An Efficient Way of Web Data Mining to Detect Online Spread of Terrorism

Akash Mishra[1], Adarsh Pandey[2]
[1,2]*Department of Electronic Communication Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad*

Abstract **- Terrorism has grown its roots quite deep in certain parts of the world. With increasing terrorist activities, it has become important to curb terrorism and stop its spread before a certain time. So as identified internet is a major source of spreading terrorism through speeches and videos. So here we propose an efficient web data mining system to detect such web properties and flag them automatically for human review. With development in technology, internet has become a medium of spreading terrorism through speeches and videos. Terrorist organizations use the medium of the internet to harm and defame individuals and also promote terrorist activities through web pages that force people to join terrorist organizations and commit crimes on the behalf of those organizations. To reduce the human effort, we implement the system which detects terrorist groups in social media. To implement this idea we are in need of a lot of human effort to gather the information and find out the terrorist groups who are involved in. It paves the way to reduce the human effort.**

I.INTRODUCTION

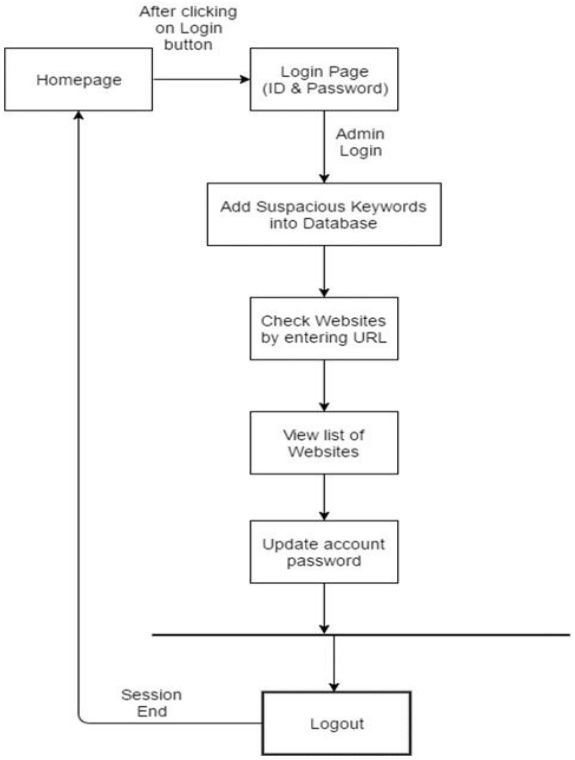
Terrorist organizations are using the internet to spread their propaganda and radicalize youth online and encourage them to commit terrorist activities. To reduce the online footprint of such harmful websites we need to create a system which detects specific keywords in that particular website and if those keywords are found then that website should be blacklisted. Data mining as well as web mining are used together at times for efficient system development. The major source of spreading out terrorism is internet through media such as images, speeches, videos and url's. Terrorism Groups use web to persuade people and youth and it also influence individuals to involve in terrorist activities through alluring web pages which inspire people and youth to involve in terrorist organization Web mining also consists of text mining methodologies that allow us to scan and extract useful content from unstructured data. Text mining allows us to detect patterns, keywords and relevant information in unstructured texts.

II. PROJECT PROPOSAL

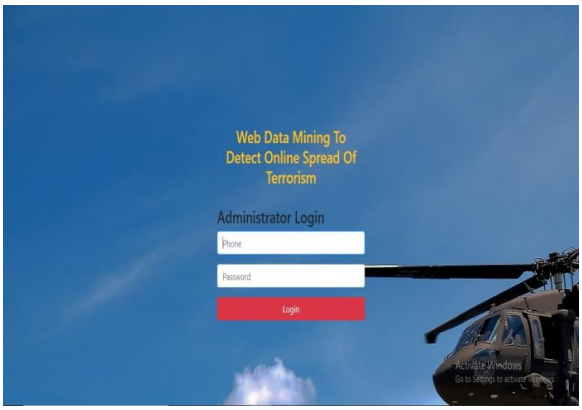
Our system will detect patterns, keywords and relevant information in unstructured texts in web page using web mining as well as data mining. Our system will mine webpage using web mining algorithm to mine textual information on web pages and detect those web pages that are relevant to terrorism. Data mining as well as web mining is used together at times for efficient result. This research have three steps- First, According to the target event, they have numerous activities based on crawl. Second, they estimate location of events and propose probabilistic activities to extract events. Finally, Alert reporting system is developed that takes earthquakes from websites and message is sent the registered user.

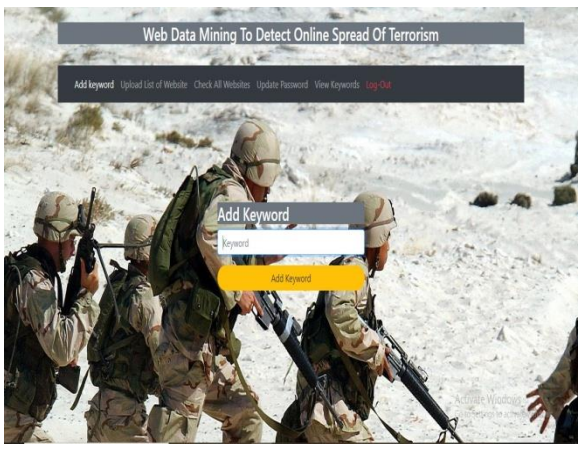
It is a method of partitioning meaningful group of clusters from data objects. After the clusters have a same similarity based on characteristics or vice-versa. Clustering is an unsupervised division of patterns(data items)which is unlabelled and there is no pre-defined labels Which is associated with set of objects .This results in compact of high sets by tiny cluster centroids. Clustering application includes text document retrieved, classification of patterns, mining such as text and data and segmentation of image.

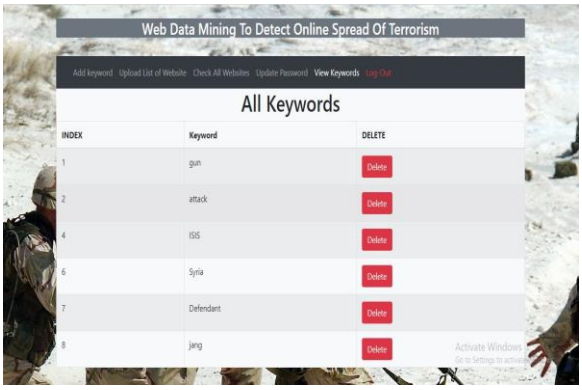
III. BLOCK DIAGRAM



IV. IMPLIMENTATION







V. HARDWARE AND SOFTWARE REQUIREMENTS

The Project is loaded in Visual Studio 2010. We used Visual Studio for Design and coding of project. Created and maintained all databases into SQL Server 2008, in that we create tables, write query for store data or record of project.

Hardware Requirement:
Processor – Dual Core
Hard Disk – 50 GB
Memory – 1GB RAM
Internet Connection

Software Requirement:
Windows
Sql
Visual studio 2010

VI. METHODOLOGY

We use web mining algorithms to mine textual information on web pages and detect their relevancy to terrorism. Websites created in different platform can be tracked using this application. This system will check web pages whether a webpage is promoting terrorism. Moreover, they as a rule overlook the unmistakable between the misclassification article costs having a place with various classes. All those assumption are totally wrong while they involving with terrorist detection on the internet. The monitored population is unbalanced so that entire population of internet users is actually close to low, that is zero This system will classify the web pages into various categories and sort them appropriately. There are two features used in this system that is data mining and web mining. System will help the cops to easily track the susceptible community who are held in terrorism.

Website will have following characteristics:
Load Balancing: Since the system will be available only the admin logs in the amount of load on server will be limited to time period of admin access.

Easy Accessibility: Records can be easily accessed and store and other information respectively.

User Friendly: The Website will be giving a very user-friendly approach for all user.

Efficient and reliable: Maintaining the all secured and database on the server which will be accessible according the user requirement without any maintenance cost will be a very efficient as compared to storing all the customer data on the spreadsheet or in physically in the record books.

Easy maintenance: Web Data Mining for Terrorism Analysis website is design as easy way. So maintenance is also easy.

Modules and their Descriptions
The system comprises of 5 modules as follows:
Login: Here, the admin or the authorized person need to enter the login credentials.
Add Keywords After successful login, admin can add various keyword which specifies terrorism.
Check Website Here, admin can add multiple URL's to scan the website for any suspicious word.
View/Check all Websites All the URL's which are entered by the admin are listed here. Can check the suspicious keyword.
Update Password System allows admin or the authorized person to update their password.

Applications
We use web mining algorithms to mine textual information on web pages and detect their relevancy to terrorism.
Websites created in different platform can be tracked using this application. This system will check web pages whether a webpage is promoting terrorism.
This system will classify the web pages into various categories and sort them appropriately. There are two features used in this system that is data mining and web mining.
System will track web pages that are more susceptible to terrorism and will report IP Address to the user who is using the system.
This System are used only by the government officials who work for country security. System will help the cops to easily track the susceptible community who are held in terrorism.

VII. CONCLUSION

To curb the menace of terrorism and to destroy the online presence of dangerous terrorist organizations like ISIS and other radicalization websites. We need a proper system to detect and terminate websites which are spreading harmful content used to radicalizing youth and helpless people.

REFERENCES

[1] Dietmar Jannach and Simon Fischer, Recommendation-based Modeling Support for Data Mining Processes, Germany and Simon Fischer Rapid Miner GmbH, Germany, Proceedings of the 8th ACM Conference on Recommender system, pages 337-340, October 2014.

Books:
[1] Jiawei Han Micheline Kamber Jian Pei, Data Mining: Concepts and Techniques 3rd Edition (22nd June2011)

Proceedings Papers:
[1] T.Sunil Kumar, Dr.K. Suvarchala, "A Study: Web Data Mining Challenges and Applications for Information Extraction", IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661, ISBN: 2278-8727 Volume 7, Issue 3(Nov.-Dec. 2012),PP 24-29
[2] Syed Ahsan, Abad Shah," Data Mining, Semantic Web and Advanced Information Technologies for Fighting Terrorism" Research paper from Department of Computer Science (Nov.-Dec. 2009).
[3] Robert Grossman, Simon Kasif, Reagan Moore, David Rocke, and Jeff Ullman," Data