consumer acceptance and ease of use.

3.2 SECURE POS & KIOSK

Limited interfaces and location within local networks, supporting kiosks and point of sale (POS) terminals can be challenging. Often they are located on networks that are not connected to the internet, making direct access impossible for most remote support tools. And even when an employee is present at the terminal, access restrictions and/or lack of technical knowledge makes communicating the solution to a problem difficult. To add complications, hackers are ramping up their efforts to steal payment card data by gaining access to POS systems and kiosks.

3.3 Reliable OSPM Schema For Secure Transaction Transaction Using Mobile Agent In Micro Payment System

The paper introduces a novel offline payment system in mobile commerce using the case study of micro-payments. The present paper is an extension version of our prior study addressing on implication of secure micropayment system deploying process oriented structural design in

mobile network. The previous system has broad utilization of SPKI and hash chaining to furnish reliable and secure offline transaction in mobile commerce.

However, the current work has attempted to provide much more light weight secure offline payment system in micro-payments by designing a new schema termed as Offline SecurePayment in Mobile Commerce (OSPM). The empirical operation are carried out on three types of transaction process considering maximum scenario of real time offline cases. Therefore, the current idea introduces two new parameters i.e. mobile agent and mobile token that can ensure better security and comparatively less network overhead.

3.4 Light Weight And Secure:PUF Key Storage Using Limits Of Machine Learning

A lightweight and secure key storage scheme using silicon Physical Unclonable Functions (PUFs) is described. To derive stable PUF bits from chip manufacturing variations, a lightweight error correction code (ECC) encoder / decoder is used. With a register count of 69, this codec core does not use any traditional error correction techniques and is 75% smaller than a previous provably secure implementation, and yet achieves robust environmental performance in 65nm FPGA and 0.13µ ASIC implementations. The security of the syndrome bits uses a new security argument that relies on what cannot be learned from a machine learning perspective. The number of Leaked Bits is determined for each Syndrome Word, reducible using Syndrome Distribution .

3.5 Building Robust M-Commerce Payment System On Offline Wireless Network

Mobile commerce is one of the upcoming research area with focus on mobile payment systems. Unfortunately, the current payment systems is directly dependent on fixed infrastructure of network (cellular network), which fails to facilitate optimal level of security for the payment system. The proposed system highlights a novel approach for building a secure, scalable, and flexible e-payment systems in the distributed scenario of wireless adhoc network in offline mode of communication for enhanced security on transaction and payment process. The proposed system uses Simple Public Key Infrastructure for providing the security in payment processes. The performance analysis of the proposed model shows that the system is highly robust and

secure ensuring anonymity, privacy, non-repudiation offline payment system over wireless adhoc network.

4. SYSTEM DESIGN

4.1 Modules:

Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as adding Categories, Adding Products for that Categories, Viewing and authorizing users, View Spam accounts details,Viewing friend request & response, All recommended posts, All posts with all Reviews, All Positive and Negative Reviews , Removing Products, Viewing All Purchased Products, viewing Positive and Negative Reviews Chart on products.

Adding Categories

In this module, the admin adds the category details such as category name. These details will be stored into the database.

Adding Products

In this module, the admin adds Product posts for categories which include details such as, product image, product name, cost, description and uses of that product. These details will be stored into the database. These details will be further searched and accessed by the users in order to recommend to their friends and to buy products.

Authorize Users

In user's module, the admin can view the list of users who all registered. In this, the admin can view the users' details such as, user name, email, address, phone number and authorize the users.

Request & Response

In this module, the admin can view all the friend requests and responses. Here all the requests and responses will be displayed with their tags such as Id, requested user image, requested user name, user name request to, status and time & date. If the user accepts the request then the status will be changed to accepted or else the status will remains as waiting.

All Recommended Posts

In this module, the admin can view all the recommended products. If any recommendations happened for particular products, those details will be shown along with products. Details include product name, recommended user name, user recommended to name and the date.

View Positive /Negative Comments

In this, the admin can view all posts with their Positive and Negative Comments posted by users based on their opinions.

Positive: If the user comment contains at least one of the word which is listed in positive words, then that comment will be treated as positive comment.

Negative: If the user comment contains at least one of the word which is listed in negative words, then that comment will be treated as negative comment.

All Comments on Products

In this module, the comments of all posts will be displayed. Comments includes Positive, Negative, Non-Positive and Non-Negative. It includes details such as, commented user name, comment and date.

All Purchased Products

In this module, the products which are purchased by users will be displayed. It includes details such as, purchased user name, purchased products, price of the products and the date of purchase.

Positive Comments Chart

In this module, the number of positive Reviews got by the particular product will be displayed in a chart.

Negative Comments Chart

In this module, the number of negative Reviews got by the particular product will be displayed in a chart.

Deleting/Removing Products

In this module, the products which have got the negative comments from more than five users will be listed and removed by the admin.

User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like viewing their profile Account details like Spam or Normal , search users and send friend request, viewing friend requests, searching posts and recommend to friends and viewing all product recommendations sent to him by his friends, commenting on posts, purchasing products and viewing their product search history.

Search Users

The user can search the users based on names and the server will give response to the user like User name, user image, E mail id, phone number and date of birth. If you wish to send friend request to particular user then click on “request” button, then request will be send to that particular user.

Searching Products and Recommend to Friends

In this, the user searches for products based on the products description. The user can recommend searched products to his friends, comment on post and he can add the products to cart to buy those added products later by using their created account.

View Friend Requests

In this module, the user can view the friend requests which are sent by other users. Which includes request sent user details with their tags such as user name, user image, date of birth,EMail ID, phone number and Address and user can accept the request by clicking on the “waiting” link.

View Product Recommends

In this module, the user can view all the products which are recommended by his friends. This includes recommended user name and his image, recommended products details.

View Product Search History

In this module, the user can view all the searched products names and categories, the keywords which he used to search the products. This includes details such as, searched product,used keyword and date of search.

View Bank Account Details In

this module, the user can create his bank account by providing details such as, account number, branch, address, email id. Later he can add money to his account and can view his account details.

4.2 DIAGRAMS

SYSTEM ARCHITECTURE:

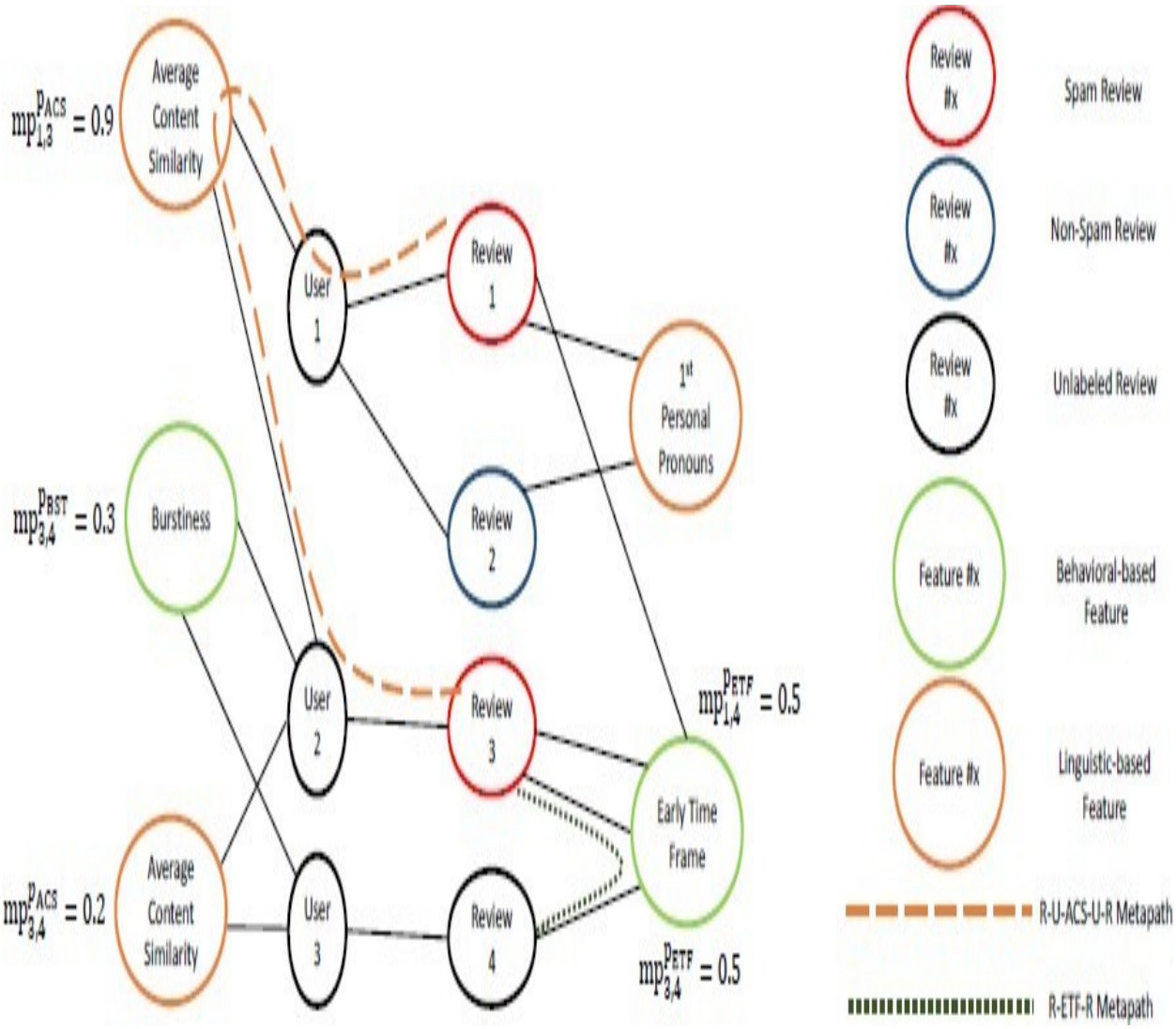


fig:4.2.1 System Architecture

DATA FLOW DIAGRAM:

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.

2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.

4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

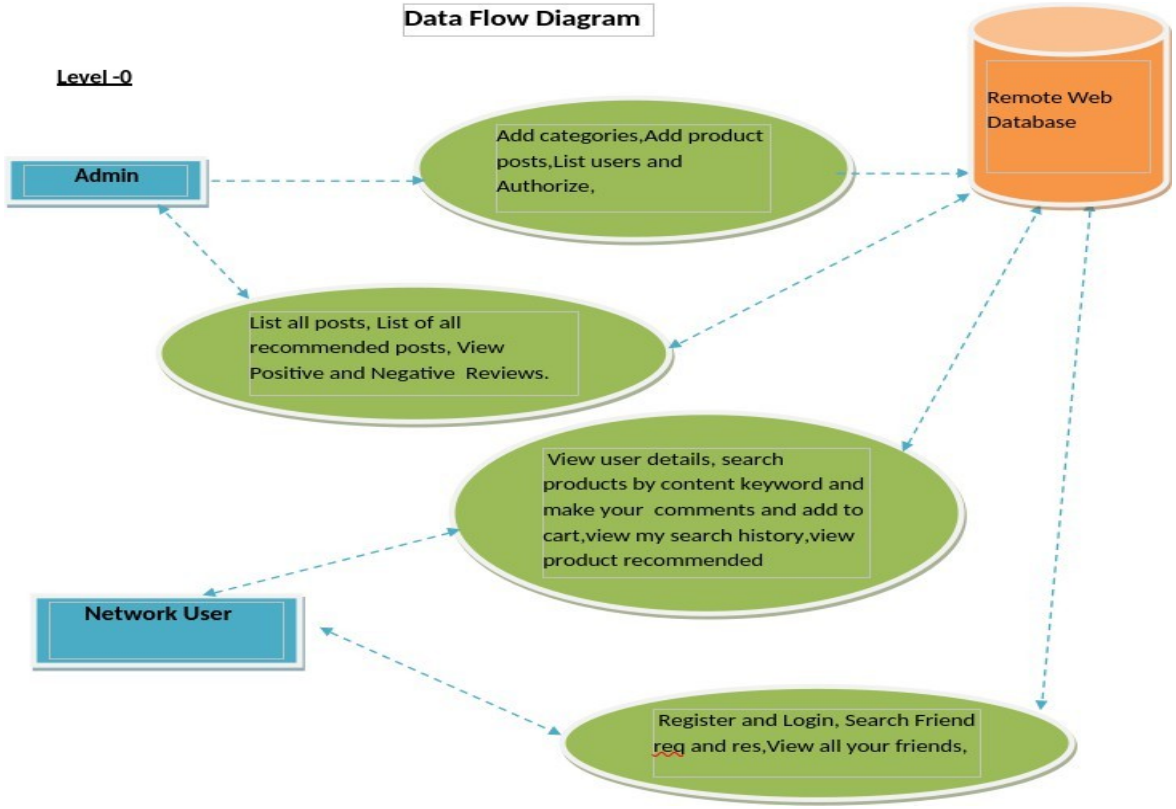


fig 4.2.2 Data Flow Diagram

USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

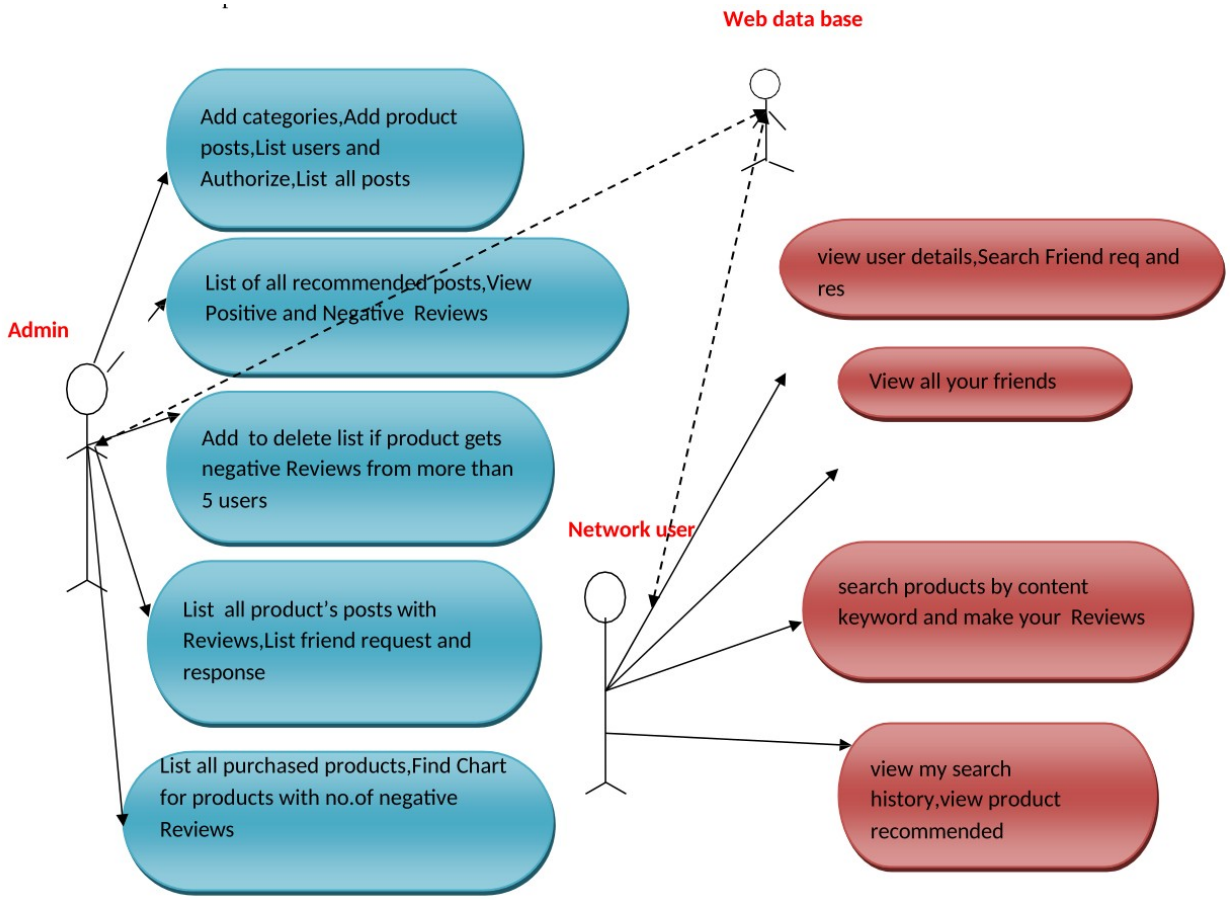


fig 4.2.3 Use Case Diagram

CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

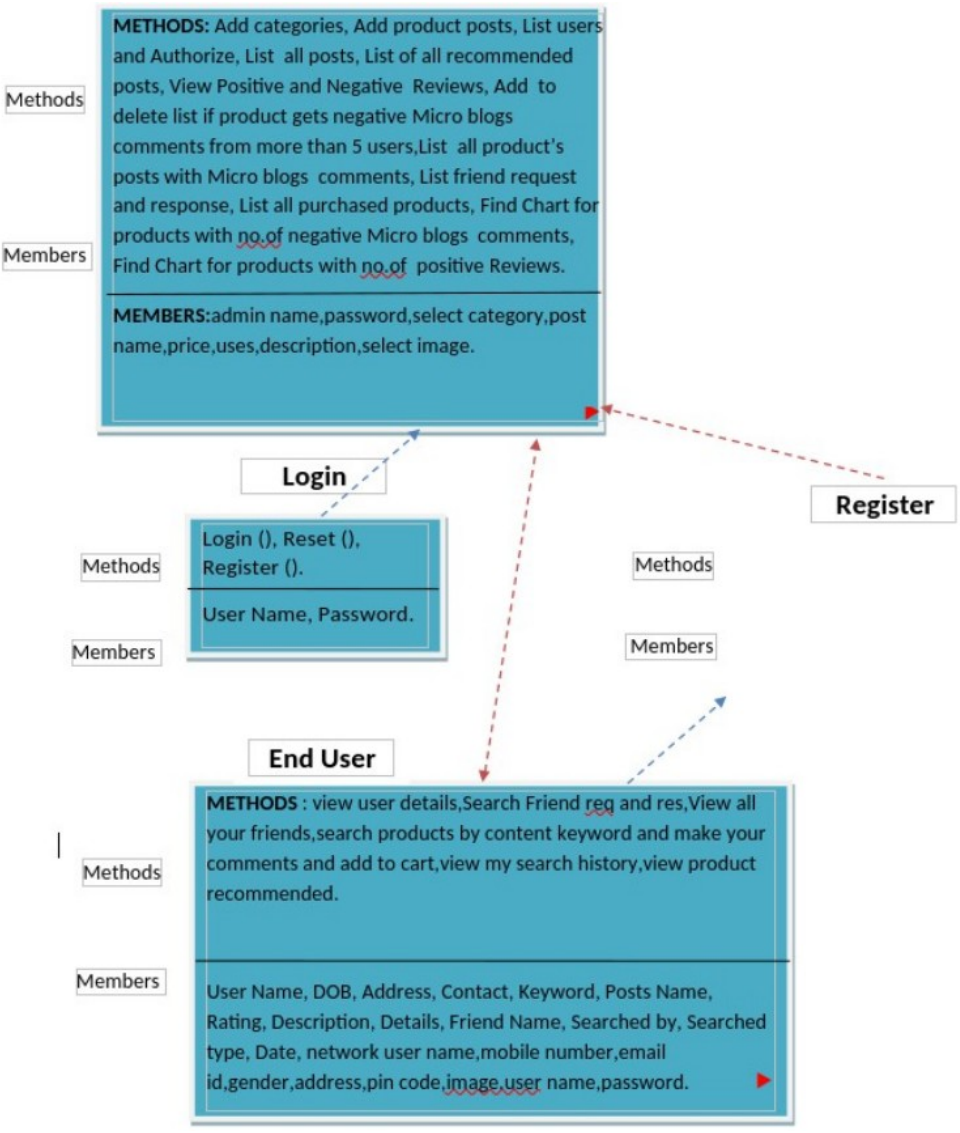


Fig 4.2.4 Class Diagram

SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

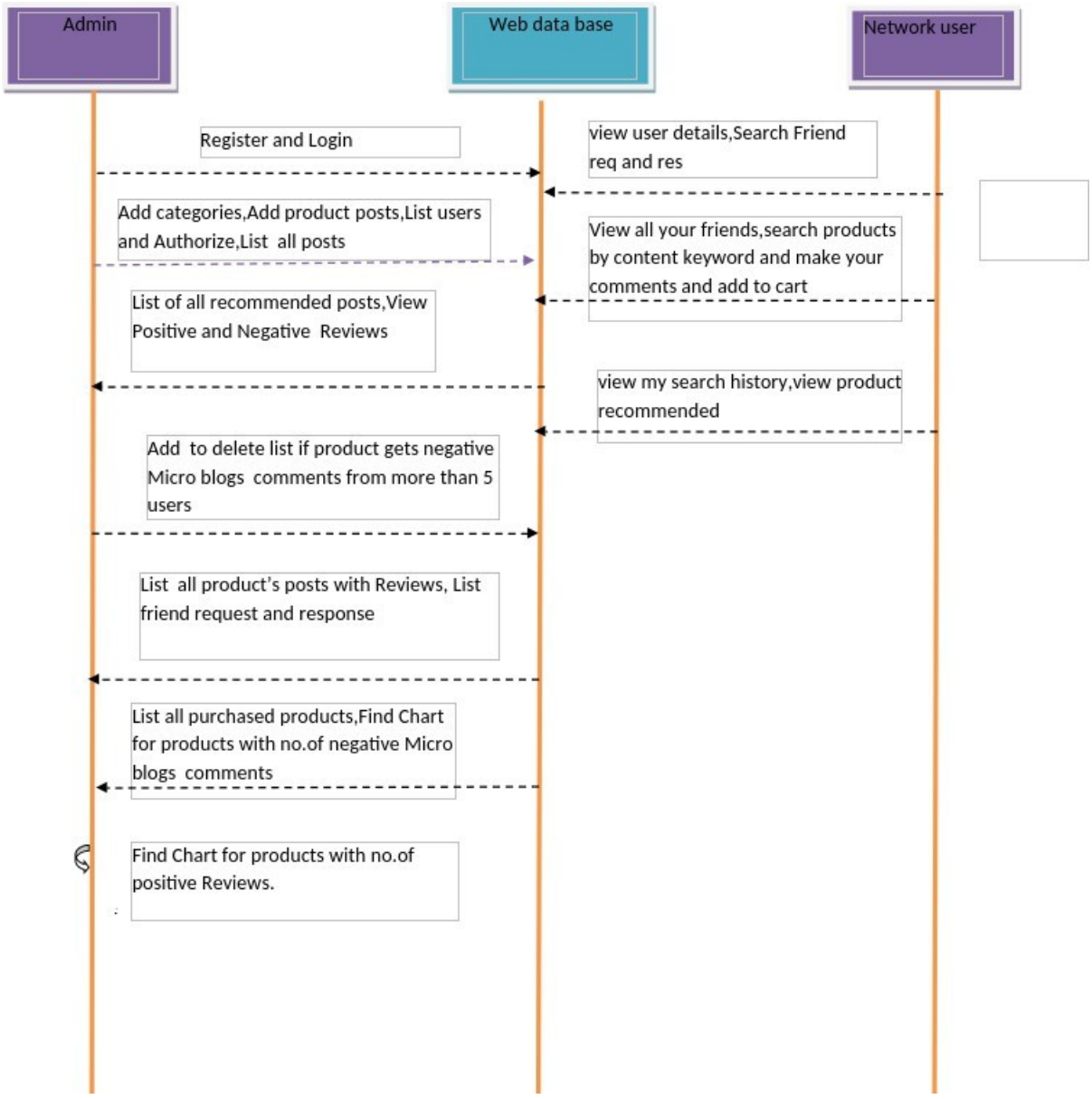


Fig 4.2.5 Sequence Diagram

4.3 DATA DICTIONARY

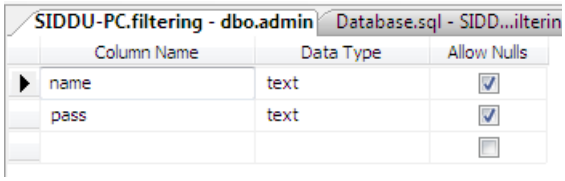


fig 4.3.1 Admin

SIDDU-PC.filtering - dbo.account | SIDDU-PC.filtering - dl

Column Name	Data Type	Allow Nulls
account_no	varchar(15)	☑
username	varchar(50)	☑
address	varchar(45)	☑
email	varchar(45)	☑
mobile	varchar(11)	☑
branch	varchar(45)	☑
amount	text	☑
		☐

fig 4.3.2 Account

SIDDU-PC.filtering - dbo.allcomments | SIDDU-PC.filtering

Column Name	Data Type	Allow Nulls
id	int	☐
p_name	text	☑
categorie	text	☑
user1	text	☑
comment	text	☑
date	text	☑
		☐

fig 4.3.3 All Comments

SIDDU-PC.filtering - dbo.categories | SIDDU-PC.filtering - dl

Column Name	Data Type	Allow Nulls
id	int	☐
category	varchar(50)	☑
		☐

fig 4.3.4 Categories

SIDDU-PC.filtering - dbo.cart | SIDDU-PC.filtering - dbo.ca

Column Name	Data Type	Allow Nulls
id	int	☐
user1	text	☑
p_name	text	☑
categorie	text	☑
date	text	☑
price	int	☐
		☐

fig 4.3.5 Add Cart

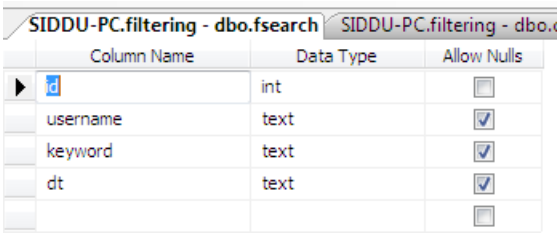


fig 4.3.6 File Search

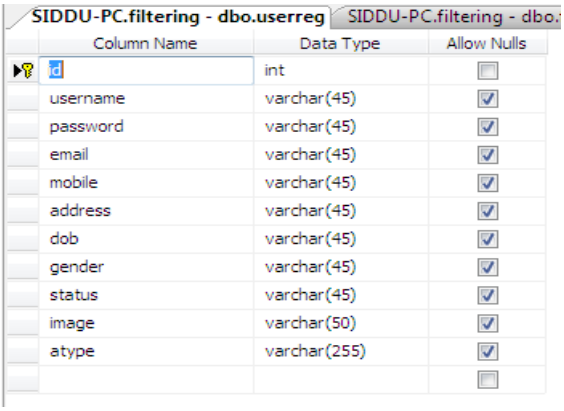


fig 4.3.7 User registration

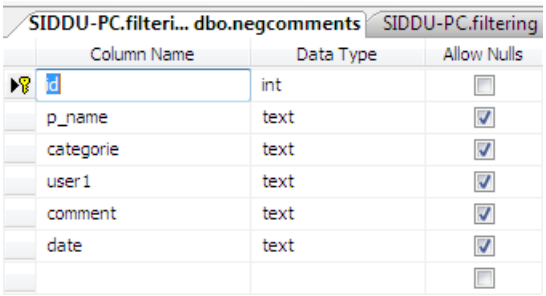


fig 4.3.8 Negative Comments

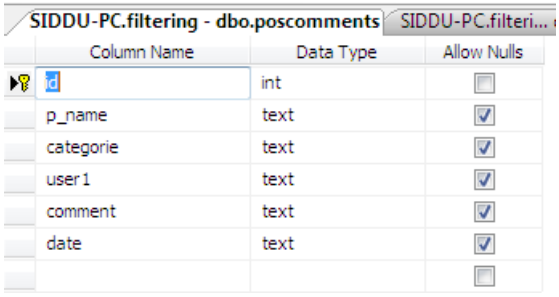


fig 4.3.9 Positive Comments

SIDDU-PC.filtering - dbo.posts | SIDDU-PC.filtering - dbo.p

Column Name	Data Type	Allow Nulls
id	int	☐
p_name	text	☑
categorie	text	☑
p_price	text	☑
p_uses	text	☑
p_des	text	☑
p_image	varchar(50)	☑
date	text	☑
positive	int	☑
negative	int	☑
		☐

fig 4.4.0 Post Search

SIDDU-PC.filtering - dbo.purchase | SIDDU-PC.filtering - dbc

Column Name	Data Type	Allow Nulls
id	int	☐
account_no	text	☑
user1	text	☑
p_name	text	☑
categorie	text	☑
date	text	☑
price	int	☐
		☐

fig 4..4.1 Purchase

5. SYSTEM IMPLEMENTATION

5.1 TECHNOLOGY USED

Client Server

Over view:

With the varied topic in existence in the fields of computers, Client Server is one, which has generated more heat than light, and also more hype than reality. This technology has acquired a certain critical mass attention with its dedication conferences and magazines. Major computer vendors such as IBM and DEC, have declared that Client Servers is their main future market. A survey of DBMS magazine reveled that 76% of its readers were actively looking at the client server solution. The growth in the client server development tools from $200 million in 1992 to more than $1.2 billion in 1996.Client server implementations are complex but the underlying concept is simple and powerful. A client is an application running with local resources but able to request the database and relate the services from separate remote server. The software mediating this client server interaction is often referred to as MIDDLEWARE.

The typical client either a PC or a Work Station connected through a network to a more powerful PC, Workstation, Midrange or Main Frames server usually capable of handling request from more than one client. However, with some configuration server may also act as client. A server may need to access other server in order to process the original client request.The key client server idea is that client as user is essentially insulated from the physical location and formats of the data needs for their application. With the proper middleware, a client input from or report can transparently access and manipulate both local database on the client machine and remote databases on one or more servers.

What is a Client Server

Two prominent systems in existence are client server and file server systems. It is essential to distinguish between client servers and file server systems. Both provide shared network access to data but the comparison dens there! The file server simply provides a remote disk drive that can be accessed by LAN applications on a file by file basis. The client server offers full relational database services such as SQL-Access, Record modifying, Insert, Delete with full relational integrity backup/restore performance for high volume of transactions, etc. the client server middle ware provides a flexible interface between client and server, who does what, when and to whom.

Why Client Server

Client server has evolved to solve a problem that has been around since the earliest days of computing: how best to distribute your computing, data generation and data storage resources in order to obtain efficient, cost effective departmental an enterprise wide data processing. During mainframe era choices were quite limited. A central machine housed both the CPU and DATA (cards, tapes, drums and later disks). Access to these resources was initially confined to batched runs that produced departmental reports at the appropriate intervals. A strong central information service department ruled the corporation. The role of the rest of the corporation limited to requesting new or more frequent reports and to provide hand written forms from which the central data banks were created and updated. The earliest client server solutions therefore could best be characterized as “SLAVE-MASTER”.

Font end or User Interface Design

The entire user interface is planned to be developed in browser specific environment with a touch of Intranet-Based Architecture for achieving the Distributed Concept.The browser specific components are designed by using the HTML standards, and the dynamism of the designed by concentrating on the constructs of the Java Server Pages.

Communication or Database Connectivity Tier

The Communication architecture is designed by concentrating on the Standards of Servlets and Enterprise Java Beans. The database connectivity is established by using the Java Data Base Connectivity.The standards of three-tire architecture are given major concentration to keep the standards of higher cohesion and limited coupling for effectiveness of the operations. features of the language used In my project, I have chosen Java language for developing the code

About Java

Initially the language was called as “oak” but it was renamed as “Java” in 1995. The primary motivation of this language was the need for a platform-independent (i.e., architecture

neutral) language that could be used to create software to be embedded in various consumer electronic devices.

➢ Java is a programmer's language.

➢ Java is cohesive and consistent.

➢ Except for those constraints imposed by the Internet environment, Java gives the programmer, full control.

Finally, Java is to Internet programming where C was to system programming.

Importance of Java to the Internet

Java has had a profound effect on the Internet. This is because; Java expands the Universe of objects that can move about freely in Cyberspace. In a network, two categories of objects are transmitted between the Server and the Personal computer. They are: Passive information and Dynamic active programs. The Dynamic, Self-executing programs cause serious problems in the areas of Security and probability. But, Java addresses those concerns and by doing so, has opened the door to an exciting new form of program called the Applet. Java can be used to create two types of programs

Applications and Applets:

An application is a program that runs on our Computer under the operating system of that computer. It is more or less like one creating using C or C++. Java's ability to create Applets makes it important. An Applet is an application designed to be transmitted over the Internet and executed by a Java –compatible web browser. An applet is actually a tiny Java program, dynamically downloaded across the network, just like an image. But the difference is, it is an intelligent program, not just a media file. It can react to the user input and dynamically change.

Features of Java

Security:

Every time you that you download a “normal” program, you are risking a viral infection. Prior to Java, most users did not download executable programs frequently, and those who did scanned them for viruses prior to execution. Most users still worried about the possibility of infecting their systems with a virus. In addition, another type of malicious program exists that must be guarded against. This type of program can gather private information, such as credit card numbers, bank account balances, and passwords. Java answers both these concerns by providing a “firewall” between a network application and your computer. When you use a Java-compatible Web browser, you can safely download Java applets without fear of virus infection or malicious intent.

Portability:

For programs to be dynamically downloaded to all the various types of platforms connected to the Internet, some means of generating portable executable code is needed .As you will see, the same mechanism that helps ensure security also helps create portability. Indeed, Java’s solution to these two problems is both elegant and efficient.

The Byte code:

The key that allows the Java to solve the security and portability problems is that the output of Java compiler is Byte code. Byte code is a highly optimized set of instructions designed to be executed by the Java run-time system, which is called the Java Virtual Machine (JVM). That is, in its standard form, the JVM is an interpreter for byte code.

Translating a Java program into byte code helps makes it much easier to run a program in a wide variety of environments. The reason is, once the run-time package exists for a given system, any Java program can run on it. Although Java was designed for interpretation, there is technically nothing about Java that prevents on-the-fly compilation of byte code into native code. Sun has just completed its Just In Time(JIT) compiler for byte code. When the JIT compiler is a part of JVM, it compiles byte code into executable code in real time, on a piece-by-piece, demand basis. It is not possible to compile an entire Java program into executable code all at

once, because Java performs various run-time checks that can be done only at run time. The JIT compiles code, as it is needed, during execution.

Java Virtual Machine (JVM):

Beyond the language, there is the Java virtual machine. The Java virtual machine is an important element of the Java technology. The virtual machine can be embedded within a web browser or an operating system. Once a piece of Java code is loaded onto a machine, it is verified.

As part of the loading process, a class loader is invoked and does byte code verification makes sure that the code that's has been generated by the compiler will not corrupt the machine that it's loaded on. Byte code verification takes place at the end of the compilation process to make sure that is all accurate and correct. So byte code verification is integral to the compiling and executing of Java code.

Picture showing the development process of JAVA Program:

Java programming uses to produce byte codes and executes them. The first box indicates that the Java source code is located in a. Java file that is processed with a Java compiler called javac. The Java compiler produces a file called a. class file, which contains the byte code. The. Class file is then loaded across the network or loaded locally on your machine into the execution environment is the Java virtual machine, which interprets and executes the byte code.

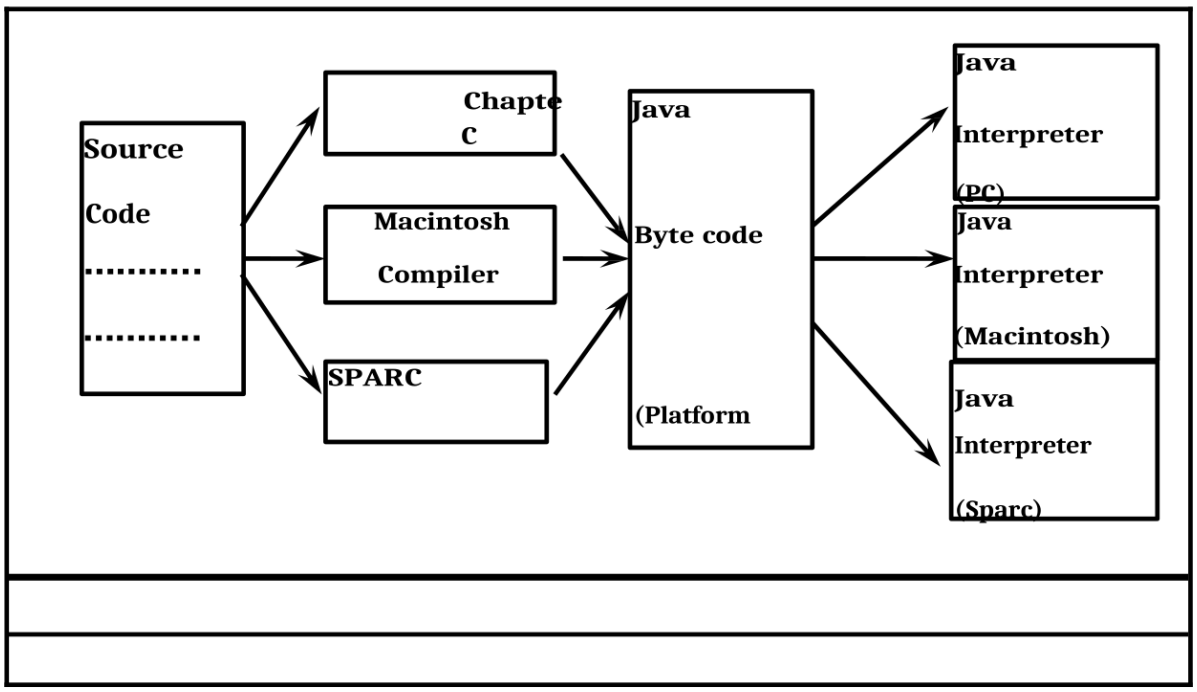
Java Architecture:

Java architecture provides a portable, robust, high performing environment for development. Java provides portability by compiling the byte codes for the Java Virtual Machine, which is then interpreted on each platform by the run-time environment. Java is a dynamic system, able to load code when needed from a machine in the same room or across the planet.

Compilation of code:

When you compile the code, the Java compiler creates machine code (called byte code) for a hypothetical machine called Java Virtual Machine (JVM). The JVM is supposed to execute the byte code. The JVM is created for overcoming the issue of portability. The code is written and compiled for one machine and interpreted on all machines. This machine is called Java Virtual Machine

Compiling and interpreting Java Source Code:



During run-time the Java interpreter tricks the byte code file into thinking that it is running on a Java Virtual Machine. In reality this could be a Intel Pentium Windows 95 or Sun SARC station running Solaris or Apple Macintosh running system and all could receive code from any computer through Internet and run the Applets.

Simple:

Java was designed to be easy for the Professional programmer to learn and to use effectively. If you are an experienced C++ programmer, learning Java will be even easier. Because Java inherits the C/C++ syntax and many of the object oriented features of C++. Most

of the confusing concepts from C+ + are either left out of Java or implemented in a cleaner, more approachable manner. In Java there are a small number of clearly defined ways to accomplish a given task.

Object-Oriented:

Java was not designed to be source-code compatible with any other language. This allowed the Java team the freedom to design with a blank slate. One outcome of this was a clean usable, pragmatic approach to objects. The object model in Java is simple and easy to extend, while simple types, such as integers, are kept as high-performance non-objects.

Robust:

The multi-platform environment of the Web places extraordinary demands on a program, because the program must execute reliably in a variety of systems. The ability to create robust programs was given a high priority in the design of Java. Java is strictly typed language; it checks your code at compile time and run time. Java virtually eliminates the problems of memory management and deallocation, which is completely automatic. In a well-written Java program, all run time errors can –and should –be managed by your program.

JAVASCRIPT

JavaScript is a script-based programming language that was developed by Netscape Communication Corporation. JavaScript was originally called Live Script and renamed as JavaScript to indicate its relationship with Java. JavaScript supports the development of both client and server components of Web-based applications. On the client side, it can be used to write programs that are executed by a Web browser within the context of a Web page.

Even though JavaScript supports both client and server Web programming, we prefer JavaScript at Client side programming since most of the browsers supports it. JavaScript is almost as easy to learn as HTML, and JavaScript statements can be included in HTML documents by enclosing the statements between a pair of scripting tags

<SCRIPTS>..</SCRIPT>.

<SCRIPT LANGUAGE = “JavaScript”>

JavaScript statements

</SCRIPT>

Here are a few things we can do with JavaScript :

➢Validate the contents of a form and make calculations.

➢Add scrolling or changing messages to the Browser’s status line.

➢Animate images or rotate images that change when we move the mouse over them.

➢Detect the browser in use and display different content for different browsers.

➢We can do much more with JavaScript, including creating entire application.

JavaScript Vs Java

JavaScript and Java are entirely different languages. A few of the most glaring differences are:

➢Java applets are generally displayed in a box within the web document; JavaScript can affect any part of the Web document itself.

➢While JavaScript is best suited to simple applications and adding interactive features to Web pages; Java can be used for incredibly complex applications.

There are many other differences but the important thing to remember is that JavaScript and Java are separate languages. They are both useful for different things; in fact they can be used together to combine their advantages.

ADVANTAGES:

➢JavaScript can be used for Sever-side and Client-side scripting.

➢It is more flexible than VBScript.

➢JavaScript is the default scripting languages at Client-side since all the browsers supports it.

Hyper Text Markup Language:

Hypertext Markup Language (HTML), the languages of the World Wide Web (WWW), allows users to produces Web pages that include text, graphics and pointer to other Web pages (Hyperlinks). HTML is not a programming language but it is an application of ISO Standard 8879, SGML (Standard Generalized Markup Language), but specialized to hypertext and adapted to the Web. The idea behind Hypertext is that instead of reading text in rigid linear structure, we can easily jump from one point to another point. We can navigate through the information based on our interest and preference. A markup language is simply a series of elements, each delimited with special characters that define how text or other items enclosed within the elements should be displayed. Hyperlinks are underlined or emphasized works that load to other documents or some portions of the same document. It can be used to display any type of document on the host computer, which can be geographically at a different location. It is a versatile language and can be used on any platform or desktop.

HTML provides tags (special codes) to make the document look attractive. HTML tags are not case-sensitive. Using graphics, fonts, different sizes, color, etc., can enhance the presentation of the document. Anything that is not a tag is part of the document itself.

Basic HTML Tags :

<!-- --!>	Specifies comments
<A>…….</A>	Creates hypertext links
<B>…….</B>	Formats text as bold
<BIG>…….</BIG>	Formats text in large font.
<BODY>…</BODY>	Contains all tags and text in the HTML document
<CENTER>...</CENTER>	Creates text
<DD>…</D>	Definition of a term
<DL>...</DL>	Creates definition list
<FONT>…</FONT>	Formats text with a particular font

<FORM>...</FORM>	Encloses a fill-out form
<FRAME>...</FRAME>	Defines a particular frame in a set of frames
<H#>…</H#>	Creates headings of different levels
<HEAD>...</HEAD>	Contains tags that specify information about a document
<HR>...</HR>	Creates a horizontal rule
<HTML>…</HTML>	Contains all other HTML tags
<META>...</META>	Provides meta-information about a document
<SCRIPT>…</SCRIPT>	Contains client-side or server-side script
<TABLE>…</TABLE>	Creates a table
<TD>…</TD>	Indicates table data in a table
<TR>…</TR>	Designates a table row
<TH>…</TH>	Creates a heading in a table

ADVANTAGES:

A HTML document is small and hence easy to send over the net. It is small because it does not include formatted information.

➢HTML is platform independent.

➢HTML tags are not case-sensitive.

Java Database Connectivity

What Is JDBC?

JDBC is a Java API for executing SQL statements. (As a point of interest, JDBC is a trademarked name and is not an acronym; nevertheless, JDBC is often thought of as standing for Java Database Connectivity. It consists of a set of classes and interfaces written in the Java

programming language. JDBC provides a standard API for tool/database developers and makes it possible to write database applications using a pure Java API. Using JDBC, it is easy to send SQL statements to virtually any relational database. One can write a single program using the JDBC API, and the program will be able to send SQL statements to the appropriate database. The combinations of Java and JDBC lets a programmer write it once and run it anywhere.

What Does JDBC Do?

Simply put, JDBC makes it possible to do three things:

- ➢Establish a connection with a database
- ➢Send SQL statements
- ➢Process the results.

JDBC versus ODBC and other APIs:

At this point, Microsoft's ODBC (Open Database Connectivity) API is that probably the most widely used programming interface for accessing relational databases. It offers the ability to connect to almost all databases on almost all platforms.

So why not just use ODBC from Java? The answer is that you can use ODBC from Java, but this is best done with the help of JDBC in the form of the JDBC-ODBC Bridge, which we will cover shortly. The question now becomes "Why do you need JDBC?" There are several answers to this question:

1. ODBC is not appropriate for direct use from Java because it uses a C interface. Calls from Java to native C code have a number of drawbacks in the security, implementation, robustness, and automatic portability of applications.

2. A literal translation of the ODBC C API into a Java API would not be desirable. For example, Java has no pointers, and ODBC makes copious use of them, including the notoriously error-prone generic pointer "void *". You can think of JDBC as ODBC translated into an object-oriented interface that is natural for Java programmers.

3. ODBC is hard to learn. It mixes simple and advanced features together, and it has complex options even for simple queries. JDBC, on the other hand, was designed to keep simple things simple while allowing more advanced capabilities where required.

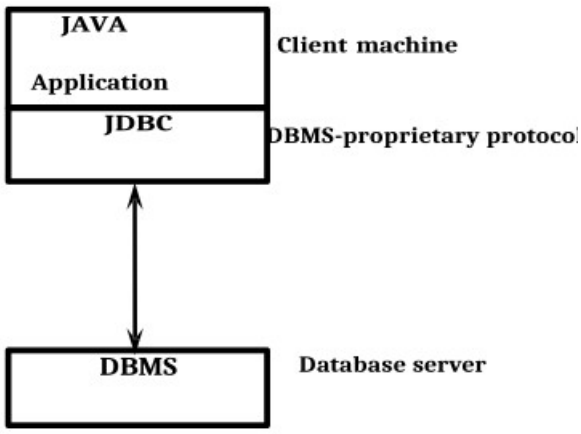
4. A Java API like JDBC is needed in order to enable a "pure Java" solution. When ODBC is used, the ODBC driver manager and drivers must be manually installed on every client machine. When the JDBC driver is written completely in Java, however, JDBC code is automatically installable, portable, and secure on all Java platforms from network computers to mainframes.

Two-tier and Three-tier Models:

The JDBC API supports both two-tier and three-tier models for database access.

In the two-tier model, a Java applet or application talks directly to the database. This requires a JDBC driver that can communicate with the particular database management system being accessed. A user's SQL statements are delivered to the database, and the results of those statements are sent back to the user.

The database may be located on another machine to which the user is connected via a network. This is referred to as a client/server configuration, with the user's machine as the client, and the machine housing the database as the server.In the three-tier model, commands are sent to a "middle tier" of services, which then send SQL statements to the database. The database processes the SQL statements and sends the results back



to the middle tier, which then sends them to the user. MIS directors find the three-tier model very attractive because the middle tier makes it possible to maintain control over access and the kinds of updates that can be made to corporate data.

Another advantage is that when there is a middle tier, the user can employ an easy-to-use higher-level API which is translated by the middle tier into the appropriate low- level calls. Finally, in many cases the three-tier architecture can provide performance advantages.

JDBC Driver Types:

The JDBC drivers that we are aware of at this time fit into one of four categories:

- ➢JDBC-ODBC bridge plus ODBC driver
- ➢Native-API partly-Java driver
- ➢JDBC-Net pure Java driver
- ➢Native-protocol pure Java driver

JDBC-ODBC Bridge:

If possible, use a Pure Java JDBC driver instead of the Bridge and an ODBC driver. This completely eliminates the client configuration required by ODBC. It also eliminates the potential that the Java VM could be corrupted by an error in the native code brought in by the Bridge (that is, the Bridge native library, the ODBC driver manager library, the ODBC driver library, and the database client library).

What Is the JDBC- ODBC Bridge?

The JDBC-ODBC Bridge is a JDBC driver, which implements JDBC operations by translating them into ODBC operations. To ODBC it appears as a normal application program. The Bridge implements JDBC for any database for which an ODBC driver is available. The

Bridge is implemented as the sun.jdbc.odbc Java package and contains a native library used to access ODBC. The Bridge is a joint development of Intersolve and JavaSoft.

Java Server Pages (JSP):

Java sever Pages is a simple, yet powerful technology for creating and maintaining dynamic- content web pages. Based on the Java programming language, Java Server Pages offers proven portability, open standards, and a mature re-usable component model .The Java Server Pages architecture enables the separation of content generation from content presentation.

Features of JSP

Portability:

Java Server Pages files can be run on any web server or web-enabled application server that provides support for them. Dubbed the JSP engine, this support involves recognition, translation, and management of the Java Server Page lifecycle and its interaction components.

Components:

It was mentioned earlier that the Java Server Pages architecture can include reusable Java components. The architecture also allows for the embedding of a scripting language directly into the Java Server Pages file. The components current supported include Java Beans, and Servlets.

Processing:

A Java Server Pages file is essentially an HTML document with JSP scripting or tags. The Java Server Pages file has a JSP extension to the server as a Java Server Pages file. Before the page is served, the Java Server Pages syntax is parsed and processed into a Servlet on the server side. The Servlet that is generated outputs real content in straight HTML for responding to the client.

Access Models:

A Java Server Pages file may be accessed in at least two different ways. A client's request comes directly into a Java Server Page. In this scenario, suppose the page accesses reusable Java Bean components that perform particular well-defined computations like accessing a database. The result of the Beans computations, called result sets is stored within the Bean as properties. The page uses such beans to generate dynamic content and present it back to the client. In both of the above cases, the page could also contain any valid Java code. Java Server Pages architecture encourages separation of content from presentation

Steps in the execution of a JSP Application:

1. The client sends a request to the web server for a JSP file by giving the name of the JSP file within the form tag of a HTML page.

2. This request is transferred to the Java WebServer. At the server side Java WebServer receives the request and if it is a request for a jsp file server gives this request to the JSP engine.

3. JSP engine is program which can understands the tags of the jsp and then it converts those tags into a Servlet program and it is stored at the server side. This Servlet is loaded in the memory and then it is executed and the result is given back to the JavaWebServer and then it is transferred back to the result is given back to the JavaWebServer .

JDBC connectivity

The JDBC provides database-independent connectivity between the J2EE platform and a wide range of tabular data sources. JDBC technology allows an Application Component Provider

- ➢Perform connection and authentication to a database server
- ➢Manager transactions
- ➢Move SQL statements to a database engine for preprocessing and execution
- ➢Execute stored procedures
- ➢Inspect and modify the results from Select statements.

Tomcat 6.0 web server

Tomcat is an open source web server developed by Apache Group. Apache Tomcat is the servlet container that is used in the official Reference Implementation for the Java Servlet and Java Server Pages technologies. The Java Servlet and Java Server Pages specifications are developed by Sun under the Java Community Process. Web Servers like Apache Tomcat support only web components while an application server supports web components as well as business components (BEAs Weblogic, is one of the popular application server).To develop a web application with jsp/servlet install any web server like JRun, Tomcat etc to run your application.

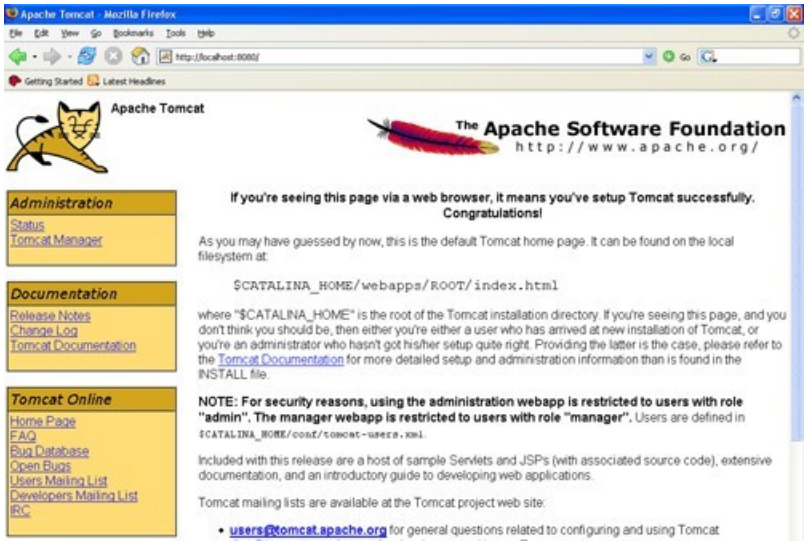


fig 5.1 Tomcat 6.0 web server

5.2 SAMPLE CODE

```
<!DOCTYPE html>

<!--[if lt IE 7 ]><html class="ie ie6" lang="en"> <![endif]-->

<!--[if IE 7 ]><html class="ie ie7" lang="en"> <![endif]-->

<!--[if IE 8 ]><html class="ie ie8" lang="en"> <![endif]-->

<!--[if (gte IE 9)|!(IE)]><!--><html class="not-ie" lang="en"> <!--<![endif]-->

<head> <!-- Basic Meta Tags -->

<meta charset="utf-8">

<title>Add / View Categories..</title>

<meta name="description" content="ucorpora demo - Free Business Corporate HTML Template">

<meta name="keywords" content="ucorpora, ucorpora demo, free, template, corporate, clean, modern, bootstrap, creative, design">

<meta name="viewport" content="width=device-width, initial-scale=1.0">

<!--[if (gte IE 9)|!(IE)]>

<meta http-equiv="X-UA-Compatible" content="IE=edge">

<meta http-equiv="Content-Type" content="text/html;charset=utf-8">

<![endif]-->

<!-- Favicon -->

<link href="img/favicon.ico" rel="icon" type="image/png">

<!-- Styles -->

<link href="css/styles.css" rel="stylesheet">

<link href="css/bootstrap-override.css" rel="stylesheet">

<!-- Font Avesome Styles →
```

```
<link href="css/font-awesome/font-awesome.css" rel="stylesheet">

<!--[if IE 7]>

<link href="css/font-awesome/font-awesome-ie7.min.css" rel="stylesheet">

<![endif]-->

<!-- FlexSlider Style

<link rel="stylesheet" href="css/flexslider.css" type="text/css" media="screen">

<!-- Internet Explorer condition - HTML5 shim, for IE6-8 support of HTML5 elements -->

<!--[if lt IE 9]>

<script src="http://html5shim.googlecode.com/svn/trunk/html5.js"></script>

<![endif]-->

<style type="text/css">

<!-- .style2 {

color: #0000FF;

font-size: 14px;}

.style10 {color: #FF00FF; font-weight: bold; font-size: 13px; }

.style11 {font-size: 30px}

.style13 {

color: #009900;

font-size: 31px;}

.style22 {

color: #0000FF;

font-weight: bold;

font-size: 13px;

}
```

```
.style26 {
color: #009900;
font-size: 14px;
}
.style93 {font-size: 25px; color: #FF00FF; }
.style57 {font-size: 15px}
.style49 {color: #FF0000; font-weight: bold; font-size: 14px; }
.style104 {color: #0033FF; font-weight: bold; font-size: 13px; }
.style105 {
color: #FF00FF; font-size:14px; } </style> </head>
```

6. TESTING

TESTING METHODOLOGIES

The following are the Testing Methodologies:

- **o Unit Testing.**
- **o Integration Testing.**
- **o User Acceptance Testing.**
- **o Output Testing.**
- **o Validation Testing.**

Unit Testing

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module's control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All important processing path are tested for the expected results. All error handling paths are also tested.

Integration Testing

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design

The following are the types of Integration Testing:

1. Top Down Integration

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner. In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

2. Bottom-up Integration

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

- The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
- A driver the control program for testing is written to coordinate test case input and output.
- The cluster is tested.
- Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

User Acceptance Testing

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

Output Testing

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

Validation Checking

Validation checks are performed on the following fields.

Text Field:

The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables. Incorrect entry always flashes and error message.

Numeric Field:

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to perform. Each module is subjected to test run along with sample data. The individually tested modules are integrated into a single system. Testing involves executing the real data information is used in the program the existence of any program defect is inferred from the output. The testing should be planned so that all the requirements are individually tested.

A successful test is one that gives out the defects for the inappropriate data and produces and output revealing the errors in the system.

Preparation of Test Data

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role

in the system testing. After preparing the test data the system under study is tested using that test data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

Using Live Test Data:

Live test data are those that are actually extracted from organization files. After a system is partially constructed, programmers or analysts often ask users to key in a set of data from their normal activities. Then, the systems person uses this data as a way to partially test the system. In other instances, programmers or analysts extract a set of live data from the files and have them entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And, although it is realistic data that will show how the system will perform for the typical processing requirement, assuming that the live data entered are in fact typical, such data generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

Using Artificial Test Data:

Artificial test data are created solely for test purposes, since they can be generated to test all combinations of formats and values. In other words, the artificial data, which can quickly be prepared by a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrote the programs. Often, an independent team of testers formulates a testing plan, using the systems specifications.

The package “Virtual Private Network” has satisfied all the requirements specified as per software requirement specification and was accepted.

USER TRAINING:

Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

MAINTENANCE:

This covers a wide range of activities including correcting code and design errors. To reduce the need for maintenance in the long run, we have more accurately defined the user’s requirements during the process of system development. Depending on the requirements, this system has been developed to satisfy the needs to the largest possible extent. With development in technology, it may be possible to add many more features based on the requirements in future. The coding and designing is simple and easy to understand which will make maintenance easier.

TESTING STRATEGY :

A strategy for system testing integrates system test cases and design techniques into a well planned series of steps that results in the successful construction of software. The testing strategy must co-operate test planning, test case design, test execution, and the resultant data collection and evaluation .A strategy for software testing must accommodate low-level tests that are necessary to verify that a small source code segment has been correctly implemented as well as high level tests that validate major system functions against user requirements.

Software testing is a critical element of software quality assurance and represents the ultimate review of specification design and coding. Testing represents an interesting anomaly for

the software. Thus, a series of testing are performed for the proposed system before the system is ready for user acceptance testing.

SYSTEM TESTING:

Software once validated must be combined with other system elements (e.g. Hardware, people, database). System testing verifies that all the elements are proper and that overall system function performance is achieved. It also tests to find discrepancies between the system and its original objective, current specifications and system documentation.

UNIT TESTING:

In unit testing different are modules are tested against the specifications produced during the design for the modules. Unit testing is essential for verification of the code produced during the coding phase, and hence the goals to test the internal logic of the modules. Using the detailed design description as a guide, important Conrail paths are tested to uncover errors within the boundary of the modules. This testing is carried out during the programming stage itself. In this type of testing step, each module was found to be working satisfactorily as regards to the expected output from the module.

In Due Course, latest technology advancements will be taken into consideration. As part of technical build-up many components of the networking system will be generic in nature so that future projects can either use or interact with this. The future holds a lot to offer to the development and refinement of this project.

7. SCREEN SHOTS

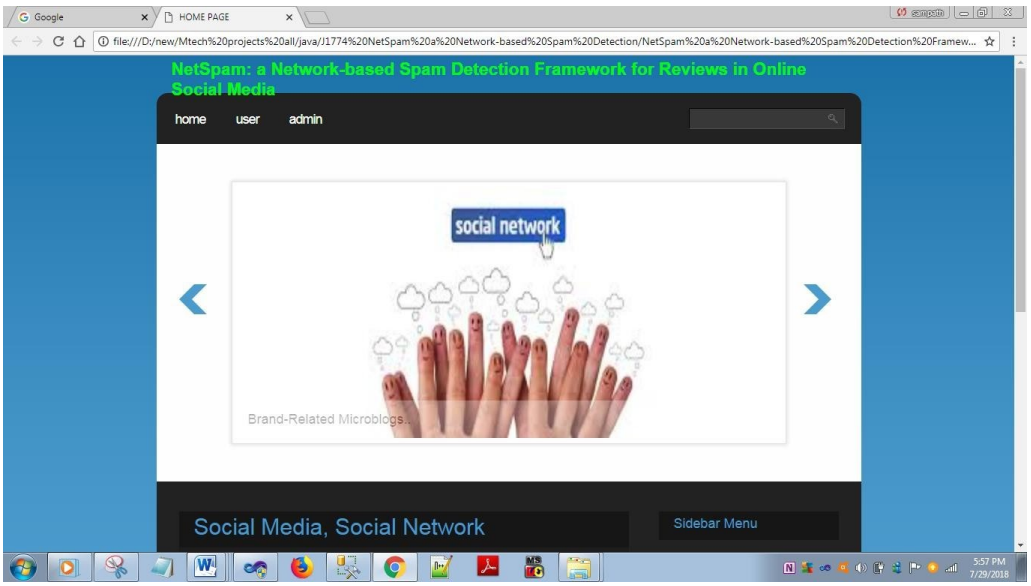


fig 7.1 Home

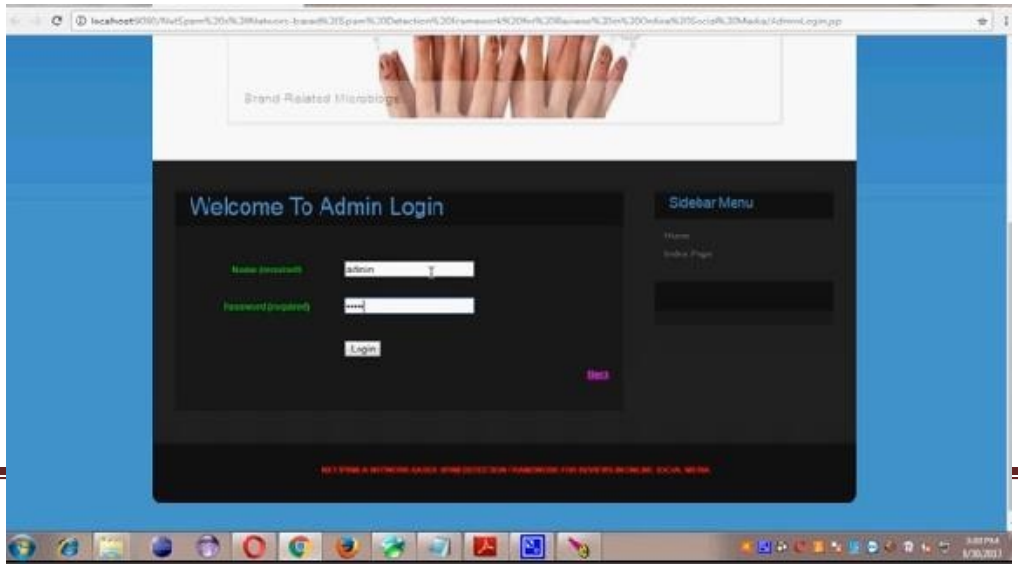


fig 7.2 Admin Login

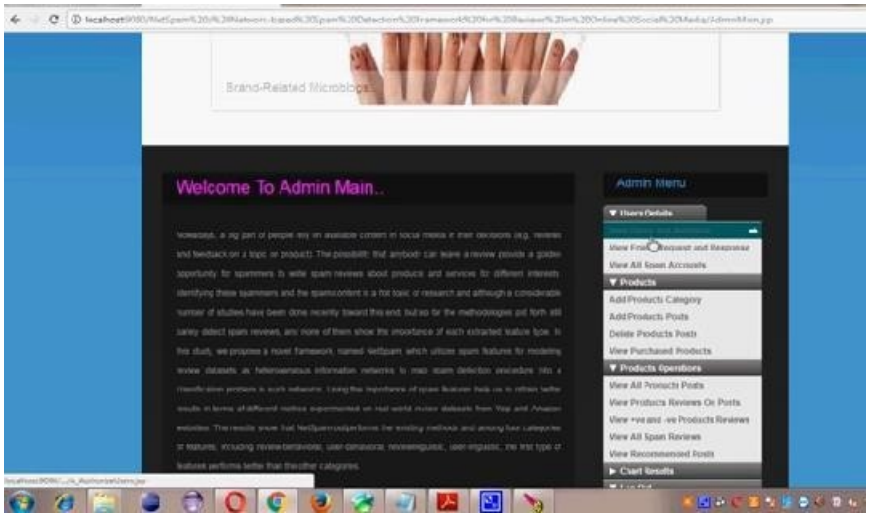


fig 7.3 Admin Home

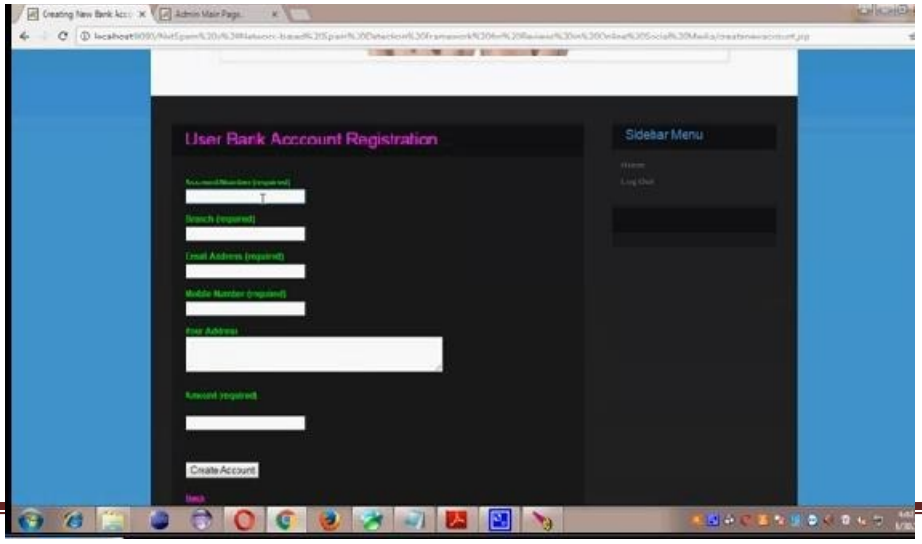


fig 7.4 User Bank Account Registration

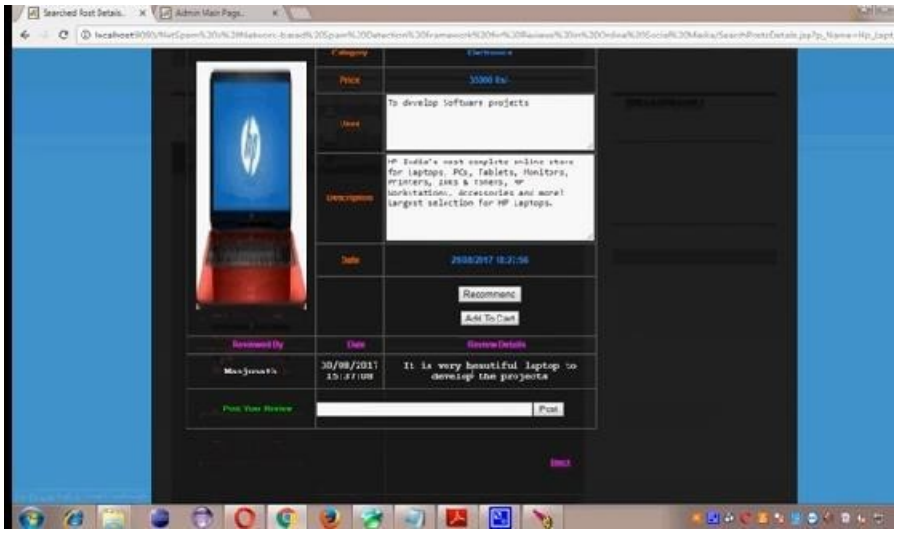


fig 7.5 Product Description

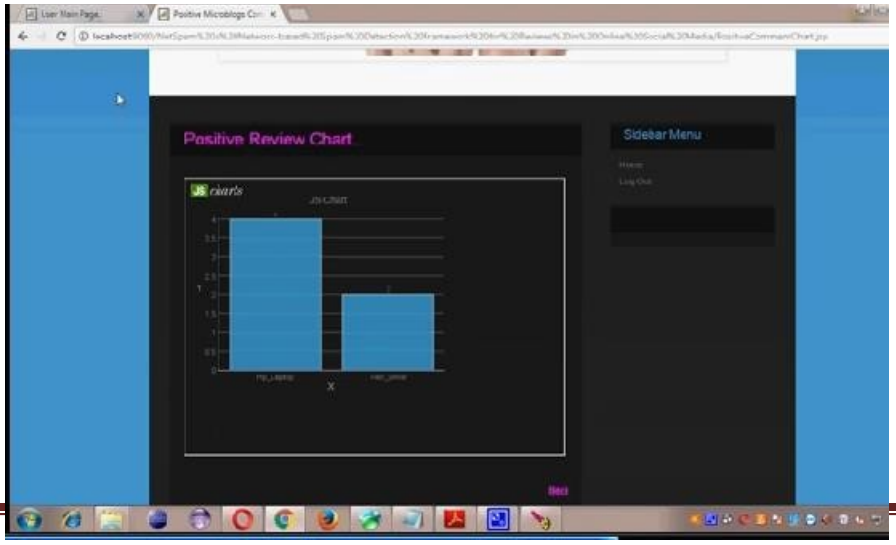


fig 7.6 Positive Review Chart

fig 7.8 Authorize Users

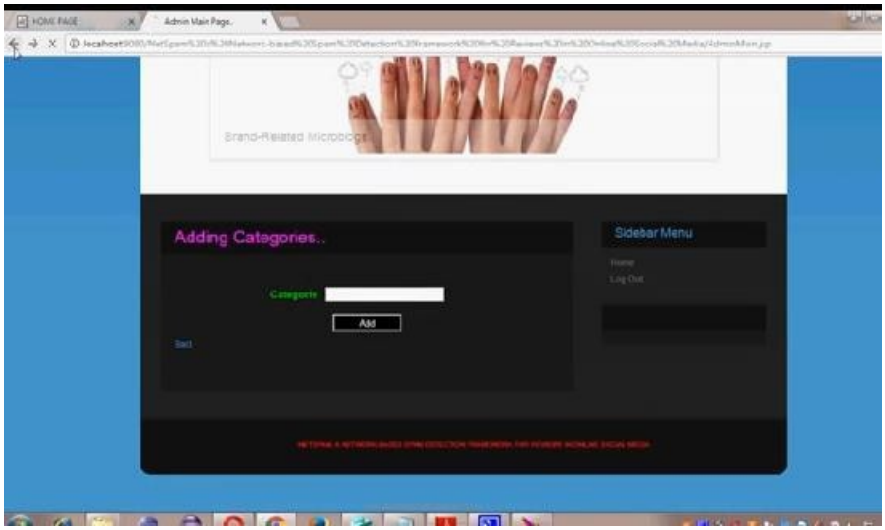


fig 7.9 Adding Categories

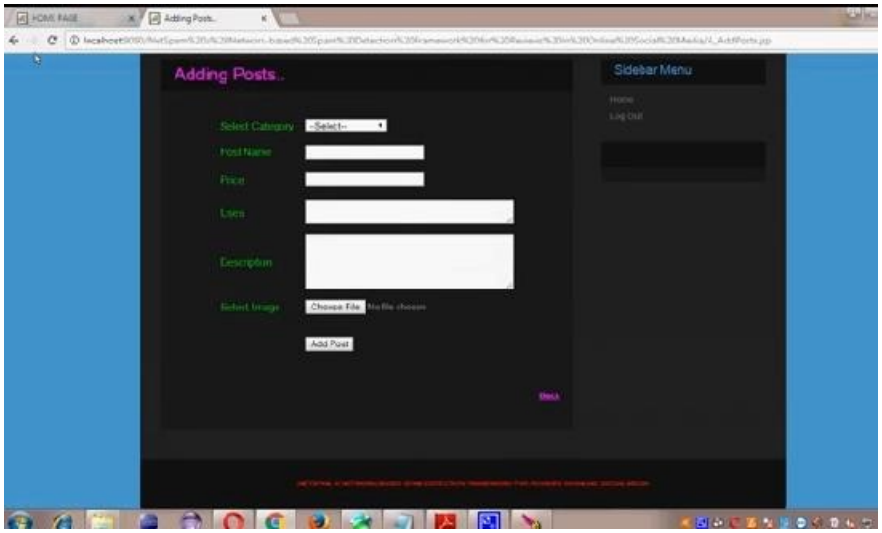


fig 7.9.1 Adding Posts

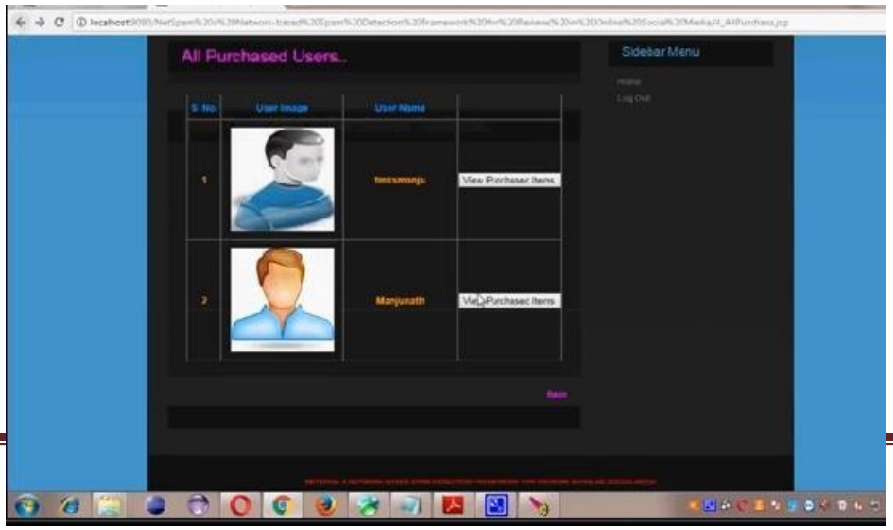


fig 7.9.2 All Purchased Users

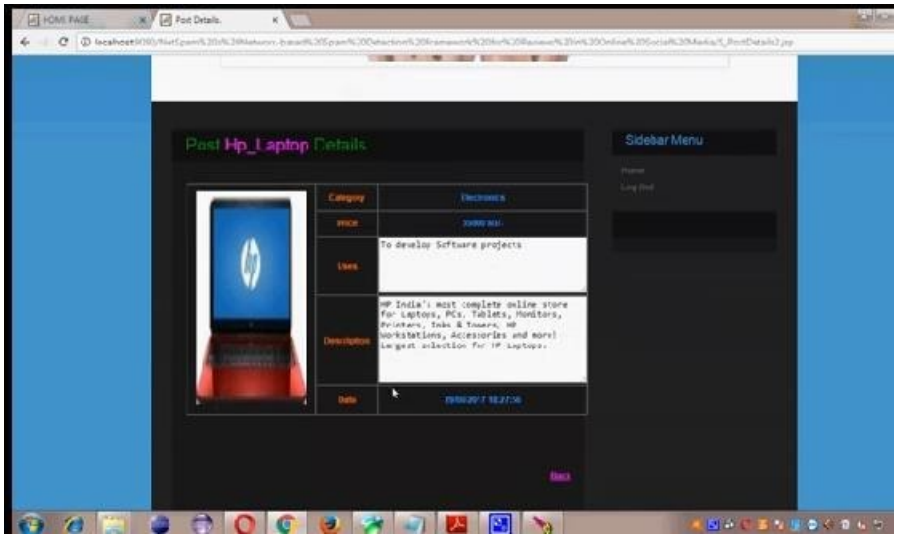


fig 7.9.3 Post Details

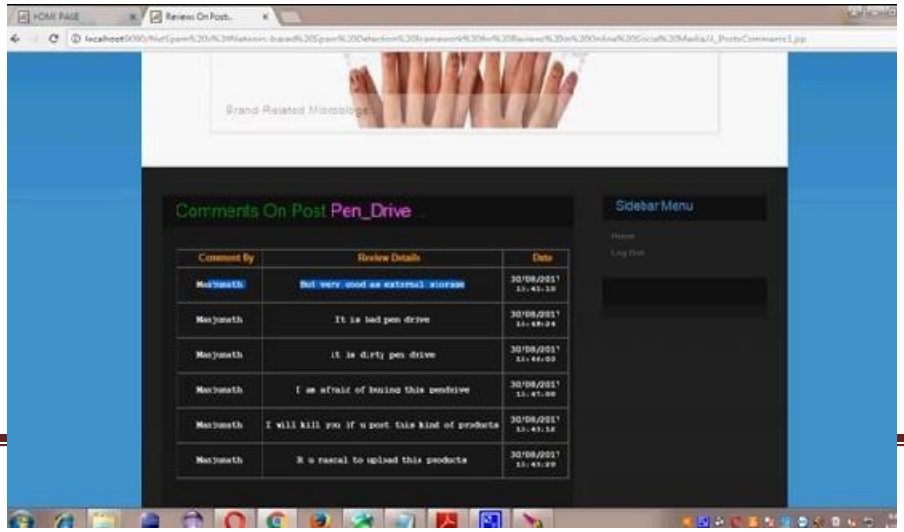


fig 7.9.4 Comments On Post

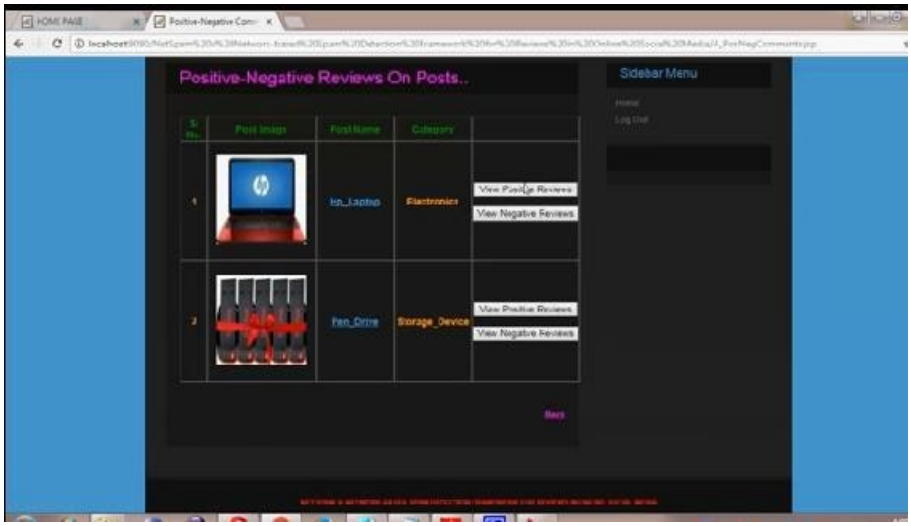


fig 7.9.5 Positive-Negative Reviews On Posts

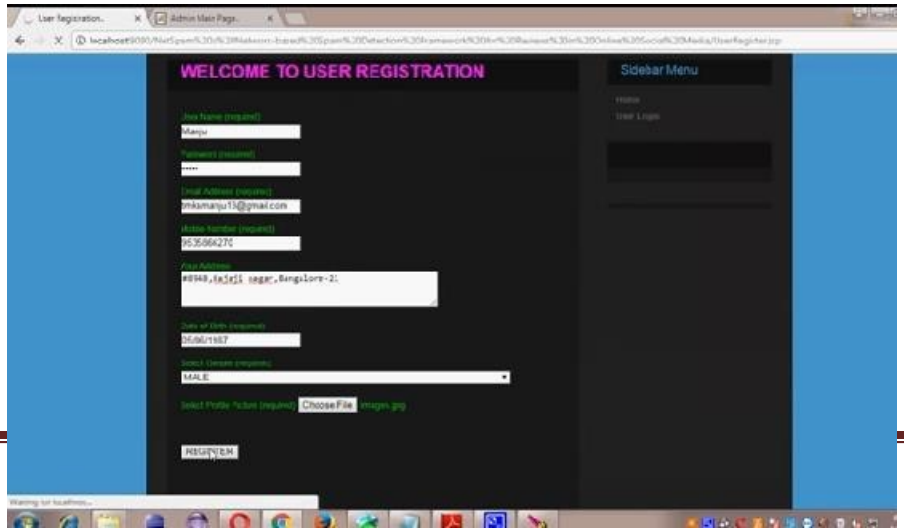


fig 7.9.6 Welcome to User Registration

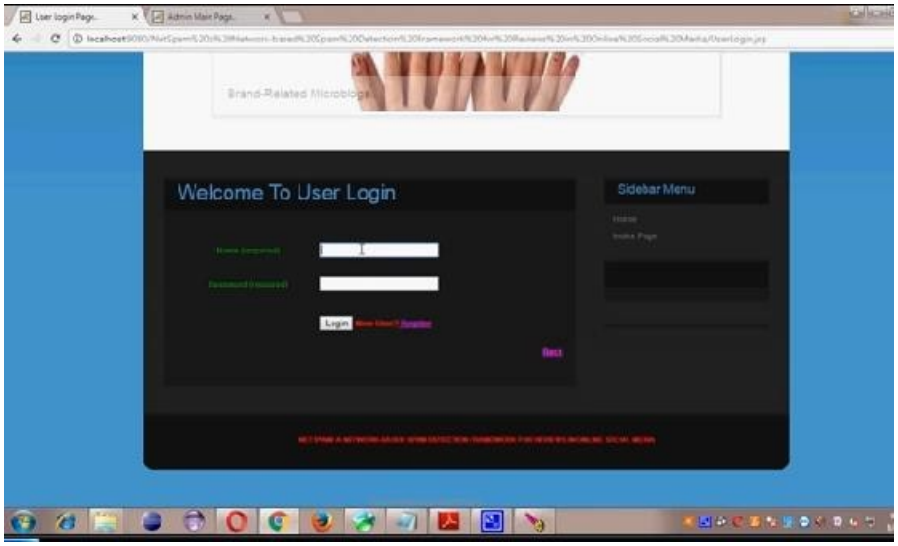


fig 7.9.7 Welcome To User Login

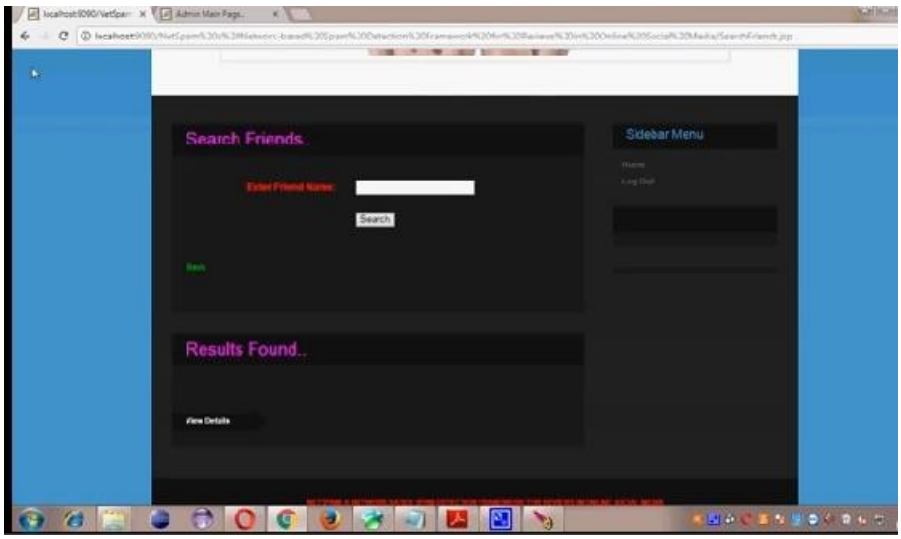


Fig 7.9.8 Search Friends

8. CONCLUSION

This study introduces a novel spam detection framework namely NetSpam based on a meta path concept as well as IEEE Transactions on Information Forensics and security volume. The performance of the proposed framework is evaluated by using two real-world labeled datasets of help and Amazon websites. Our observations show that calculated weights by using this meta path concept can be very effective in identifying spam reviews and leads to a better performance. In addition, we found that even without a train set, NetSpam can calculate the importance of each feature and it yields better performance in the features' addition process, and performs better than previous works, with only a small number of features. Moreover, after defining four main categories for features our observations show that the reviews behavioral category performs better than other categories, in terms of AP, AUC as well as in the calculated weights. The results also confirm that using different supervisions, similar to the semi-supervised method, have no noticeable effect on determining most of the weighted features, just as in different datasets.

For future work, metapath concept can be applied to other problems in this field. For example, Similar framework can be used to find spammer communities. For finding community, reviews can be connected through group spammer features and reviews with highest similarity based on metapth concept are known as communities. In addition, utilizing the product features is an interesting future work on this study as we used features more related to spotting spammers and spam reviews. Moreover, while single networks has received considerable attention from various disciplines for over a decade, information diffusion and content sharing in multilayer networks is still a young research. Addressing the problem of spam detection in such networks can be considered as a new research line in this field.

9. REFERENCES

[1] J. Donfro, A whopping 20 % of yelp reviews are fake. http://www.businessinsider.com/20-percent-of-yelp-reviews-fake-2013-9. Accessed: 2015-07-30.

[2] M. Ott, C. Cardie, and J. T. Hancock. Estimating the prevalence of deception in online review communities. In ACM WWW, 2012.

[3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock. Finding deceptive opinion spam by any stretch of the imagination.In ACL, 2011.

[4] Ch. Xu and J. Zhang. Combating product review spam campaigns via multiple heterogeneous pairwise features. In SIAM International Conference on Data Mining, 2014.

[5] N. Jindal and B. Liu. Opinion spam and analysis. In WSDM, 2008.

[6] F. Li, M. Huang, Y. Yang, and X. Zhu. Learning to identify review spam. Proceedings of the 22nd International Joint Conference on Artificial Intelligence; IJCAI, 2011.

[7] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. Exploiting burstiness in reviews for review spammer detection. In ICWSM, 2013.

[8] A. j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos. Trueview: Harnessing the power of multiple review sites. In ACM WWW, 2015.

[9] B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Towards detecting anomalous user behavior in online social networks. In USENIX, 2014.

[10] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao. Spotting fake reviews via collective PU learning. In ICDM, 2014.

[11] L. Akoglu, R. Chandy, and C. Faloutsos. Opinion fraud detection in online reviews bynetwork effects. In ICWSM, 2013.

[12] R. Shebuti and L. Akoglu. Collective opinion spam detection: bridging review networks and metadata. In ACM KDD, 2015.