1.0 Introduction

In today's modern world internet is being used by different terrorist organizations like militant and sectarian groups for their purpose. In last ten to fifteen years it increases much, the purposes includes recruitment and training of militants, financing for their activities, propaganda, provoke others to commit different type of terrorism, and preaching for terrorist purposes. These organisations also use internet for exchanging information and communication with each other. With increasing cyber terrorist activities it is important to monitor and curb these activates. Terrorist groups use different sources for spreading terrorism on internet like through literature, and hate speeches etc. Basically they use internet for brain washing of those individuals who haven't proper knowledge of religion and pessimist from life and society. They also promote their specific ideology and propaganda through web pages that attracts people who have some extremist views and who have some other social problems to join their organizations.

After 9/11 attacks when two airplanes hits American World Trade Centre, from that day the world has faced many terrorist attacks like attack in hotel of France etc. This was also the time when terrorism starts grown their roots in Pakistan and some other countries especially in Afghanistan. Everything was changed in Pakistan when American and NATO forces attacks in Afghanistan in context of 9/11 attacks. Terrorism was now a major problem and it was raise to the top of the political and military agenda in Pakistan.

ISIS, Al Qaeda and TTP are three large organizations which use internet for their purpose of preaching, recruitment and funding etc. According to latest reports FBI America and BBC number of educated young girls and boys join them and went to them in Afghanistan, Syria, Iraq, Yemen and other countries for fighting, mostly of them are from European countries. Mostly young age girls and boys are recruited through social media platforms where they built relationships with agents of terrorist organisation and talk on Islam and challenges face by Islam from modern/liberal media and countries. Extremist groups mostly looking for people who have not exact knowledge about different subjects of Islam because they think they can guide them on their path easily. They also search for people who belongs from isolated and backward areas because they have a naturally disappointed from their social issues and it is easy to convince them and to make their mind according to their agenda. College and university students are also their major target because this age is full of youth, thrill and anti-state acts.

1.1 Problem Domain

Many Software's were developed for analysing terrorist websites working on internet for example "Black duck software", "Raytheon's Riot Software" they have ability to identify the movement of people and predicts their behaviour by mining data from social media websites. These software's are basically analyse "text", "audio" and "video" data which are used by militant groups for their dirty aims and objectives. However the Existing software's in Pakistan for analysing black websites are working manually and also not working 24/7. Moreover lack of capability of storing large amount of information about black websites fully and quickly. And due to high level security issues, high volume of data should be maintained and analysed. There is no latest automated way to analyse black websites efficiently and effectively in Pakistan.

1.2 Problem Statement

Today the terrorist/black websites are being maintained and analysed manually by security agencies and specific state authorities, which is not an efficient way it takes much time and they face more difficulties and not able to search and analyse each and every website. The result generated by these systems was not much accurate. This proposed system could help to analyse as well as report about terrorist/black websites to specific authorities in a fast and efficient. This system can be used by the security special government authorities who work for the security of country. After registration in system these users can be authorized to analyse websites by giving domain name or URL of website to the system, after analysing system can give ratio rate of sentiment of text and also provides IP address of the websites can also view information in a meaningful way.

1.3 Aims and Objectives

The main objective of developing and implementing this project is to contribute with state authorities to fight against cyber terrorism in the country. It provides fast and efficient way to analyse websites. In present time security agencies analyse the manually which takes a lot of time and many websites remain unchecked. Thus, main focus of this project is to analyse websites in fast and automated way and gives optimal results.

1.4 Resource Requirements

There are two types of resources which are required for developing the system.

1.4.1 Software Requirements

The software resources required for system are windows 7, 8, 8.1 and 10, windows XP, Microsoft Visual Studio and Microsoft SQL Management studio.

1.4.2 Hardware Requirements

The hardware resources required for system are dual core or latest processor. The device on which this system should have at least 1 gigahertz (GHz) or faster processor and at least 2 GB of RAM. System should have 50 GB hard disk space and have Internet connection.

1.5 Tools and Technologies

For this project the latest and efficient versions of tools are used for the development of this system. The system used Microsoft visual studio 2013 and SQL Management studio 2012. Details are as follows:

1.5.1 Microsoft Visual Studio

Microsoft Visual Studio is one of the strongest development tool launched/introduced by Microsoft in 2002. It is used for developing different type of computer programs for windows like desktop applications, as well as websites, web applications and web services. Visual Studio uses different platforms of Microsoft software development such as, windows and web forms, windows presentation etc. It produces two types of codes one is “native code” and second is “managed code”. [10].

1.5.2 SQL Server Management Studio

The SQL Server Management Studio also known as “SSMS” is a software application first deployed in 1998 with Microsoft SQL Server that is used for different purposes like controlling, organizing, and managing all the attributes present in the Microsoft SQL Server. It has also features like configuring, and administering all components within Microsoft SQL Server. The tool includes both script editors and graphical tools which work with objects and features of the serve [6]

1.6 Project Scope

The scope of project is that it only analyse the websites which have extension ".pk" and ".com". The proposed system extract the contents of webpage using web content mining algorithm to mine only "textual information" from web pages. Web mining also consists of text mining methodologies that allow us to scan and extract useful content from unstructured data and analyse the content of pages using Data mining technique. It detect web pages that are involve in spreading terrorism by matching sentence which are properly maintained in database. This system classify the web pages into various categories and sort them appropriately. The system gives IP address and URL to users/agencies from that they are capable to detect the location of these websites.

1.7 Report Layout

The outline of project report is described below:

The first chapter describes the introduction of project. It includes problem domain, Problem statement, development methodology, aims and objectives, project scope, tools and technologies and resource requirements. Second chapter includes discussion on literature review. It includes background of terrorism, Existing work, what are black sites, Methodologies used for Content mining and web mining used in terrorism analysis and project limitations. In third chapter explain requirement specification. Functional and non-functional requirements, modules of the system are explained in details. Next chapter consists of proposed system and system modelling. System analysis, evaluation of existing work, proposed system, use case and Use case diagram, sequence system with diagram, data flow diagram.

The fifth chapter explains the result, testing, validation and analysis of the system. Sixth chapter consists of conclusion and future work. Achievements are briefly described. Future recommendations are also explained in this chapter.

2.0 Background

The word "terrorism" comes from the French word, which is based on the Latin verb “terrier” (to cause to tremble). In year 1795 when it was used to describe the actions of the Jacobin Club in their rule of post-Revolutionary France, the so-called "Reign of Terror". Jacobins are rumoured to have coined the term "terrorists" to refer to themselves. It is basically a strategy to create violence, threats socially, and attacks on population, in order to create fear, cause disruption, and ultimately, brings about compliance with specified political, religious, or ideological demands. The European Union includes in its 2002 definition of “terrorism" the aim of destabilizing or destroying the fundamental political, constitutional, economic or social structures of a country.” Terrorism is defined in the U.S. by the Code of Federal Bureau of Investigation as: “.the unlawful use of force and violence against persons or property to intimidate or coerce a government, the civilian population, or any segment thereof, in furtherance of political or social objectives” [1].

Terrorism of present days grown their roots after 9/11 incident when 3 to 4 fighter planes hit world trade centre in United States of America in this incident thousands of people killed. After that USA and UN start fighting against different groups most probably “Al-Qaeda” in Afghanistan. Most people, groups, authorities claim that 9/11 attacks were planned in America by American state for completing their different agenda’s. After the attack of USA and NATO forces in Afghanistan extremist forces grown and stand in Pakistan and all over the world against America and their allies’ action in Afghanistan. Al-Qaeda and other organizations promote their agenda using internet, they use it for funding, preaching and recruiting purpose [2].

2.1 Black Websites

There are millions of websites present on internet and hundreds of them are those which contain content related to terrorism, sectarianism and violence etc. Terrorist organization used these websites for different purposes like for spreading terrorism, recruiting terrorists, promoting their agenda and literature for their aims and objectives. These groups convey their message in the form speeches and literature. Majority of terror websites not openly announce their acts and success, however they focus on two basic issues one on the restrictions which placed on the freedom of expression and the situation of companions who are political prisoners [3].

Al-Qaeda is the first organization which used internet for promoting their agenda and it was the "Al-Neda Centre for Islamic Study and Research" which was created by the Sheikh Yousaf who was the bodyguard of Osama Bin Laden. He made a video to spread propaganda on the internet which contains messages from Sheikh Yousaf in which he said "In the first stage, the stage of attrition and engaging the enemy in battle, you need to make your enemy tired more than you need to kill a large number of its members. You need to scatter the enemy, demoralize it, spread it out over a large area, and cause it to get tired. If the enemy is spread out, it needs a supply armies and a lot of other things. You need to make the enemy reach this stage. In this stage, you should strike, run, and disappear, strike at the weak points." This man also published a book on the website of "Al-Neda Centre for Islamic Study and Research". So, he was the person who created the video and published the book for spreading propaganda, and through cyber acts they had trying to attract more peoples to them for their aims and objectives, and giving them agenda to work for them on different places [4].

2.2 Existing Work

Today by using different manual methods the black/terrorist websites are being monitor and analyse by different state authorities and security agencies. These methods take a lot of time in detecting and analysing websites and moreover they are unable to visit and monitor each and every terror related website. This system could focus to monitor as well as report about websites to different authorities and agencies and this is the fast and efficient way to recognize the illegal, unethical and terror related websites (Black websites) and this proposed system works at any time and active on internet like Botnet malware and visit types on websites on the base of their extension and report to its Botnet Master. This system can be used by specific users which are authorized by admin like cops and other government officials who work for country security.

2.3 Methodologies used for Content Mining

There are two methodologies used for content mining which are following:

Data mining is first technique which is used to extract specific type of data from large set of information. It is used in both web applications and desktop application. It is an application of web mining. Second is web mining which is used to mine content from web. There are different are different algorithms used for both technique's which work

according to given criteria. In both cases proper database is maintained though datasets. C4.5, SVM and KNN are some algorithms that are used.

2.3.1 Data Mining:

It is the method and way used for analysing the data from large dataset present on internet. There are different software's designed for data mining purpose from numbers of analytical software's/tools which allows users to analyse data from different aspects.

2.3.1.1 Data Mining Procedure:

It includes two stages which includes pre-processing in which data set must collect before data mining algorithms are being use. As data mining algorithms are used to display/discover the patterns which are in the form of groups present in large data. Before using data mining technique pre-processing is necessary to analyse the data sets. And second is result validation in which unintentionally data mining technique used in wrong way, and then it produced results which appear to be remarkable, but these results actually do not predict upcoming behaviour and cannot be regenerate on a new sample of data. Often the results from researches formed many assumptions and not performing proper statistical testing based on assumptions. Machine learning is a kind of this problem known as over fitting but the same kind of problems may rise up during different phases of the process and then test breaks [9].

2.3.1.2 Data Mining Techniques:

There are two different techniques used for data mining one is descriptive is a technique used for data mining. It has further two types which are clustering and association. This technique is basically extracted and analyse information which is working in present scenario. And also for inter linking, and to find frequency and to make a regularities etc. Predictive is second technique used for data mining purpose. It has also further two types which are classification and regression. This technique is used to analyse the future behaviour of given data or information. The result of prediction can be alphanumeric value or may divide in different categories. Data mining can unintentionally be misused, and can then produce results which appear to be significant; but which do not actually predict future behaviour and cannot be reproduced on a new sample of data and bear little use [9].

2.3.2 Web Mining:

It is used to find data patterns from web documents and also used to find out knowledge from them. It is categorized among the applications of data mining.

2.3.2.1 Web Mining Techniques:

There are three techniques used for web mining which includes web content mining the technique which is used to discover different type of meaningful information from large set of information which is present on different web documents. Different techniques of data mining apply in web content mining. It basically deals with data which is in unstructured form. Second is web usage mining technique which is used for extracting meaningful information from different web pages. It is also use to find behaviour of user what they want to do on internet. For example someone may search for textual data and other for images. And last one is web structure mining. It use graph theory is used to analyse the nodes and connection structure of the websites. There are two further techniques of web structure mining one is extracting different patterns from the hyperlinks and other is mining the documented structure [8].

2.4 Web Mining In Terrorism Analysis

In this system terrorism analysis web content mining is used for extracting the textual information.

Cluster data mining technique is used in web content mining to discover textual information. Sentiment analysis performs on extracted data to find the nature of data. Sentiment analysis tells the attitude of person while writing anything is either positive or negative. By using text mining algorithm the textual information extracted from web pages and this is free from tags. The information categorized in different groups before storing in data base. The set of information present in a data base have same characteristics that have extracted data. Hence, it is basically used for finding groups of same data.

2.5 Limitations of project

The system is not without limitations. This project has some limitations that it is limited on some websites and it does not support all the domains of internet. This system is only specify for websites with extension like ".pk" and ".com". It only analyse single website at one time. The system does not extract content in audio, video or pictorial form It only mine textual content from web pages by using data mining techniques. Without the

permission of user the system itself unable to block websites. The system is not going to use web service. The system analyse the websites which contains content in English language other than English language it is unable to analyse. The current system is only access by specific users like government agencies etc. The system is not capable to analyse and detect the code words used in black/terrorist website.

3.0 Requirements Specification

In this chapter, the Software Requirement Specification of the proposed system should be discussed. Functional requirements are the requirements which are need by the system to perform proposed application e.g. in this case users add the key words, mine the data to extract specific data and information, select the different methods and apply data mining algorithms to achieve the specific results. While non-functional requirements are those requirements which do not change the functionality of the system, but these are the properties which the system must have for e.g. maintainability, reliability, efficiency etc.

3.1 Functional Requirements

The definition present in systems engineering is that, a functional requirement defines the functionality of a system or its component. A function may include: inputs, processing, and outputs. There are different functional requirements such as mining, data manipulation and processing and other specific functionality that define what a system is supposed to achieve. Functional requirements are identified in the beginning and derived from pre-defined use cases. Non-functional requirements provide a base to functional requirements, which represent limitations on the design or implementation, such as performance requirements, security, reliability etc. In this proposed system, each functional requirement has a unique reference ID which is defined as:

3.1.1 Log-in

The first module of the system develops to register the specific user. When users are registered then they have access to use the application. The login page is created through database of the system, which can handle any type of user. The basic information of any new user would be inserted into table located in database of the system. Through this process user login get secure and verified. All the users go through registration process individually. Login can only be used for moving in main menu of the system. When a user login to the system they can perform some action from home page of the system. If users want to analyze any type of website then he/she must be registered first to system and then they can use system according to their need.

Table 3.1: Requirements for Log-in

Sr. No.	Description
R1.0	User can easily get access to the registration form for registration.
R1.1	User can easily perform the action of login.
R1.3	The details of the users registration was inserted into the database table.

3.1.2 Main Screen

After login user redirect to home screen of system where they have many options regarding the system and have proper guideline how to use the system, but the important section is input the domain name of website for analysing the content in it. The users write domain name in text field available on screen. The result was generated on this screen in grid and in the form of graphs. The home screen is display on different platforms like mobiles, tablets, laptops and PCs etc. The
interface of home screen is very attractive and easy to understand by users.

Table 3.2: Requirements for Main Screen

Sr. No.	Description
R1.0	User can easily understand home screen features.
R1.1	User can easily perform their tasks on home screen.
R1.2	User can easily understand how to use the system.
R1.3	The user input the domain name of website for analysing.

3.1.3 Website analysis

The website analysis is an important section of the system where user analyse website by extracting textual content from and match it with data base value to check either the content in website gives negative meaning or positive. This system can be easily understood by users. The interface of the current system is user friendly which provides easiness to users. User can perform different action in this module like extract textual content by eliminating tags and also perform sentiment analysis on this content. In analysis module it tells the extract web content and request sentiment analysis by the

system. It tells the user that the extracted content in either have positive meaning or a negative [7].

Table 3.3: Requirements for Website Analysis

Sr. No.	Description
R1.0	User can easily access service of web content mining And analysis.
R1.1	User can extract web content and performs sentiment analysis.
R1.2	Through website analysis user can easily perform its task.
R1.3	User can easily use and understand the system.

3.1.4 View Records

The next functional requirement is view record by user. In this module user can view record of all websites he checked and analyse in managed way. This view record module contains record of four parameters which are searched words, web URL, match ratio and status. The current system "Terrorism Analysis" shows the accurate results to all the users and also show results in graphs. This module is also the part of terrorism analysis system and researchers also give importance of this in their researches.

Table 3.4: Requirements for View Records

Sr. No.	Description
R1.0	System gives accurate information in record.
R1.1	The task was performed by the system efficiently.
R1.2	The user can use this system without any failure and error.
R1.3	The data show in record was secured.

3.1.5 Manage Websites

In fifth module user can managed checked websites. In this section user can see three options in grid which were "URL of website", "IP address" and "status". In status there are three options black listed, pending and not pending which was used by user in order to give status to websites and the color of URL of websites change according to these three options. By the help of this information user has option to block the website or not

Table 3.5: Requirements for Manage Websites

Sr. No.	Description
R1.0	User can easily manage websites.
R1.1	System gives accurate results to the users.
R1.2	User has option to block the black listed websites.
R1.3	User put website in pending status for further monitor.

3.1.6 Report Generation

The report generation is important section in any project which gives result of all processes in one place. In the current system user can also generate report at the end. In report user have all the information about check websites which include status of website, match ratio, polarity graphs etc. Users have also option to print report in pdf, docs or xml format [5].

Table 3.6: Requirements for Report Generation

Sr. No.	Description
R1.0	System generates the report of required task.
R1.1	User input the website domain name to check whether the system is available or not.
R1.2	The user must assure the availability of internet connection.
R1.3	The generated report also shows results in graphical form.

3.1.7 Storage Module

In this module user can store result in hard drive of system. Storage Module consist of three types datasets which can also be access by user, according to the criteria of the data shared in websites storage module consists websites in its datasets. The advantage of Storing data is to keep record of the system. The system can not repeat the specific websites which is already visit by the user.

Table 3.7: Requirements for Storage of data

Sr. No.	Description
R1.0	The user can store result in the local hard drive.
R1.1	The user submits their task through local hard drive.
R1.2	The user analyses the websites through internet.

3.2 Non Functional Requirements

Non-functional requirements have different features and functions that are offered in given proposed system, these are the some extra requirements that actually do nothing. These requirements supports functional requirements and ought to provide specific dimensions that the software has to meet like Availability, Accuracy, User friendly, Reliability, Efficiency, Flexibility etc. The non-functional requirements for the application that are meeting in the project are as follows.

3.2.1 User Friendly

During use users can easily performs all actions like input domain name, extract content, find IP address, get report etc. This system was made to give satisfaction to user, system performs tasks without wasting time give desire output to the user. The system is easy to use and understandable by users.

3.2.2 Efficiency

The designed system “terrorism analysis” is more efficient and effective in all directions as compared to existing system. The existing system works manually and don’t have featured like give IP address, generate reports etc. In existing system user can manually goes to every web page and checks content which is time consuming activity. There is not any automatic system available to specific users (agencies) to detect websites related to terrorism. The current system measures the positivity and negativity in content through different algorithms of web mining and data mining and gives the best result. The output was generated in the form of managed reports with graphs.

3.2.3 Accuracy/Reliability

Accuracy refers to the degree of similarity between values which was measured to values which was known. All the functionality should be performed accurate and in case failure

of any module does not cause whole system failure. This system provides the best tool for analysing the websites containing content related to terrorism, and gives best and quick result which means that the degree of accuracy of project is high.

3.2.4 Availability

The Terrorism analysis system is used specifications in which discussed that the availability of system is important factor. System is available and works 24/7 for specific users, this system provides an output of different sections in the form of proper report with graphs. For the availability of this system internet connection is necessary.

3.2.5 Flexibility

The current system is flexible, because it performs multiple functions like extracting content, analysis of content and other specifications. These all functions are used to analyse the content. There is a registration form where users can registered them and after that user move to main page of the system where user can get guideline about the use of system. The user can input domain name and extract textual content from it and analyse it, with all these there are some other characteristics in the system which gives best results to user.

3.2.6 Consistency

Project need hard work to develop and maintenance of project is also not easy. It also develop

users trust on system. If there is no internet connection available to users then it can't work according to their need.

3.2.7 Security Concerns

Extracting and analysing the web content is very useful system for specific users like government agencies or any other users according to their own need. The users can easily to understand how to use the system through simple interface of system and guideline before using it. The users simply registered them and use the system to analyse the websites according to their parameters. Users can individually use the system therefore login authentication is created.

3.2.8 Performance

The current system performs each function is fast. When the current system is tested, the processing time, output time and its response to other functionality is satisfactory and efficient.

4.0 System Modelling

Now in this chapter there is detail discussion on the procedures of evaluation in detail. In this software the user can find the black sites according to their quality of services. Internet connection is needed to open the application main page. Its solve problems of users, and work efficiently and effectively. The field of system analysis is relates to research operation. Many researchers work on this that through these all parameters it is the not so easy tasked. The researcher explains its advantages one by one in their papers. This application is distinctive and work efficiently. The directly access the system through webpage. This system offers the services to specific users, terrorist radicalization can happen at any age, young people in search of a sense of belonging, purpose, and identity may be more vulnerable to violent extremism and terrorist radicalization. This is often compounded by negative perceptions of state authorities, especially of the police and security forces which are perceived to belong to or support 'the other'. State authorities are defending theses violent negative extremism and sacrificing to hold the law and order. The prime reason of developing and implementing this project is to contribute with state authorities to fight against terrorist radicalization.

4.1 Evaluation of Existing Work

Today the Evaluation of the terrorist websites are being monitor manually by specific state authorities and security agencies, that take a more time and of course they can't able to visit and monitor each and every website. This system help in monitoring as well as report to specific authorities which is fast and efficient way to recognize the illegal and unethical(Black websites) and the proposed system work 24/7 active on internet like Botnet malware and visit types on websites on the base of their extension and report to its Botnet Master. This system can be used by the cops and other government officials who work for country security.

4.2 Proposed System

The proposed system is not same as the existing system. The proposed system extracts the contents of webpages using data mining algorithms to extract textual information present on web pages and detect those web pages which are cause for spreading terrorism. In this system, system tracks susceptible IP Address and provides the information to the users. This system detect patterns, keywords and relevant information in unstructured text present in web pages using web mining as well as data mining algorithms.

The back end of this project is SQL server which stores data and other details which is related to this project. There are basic requirement of hardware to run this application. This system is developed in .Net Framework using ASP.NET. This application is online so this application can be accessed by using any device like (Personal Computers, Laptop and with some hand held devices). This system can not specify for international websites with extension like ".gs", ".us" etc. The proposed system diagram of the developed system is shown in the figure 4.1.

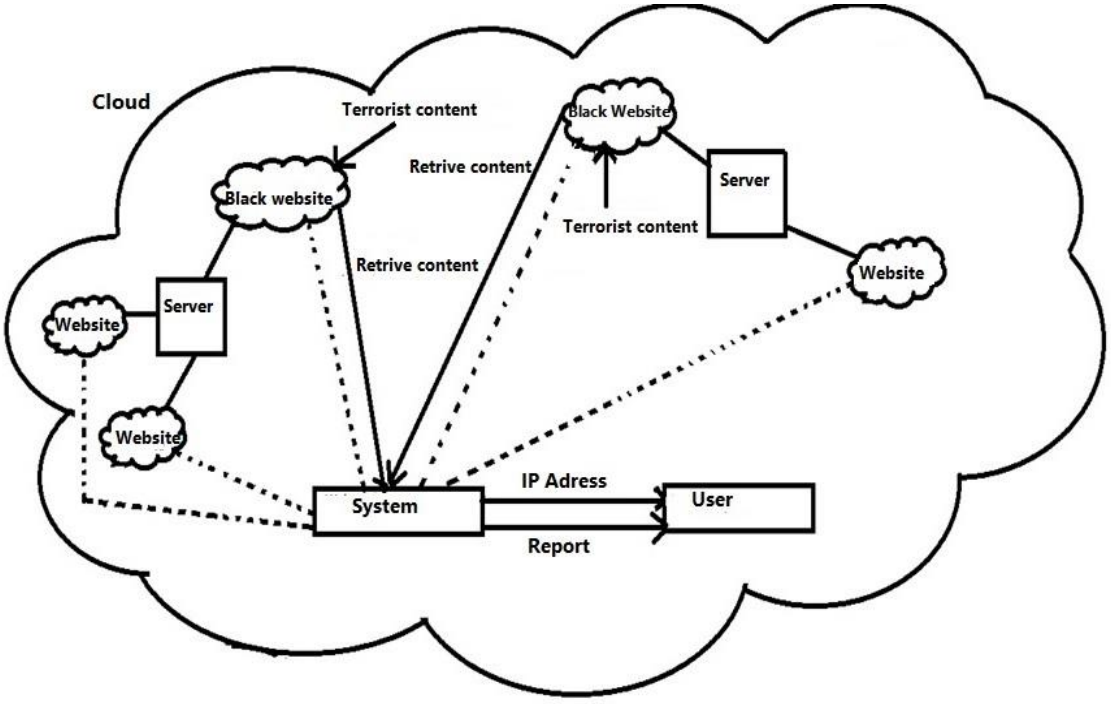

Fig 4.1: Proposed System Diagram

4.3 Use Case

Use cases are an integral part of software engineering that demonstrate how a software as a whole or (preferably) by part interact in a way that helps or benefits the user. It is also used as a way to pin point deficiencies / errors which might have been missed during the development plan. A use case in programming is a depiction of a potential arrangement of communications between a product module and outer specialists, which lead the operators towards something helpful. Use cases are utilized as a product demonstrating method that offers engineers some assistance with determining which elements to actualize, and decide how to effortlessly resolve blunders.

4.3.1 Main Use Case Diagram

As everyone know that use case diagrams are used to show the interaction between user and system. Basically use case tells that how users use the system and how user got interacted with system. Users can use the system as per their own choice to get the best result. In this system user uses a web application for finding terror related content in websites. Also in this system there are seven to eight main section take place to perform task. Firstly the user login to system for accessing the system. After login home page of web site appears where users get welcome note and guideline to use the system. In third section user input domain name of website for extracting textual content from website. In fourth section after extracting content from web page system performs sentiment analysis on extracted content to check the ratio of positivity and negativity in content. In fifth section user can find out IP address of website. In sixth section user can view record of those websites he/she checked. In next section user can manage websites which have content related to terrorism user have also option to block website. In last section user can get report of all checked websites.

The first module of this system is login form. First of all users must register themselves through registration form. For get access to the main menu of the system user must login first and validation checked is applied on the form. The main purpose of registration and login is to give access individuals to system. Login can also be used for security purpose of system.

In second module user input domain name of website. User can put domain name and extract textual content from web pages of website by eliminating tags from it. This can be done through data mining and after this user can do sentiment analysis of extracted data to check whether the content gives negative meaning or a positive, it is done by using database values. In same module user can also find IP address of same website. The third module is view record in which user can view the record of checked website in managed form. The next module is to manage the websites who are recognized as terrorist websites users have option to block them or to add them in category of under monitoring. In the last module user can get final report of checked websites in managed and graphic form. In this way users can get best results as shown in table 4.1.

Table 4.1: Use Case

Use Case ID	**UC1**
Actor:	Actor is person who used this system according to his need.
Purpose:	Extracting web contents through user input by domain name of website. Eliminate unnecessary tags from the web content. Perform sentiment analysis by using database values. Manage websites and generate reports.
Overview:	This application is being used by specific users. These users can extract textual contents from web pages by input domain name of website. This system can save the time of users and help them in finding websites related to terrorism.

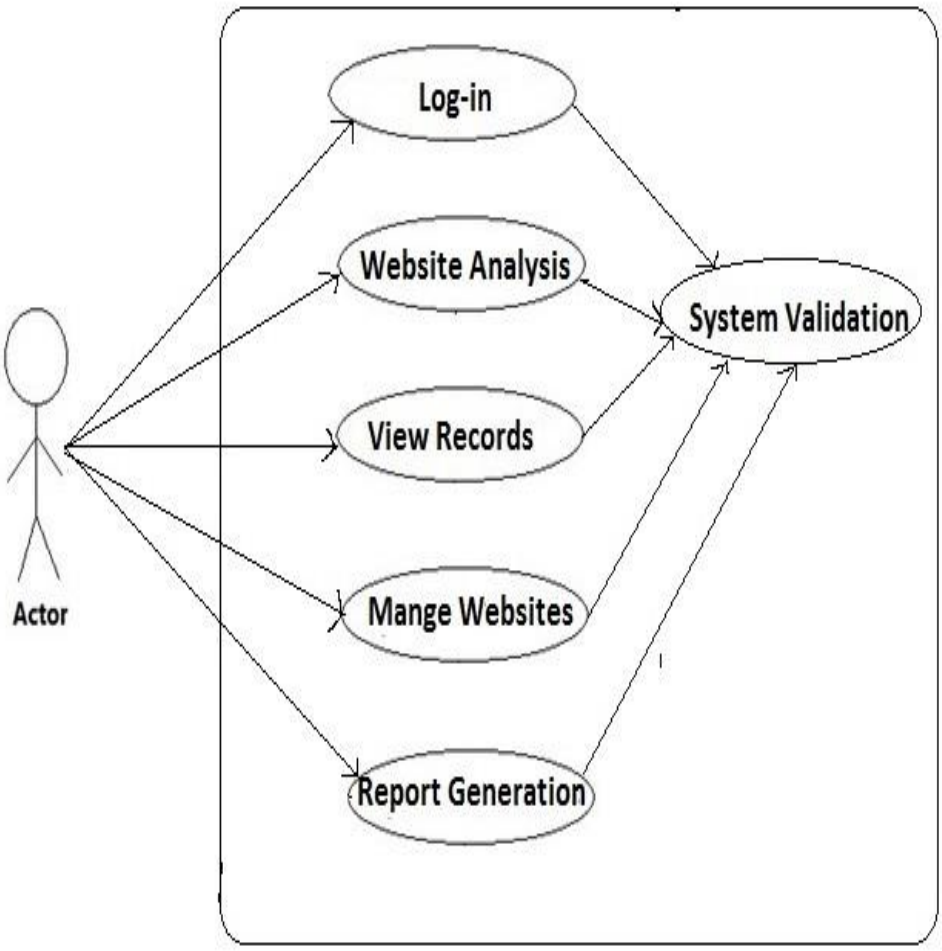

Fig 4.2: Use Case Diagram

4.4 System Sequence Diagram

The result was obtained by interact different commands with each other. By using commands are interacted with each other and result was obtained. The report was generated in order to get a work in meaningful way. A sequence diagram shows the operation and order of processes used in system. The ranging time of objects interacting illustrate in sequence diagram. It also explains the system in logical manner. Users first registered him through registration page by putting information, after login user input domain name of website for extracting textual content from it and do sentiment analysis on it. User can view records of analysed websites and also managed these websites. The sequence diagram also represents the objects on the designed system with each other. The sequence of the current system is shown in Figure 4.3.

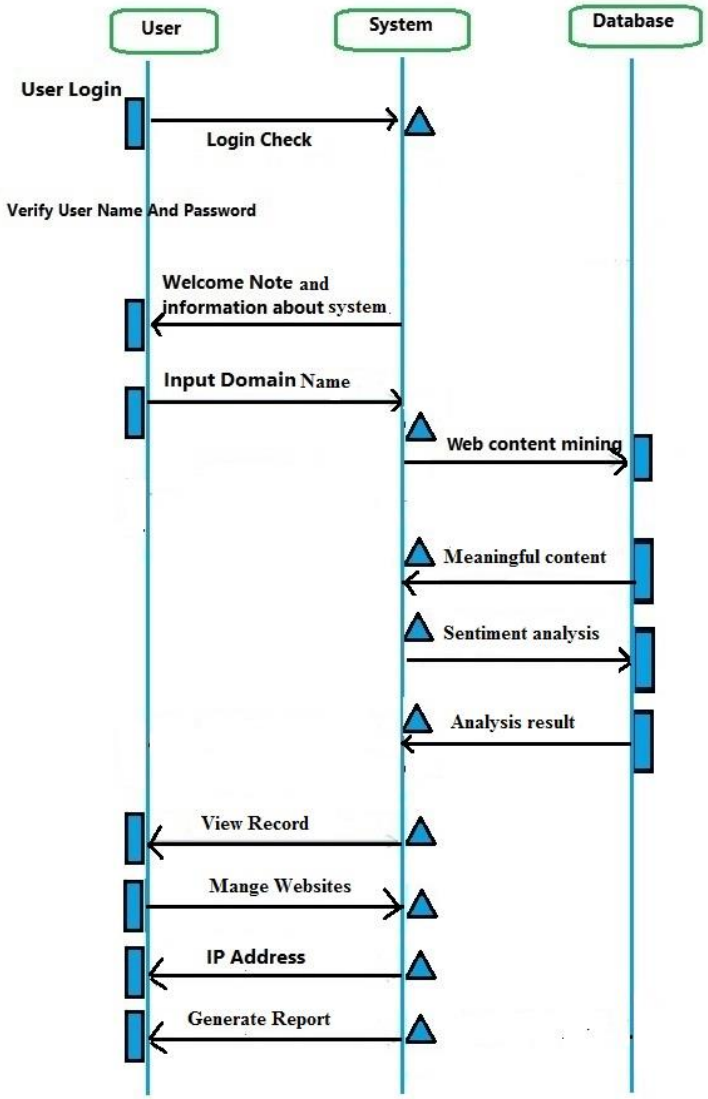

Fig 4.3: Sequence Diagram

4.5 Architecture Diagram

The term architecture is being used for defining the system, in other words an architecture diagram is a start to end representation of the design of a specific system. As this system is based on web and data mining and the architecture diagram shows the components of purposed system. The architecture diagram of this system divided into two parts one represents user's side which consist of different users. And the second side represents a cloud environment which have server at frontend. The front end server is web server which contains a set of websites. When user input domain name of website system use different algorithms of web mining and data mining for extracting textual from websites. The system also find out IP address of these websites by using find IP address using findIpdx algorithm. The architecture of the current system is shown in Figure 4.4.

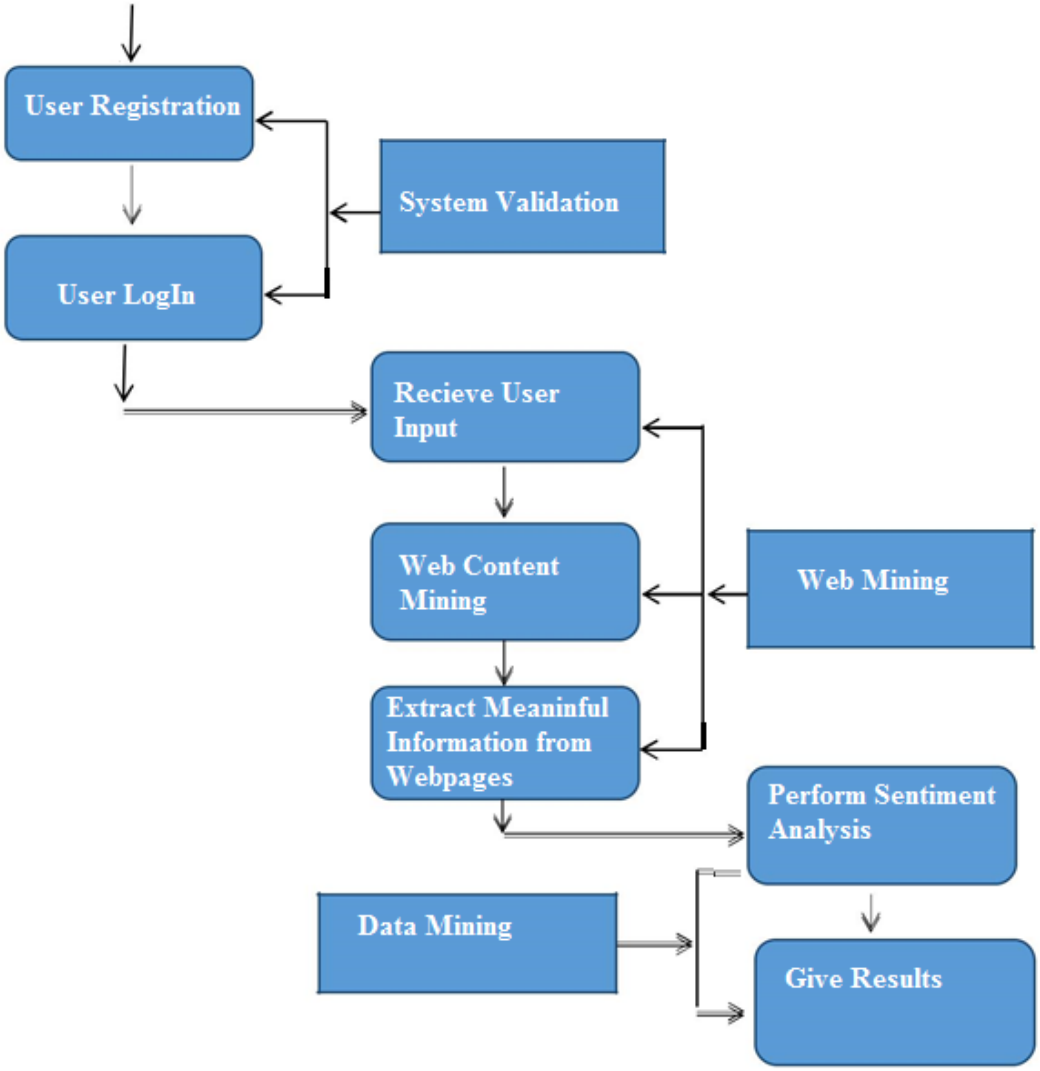

Fig 4.4: Architecture Diagram

4.6 Data Flow Diagram

The data flow diagram of this system is stated as a user login to the system first for that he must registered himself first. After login user moves to main interface of the system where he first goes to website analysis section where he inputs the domain name of the website for extracting content from it and do sentiment analysis from it. User also gets IP address of this website in this section. After this next section is view records in user can view records of checked websites in a managed way. In this section user van manage record by giving status to website according to analysis results. After that user can manage websites by blocking them according to status he gives in previous section. At the end user can get crystal report of all analyse information in managed and graphical form. The flow of the current system is shown in Figure 4.5.

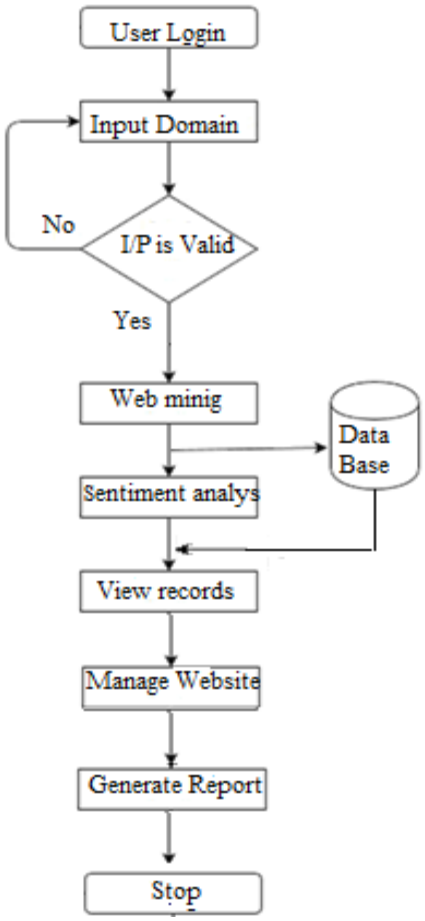

Fig 4.5: Data Flow Diagram

4.7 Development Objective Achieved

Our objective was to find websites which contains content related to terrorism, first of all login module is discuss. In this module firstly users registered their self through sign-up page and after login to system user move to home page of website which contain system. On main page the user get some guide lines about the use of system. The first work user do is he/she inputs the domain name of the website for analysing it from different dimensions. The data mining technique and sentiment analysis technique applies to check that either website contains terror related content or not. The record of searched sites was displayed to user after this and user can also manage websites which contains terror related content. At the end after performing all these steps system generates a report about searched websites to user.

5.0 Introduction

In this chapter the result are discussed in detail by running the system through testing phase. The system is error free and all the features are completed and work properly. The objective of this process is to find out the errors and mistakes and try to overcome these errors/bugs in the way that the efficiency of the system increases. When testing process is completed the second process is analysis process in this process the results of previous process are properly examine and also knows that either this system provides convenience to user or not. And the outputs are examine after this process there is third process called validation process all the results are analyse and try to resolve the bugs/errors and then implement them in the system. A brief discussion on the testing process is below**.**

5.1 Testing and Validation

Software testing, and software validation is the process of checking that a software system meets specifications and that it fulfils its intended purpose. It may also be referred to as software quality control. Also the testing mean is to find out the errors and bugs present in the system and assure that the system should be error free before the deployment. There are some bugs and errors which cannot be found out without testing process. Through testing knowing all about the system and its working. Testing process is not only decreasing the presence of errors in system but it also helps in improvement of the system in different ways like reliability and accuracy of the system and also satisfaction of the user. All the components of the system are check one by one in testing phase and if there are some bugs or errors in the system then these bugs or errors are removed and then this testing process is performed again and this process performs till the system become error free system. The most important of all this is that checks for users validation, user might be perform some actions that the proposed system may not accept like if user enters web URL in the mention websites URL text fields if that happens then the proposed system displaying the error messages that show the message (Please enter only domain name).

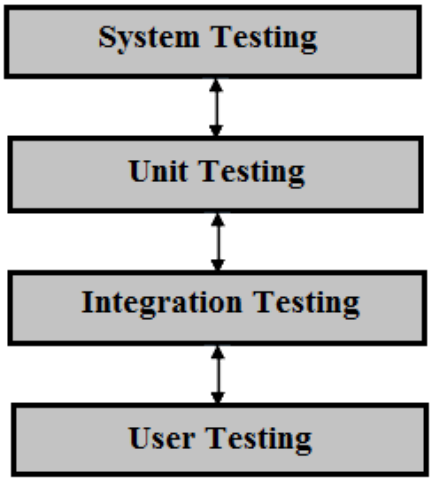

Figure 5.1: Testing Technique

5.1.1 System Testing

The process of testing is completed by using different methodologies and in this case framework testing is adopted. The aim of the system remains same and it doesn't matter which method is used to test the limits of the proposed system in the testing process, it only assure that the system works correctly/properly and there is no bug or error present and also gives us correct results which is wanted. In short testing the whole system through different testing methods to find out the errors and bugs in the system.

5.1.2 Unit Testing

In this process of testing, system is divided in components and then tested one by one. This approach is purely engineering approach which helps in debugging of the each and every component of the system for better and faster performance. Modules are the components of the system and these components of the system are login page, functional module, user module, extracting the web content from the given web site and output are the main components of the system where user can found the results. In this phase all the components of the system are checked one by one properly, and if the bugs or errors are present in any component then in testing process it can be easily find out and it can resolve the errors or bugs in sequence.

5.1.3 Integration Testing

After the completion of the component processes the next process is integration in which developer debugged components integrate means when developer debugged them then it's again tested whether they are working correctly or not. If the errors are found again

then testing process for components performs again and then integrates them again until the system or component became error free. The integrating process is so important for the convenience and satisfaction of user.

5.1.4 User Testing

In user testing process the user tests whole system or application, user tests overall layout and all the parts whether he/she understand it easily or not when the user checked the system then they give their views about the system whether their expectations are fulfilled properly or not. This process is very helpful for the proposed system in present and for the future development. If users are not satisfied with the work and give negative remarks about the system then integrates all those parts that user might not like.

5.2 Test Case Scenarios:

It is basically a documentation that is filled by the tester for certifying the scenario. For testing the Framework, several experiments are performed to test the requirements that are well known. In this process of testing all the cases are tested which are use in the proposed system either the results of all the cases are correct or not. If the output of required results is not satisfactory then the proposed system is again revised for the testing. For checking the functionality of the proposed system there are some test cases which are discussed here which are performed for checking the functionality of the proposed system. Each module of the system is being tested separately in order to check whether it works according to user requirement or not.

5.2.1 TCS 1: Log-in Testing

When users start using system they first login to the system by registering himself. In login testing it is checked that is it works properly or not before system modules starts, the login page opens first for login or registration. The working of these functions is that login testing is checking how system works in this case scenario. When this system is run in debug mode then it is analyse. For login testing the system first checked the users who already registered to the system. The entire query for the “Data Table” login authentication is created in this testing. The system checks the correct format for “Sign-up” if user enters the correct information then he/she redirect to home page. Login is used for the verification of different users. After login the user get access to system. All users access the system individually.

It gives proper information to users about the system and how to use it after “Sign-in” to the system. Also the transition of screen from the Login screen to next menu/main page must be made. Test drivers used for this module are Microsoft Visual Studio it is a language in which system was developed. To use this application there is no need of specific device this application can run on pc’s, laptops and mobile devices without internet connection this system not run. Objective of testing is achieved login page is tested and it is free from any type of errors and bugs, never be crash in running state. And the transformation from login page to home page is reliable and fast. The home page can only be access after login.

5.2.2 TCS 2: Main Screen Testing

The main screen appears after user successfully log- in to the system. The main screen module includes all the functionalities like extraction and analysis of the content of website and show record of checked websites, manages websites and report generation. If user moves towards website analysis section of the system then he/she can extract textual content from websites and Perform sentiment analysis on it. The user just inputs domain name of website and further moves towards next step. The execution procedure of this module is explain. In Main screen module tests

the proper functionality of URL authentication and search contents nature. The main screen modules are accessed only after user login. The system checked the correct working of the screen

by “On click” and “On get” source button to validate the URL and extract contents of specific website. User choose get source button then the application indirectly Extract source code of specific website contents, after validation of URL and Extraction the user may click on analysis button for further operation . In this module the screen appeared or having options of get source button once URL is entered by taking input from user, he/she clicked on get source button to validate and Extract contents of specific website. Objective of this is to get source button must validate the URL. And extract contents: The Contents must be in correct format. Test drivers are Microsoft Visual Studio it is a language in which system is develop. To use this application there is no need of specific device this application can run on pc’s, laptops and mobile devices without internet connection this system not run. Objective of testing is achieved to get source button is ready to taking accurate and validate the URL. And the Contents must be extracted before

analyses and it must be in correct format and main screen also contains some other parameters. The result is display in the rich text box.

5.2.3 TCS 3: Website Analysis Testing

When users receive contents through the activation of get source function after the authentication and accessibility of specific websites then the user click on website analysis button present on right below of the rich text box of the system. After user gets a filled rich text box where he/she click on the analyse button to analyse the Contents. The execution for this module is that once System extract the contents of website through the domain name input by the user, Now it is time to analyse these contents User click button name ”Analyse Output For Sentiment Analysis” and the method of data mining and sentiment analysis execute successfully provide polarity confidence of web contents. Entering user manual contents directly in the rich textbox and analysing check the algorithms is working correct as well as system is analysing the data base values with the contents show that data connectivity is active and working correct. Basically in this module the user can extract textual information from web pages. The extracted text is free from tag. Sentiment analysis should perform by the system to check the polarity of content. Also find out the ratio between negative sentences and positive sentences. The objective of this was that the system is able to extract web contents. System is also capable to remove unnecessary tags from contents and also recognize polarity confidence of web contents. Test drivers are Microsoft Visual Studio it is a language in which system is develop. To use this application there is no need of specific device this application can run on pc’s, laptops and mobile devices without internet connection this system not run.

5.2.4 TCS 4: View Records Testing

Polarity Confidence, Subjectivity Confidence, IP address and word match ratio can be find through the system now system store these attributes along with their values match ratio is find out by the system now system store these attributes along with their values in data base and generate or display Grid view of these attributes to ensure user that data has be collected and store in data base. By using testing techniques the functions are test step by step. Execution procedure for this module is that the terrorism analysis parameters are tested by system to check out the correct result and show in the grid view. The data is correctly store in data bases according to their fields. All the contents of websites are match with the stored data variables and text to validate the required data. Users can

easily view the record of websites which are analysed by system. In view records user have 4 columns “searched words” which contains the list of words which are extracted from web pages, “web URL” it contains the list of URL’s of checked websites, “match ratio” in this match ratio between negative and positive words shown and “status” in status users have three options for giving status to websites which are black listed, pending and not pending. Objective of this module is that it shows the record of all checked websites in meaningful way. And the result shown in this is meaningful.

Test drivers used for this modules are Microsoft visual studio in which this system is developed and there is no need of any specific device, the user use this system in all devices like Pc’s, laptops or smart phones etc. Objective of testing is achieved by showing record to users can increase the efficiency of the system can be increase. This provides easiness to user in analysing the websites for the next time.

5.2.5 TCS 5: Manage Websites Testing

After viewing the record of checked websites now it is easy for user to manage websites. By checking the record user can block the websites which are in black listed category and put other websites in under monitor category which are in pending status. Execution procedure for this module is that in this section there is two columns one is URL which contains URL’s of checked websites and second is Manage websites which contains an option of block websites with checkbox. If user wants to block website he/she can tick the checkbox and if not then checkbox remain unchecked. The description of this module is that users can manage websites by reviewing record of websites which are analysed by system. He has option to block them or put them in under monitor category. The objective is by managing the websites the efficiency of the system is increased. And this can help in overcoming black website present on internet. Microsoft visual

studio is used as test driver. And there is no need of any specific device, the user use this system in all devices like Pc’s, laptops or smart phones etc. Objective of testing is achieved by managing the websites can increase the efficiency of the system can be increase. This provides easiness to user in analysing the websites for the next time.

5.2.5 TCS 6: Report Generation

In this Module system can generate the reports of all the record. The generated report is in managed form and also shows results in graphs form. To generate the report crystal report technique used by taking data from different tables of database. Execution

procedure for this module is that report generation test is checking how system works in this scenario. Crystal report viewer was drag to the web form. From view records form new item add which is crystal report by selecting its connection was established with database. The values stored in tables of database show in report viewer through query. Description is that in this module system can generate the reports of all the record. The generated report is in managed form and also shows results in graphs form. To generate the report crystal report technique used by taking data from different tables of database. Objectives are that in this module the different records are shown on the web browser for user they select (click) on generate report by their own choice if he/she wants to get. And on the basis of parameters selection the system move towards the solutions.

Test drivers are Microsoft Visual studio. And there is no need for any specific device, the user use this system in all devices like Pc's, laptops or smart phones etc.

5.3 Testing Results and Analysis

The basic goal of this web application is to analyse the websites contains the content related to terrorism. This application is specifically design for government agencies or any other special user. This system is easy to use and very useful for security agencies to cop the websites related to terrorism. It helps users to analyse websites automatically and get best results. This system also shows the results of checked websites in the form of graph. The users just input the domain name of website, than system extract textual content from it and match it with negative and positive words and sentences present in database to find match ratio. It also performs sentiment analysis of content. System also finds out IP address of websites. The design of the system is simple which is easily understandable by the users. Testing of the system is successful at the system level, since it generates result in few minutes.

6.0 Introduction

This section consists of conclusion and future work. In this chapter detail explanation of the amount of work that which is done and the amount of work that can be possible in future.

6.1 Conclusion

This system is design for the Government agencies, Security forces or for those who related with the field of security so it is a great security tool. It is latest addition to the security section. It designs to facilitate the security agencies to overcome the terrorist related content present on internet and it acts like security system at work. Finally it is conclude that this system which is developed in this project probably the best system for analysing terrorist related websites on internet. This system has following characteristics and qualities.

It is loaded with instinctual menu. It has capability to detect the terrorist related content automatically or manually as mention terrorist related content means any article that support and encourage hate rate, terror, adult and any other dangerous activity which panic or encourage people to attempt any unethical or illegal activity. System is design to select the specific functions to attend easily results as required ,It is developed using some of the most upgraded new era of algorithms like Artificial intelligence algorithms, that's not only find the required contents but find the sentiment of that contents according to the given contexts. System is simple to used first of all main menu page generate simple description of a System with login and Sign in buttons, As User sign In new page appears and there is different click-to-call buttons that let user to perform different functionalities such as analyse the websites to find hate rate or terror or adult etc. Contents in specific region, functionality to store searched websites and statues of blocked and unblock. All the functions are working properly or in proficient manner

6.2 Limitations

This project is limited since it is not in support of all domains of internet. This system is only specify for websites with extension like ".pk" and ".com". The system does not extract content in audio, video or pictorial form It only mine textual content from web pages by using data mining techniques. Without the permission of user system itself unable to block websites. The system is not going to use web service. The system analyse the websites which contains content in english language other than English language it is

unable to analyse. The current system is only access by specific users like government agencies etc. The system is unable to analyse code words black websites.

6.3 Future Work

In the existing system, here are some limitations since of project have a few deficiencies. One of the constraints is that it does not mine information from multiple domains like it does not search the web pages which have domain other than ".com" and ".pk" and it only mines "textual information". Keeping in view the limitations the system has, in future these limitations is overcome and system can be improved by adding the following features, like it can also mine "audio", "video" and "pictorial" information from web pages. This system can mine information from other domains like ".org" ".edu" etc. This system achieved 65% accuracy. In future, the system also perform better by increasing number of keywords which is use and also use code words in system which are used by terrorist organisation.