IJARIIT INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 2)
Available online at: www.ijariit.com

Online terrorist detection system

S. Aishwarya
aishuk1105@gmail.com
SRM Institute of Science and Technology, Chennai, Tamil Nadu

A. Janani
janu.ajay1998@gmail.com
SRM Institute of Science and Technology, Chennai, Tamil Nadu

Manepally Alekhya
alludad19@gmail.com
SRM Institute of Science and Technology, Chennai, Tamil Nadu

ABSTRACT

Data mining is the extraction of pattern and knowledge from a large amount of data. This concept can be utilized in various real-life applications. One such congenial application is the online terrorist detection system. With expanding terrorist activities, it has turned out to be essential to control and stop its spread before a specific time. Internet is a noteworthy wellspring of spreading fear based oppression through videos and speeches. They use websites to mentally program people that would move vulnerable individuals to join terrorist associations. So here we propose an effective software to recognize such web properties and ban them naturally for human survival. In present situation utilization of the web is in the blast. Be that as it may, each coin has opposite sides, moreover, the utilization of web is valuable just as hurtful to individual. Late days numerous fear assaults were there on the net. Such fear related exercises are risky for people groups, association, and nations. The fear-based oppressor is utilizing the web to spread dread and structure psychological militant gatherings. By utilizing web they effectively do likewise. To trade data Internet foundation is utilized by various Terrorist cells and they enlist new individuals and supporters. To deal with such circumstance we propose a framework called ATDS.

Keywords*— Data mining, Text mining, Web mining, Field extraction, Clustering, Word extraction, Terrorist trend detection, data mining, User modeling, Anomaly detection, Log mining, Vector generator, Threshold, Clustering, ATDS (Advanced Terror Detecting System)*

1. INTRODUCTION

Terrorist associations are utilizing the website to spread their propaganda and radicalize youth on the online and urge them to commit terrorist activities. To decrease the online spread of such destructive sites, we have to create a software which recognizes such harmful phrases in that specific site and if they are being detected then that site has to be boycotted. Data mining is a technique used to mine out patterns of useful data from large datasets. Web mining comprises of content mining strategies that enable us to output and concentrate valuable data from unstructured information.

Text mining permits us to recognize examples, catchphrases and significant data in unstructured writings. Data mining algorithms are efficient and manipulate organized data. Web mining algorithm is used to mine unorganized data from internet. Websites are made in various platforms and hence it is difficult to be read by one algorithm. In order to solve this problem, web pages are created using HTML. Since HTML is used, web mining is done effectively, hence web mining is designed in a manner to mine textual information on web pages and check if they may be promoting terrorism.

Our software will recognize suspicious words and pertinent data in unstructured message in web pages utilizing web mining just as information mining. Our framework will mine pages utilizing web mining calculation to mine literary data on pages and distinguish those sites that are significant to a terrorist. Data mining just as web mining is utilized together now and again for a productive outcome.

2. LITERATURE SURVEY

Online Content information is the most widely recognized substance type on the net with regards to the creator's sentiment. As of late, after the advancement of remote web and cell phone gadgets, iPhones the measure of information on the web is significantly expanding with no compel to time or area. In this paper, we proposed the technique for separating the words from record names as Word Net Chain of command [1]. This strategy was tried with the inspected New York Times articles by questioning four particular words from four distinct territories. Exploratory outcomes demonstrate our proposed strategy successfully extricates setting words from the content and distinguishes psychological oppression related records. The content investigation is utilized to find obscure, legitimate examples and connections in vast informational collections. Indeed, even content examination has an extraordinary potential for recognizing obscure content archives, there is a constraint that human composed language is as yet confused for the machine to comprehend semantic implications of it [1].

The learning Typical– Fear based oppressor Conduct some portion of the strategy characterizes and speaks to the run of the mill conduct of psychological oppressor clients dependent on the substance of their Internet exercises. It is expected that it is conceivable to gather Website pages from fear related destinations. The substance of the gathered pages is the contribution to the Vector Generator module that changes over the pages into vectors of weighted terms (each page is changed over to one vector). The vectors are put away for future preparing in the Vector of Fear mongers Exchanges DB. The Grouping module gets to the gathered vectors and performs unsupervised bunching bringing about n clusters speaking to the ordinary themes seen by psychological militant clients [2].

One noteworthy issue of today is the portrayal of the printed substance of Site pages. All the more explicitly, there is a need to speak to the substance of dread related pages as against the substance of a right now gotten to the page so as to productively figure the likeness between them. This examination will utilize the vector-space show regularly utilized in Data Recovery applications for speaking to fear mongers' interests and each gotten to the Site page.

3. PROPOSED SYSTEM

The online terrorist detection system is aimed at detecting online access to abnormal content that could embrace terrorist-generated websites by analyzing the content of knowledge accessed by the net users. This operates in two modes: coaching mode and detection mode.

Within the coaching mode, it determines the everyday interests of a pre-specified cluster of users by processing the web pages accessed by these users over time.

Within the detection mode, it performs a periodic observance of the network traffic generated by the monitored cluster, analyzes the content of the accessed web content, and gives an alarm if the accessed information isn't among the typical interests of that cluster and almost like the terrorist interests. This can be represented systematically by the following system architecture.

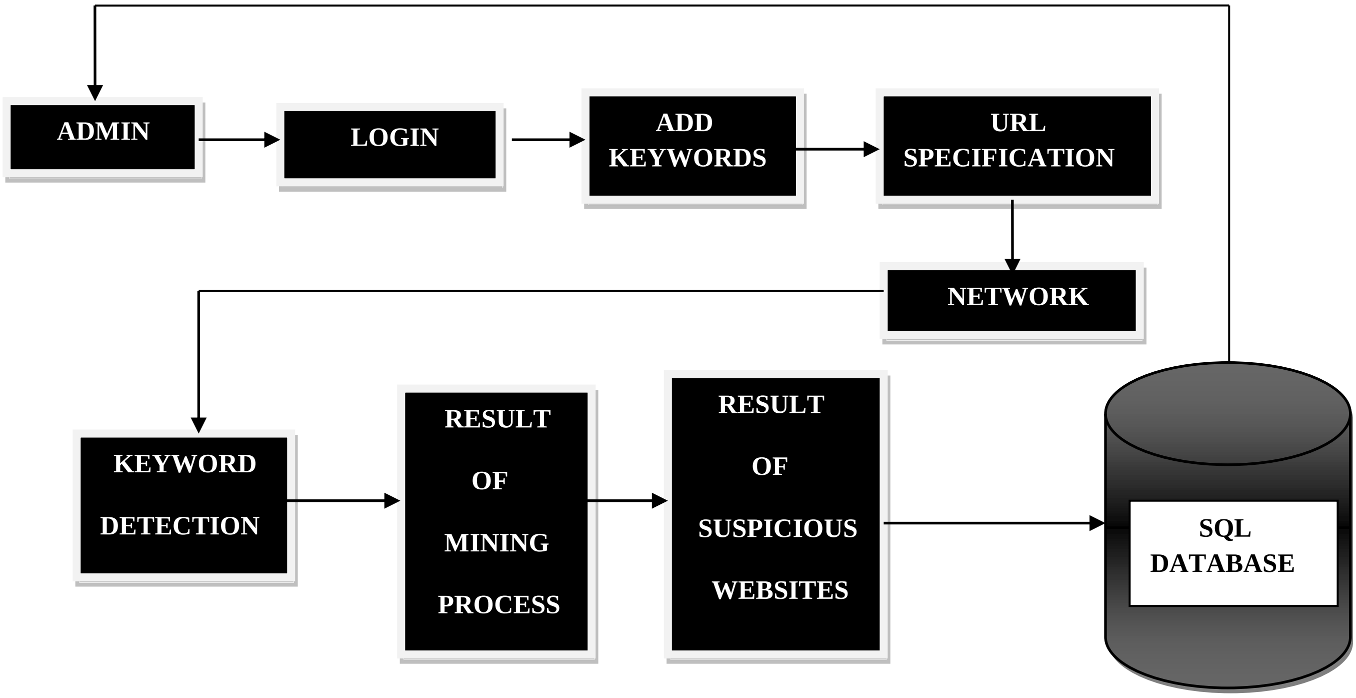


Fig. 1: System architecture

In the first step, the admin should log in using user id and password. If the admin is new, then the new admin should register the user id and password to access further. After the logging in the process the admin can add suspicious keywords and the websites to the database. In the next process, it includes the detection of suspicious keywords from the user's browser history, if a user is found to be using any kind of suspicious keywords or websites continuously, then the particular user's IP address will be mined and get stored in a database, which will be helpful to detect the particular user and stop the spread of terrorism.

Our software consists of two modules, the first module is updating the information and the second module is the detection of suspicious user.

3.1 Updating the information

In this module, first the admin login into the software by entering user id and password this process is carried on the bases of the algorithm of field extraction, then the suspicious words and keywords are clustered based on the similarities through the algorithm of clustering. Clustering is the process of portioning data into groups of clusters. Data with similar characteristics are clustered into a group. Clustering represents a large set of data into small centroids. The clustered data will be stored in a database, to suspect the suspicious user.

3.2 Detection of suspicious user's

In this module, the clustered suspicious words and websites will be checked with the user's browser history and detected. The detection process is done through an algorithm of word extraction. The user's browser history is checked through the data mining process. If any kind of suspicious words or website is detected to be overused and matches the cluster stored in the database, then the user's IP address will be reported to the admin and notified through an alarm.

4. RESULT AND ANALYSIS

Initially, we run the program to get the login page where the cybersecurity officials enter their details to get access to the software. Using this they can add new keywords or websites. And also they can update already existing data and get connected with the database. They can also detect such websites to block them.

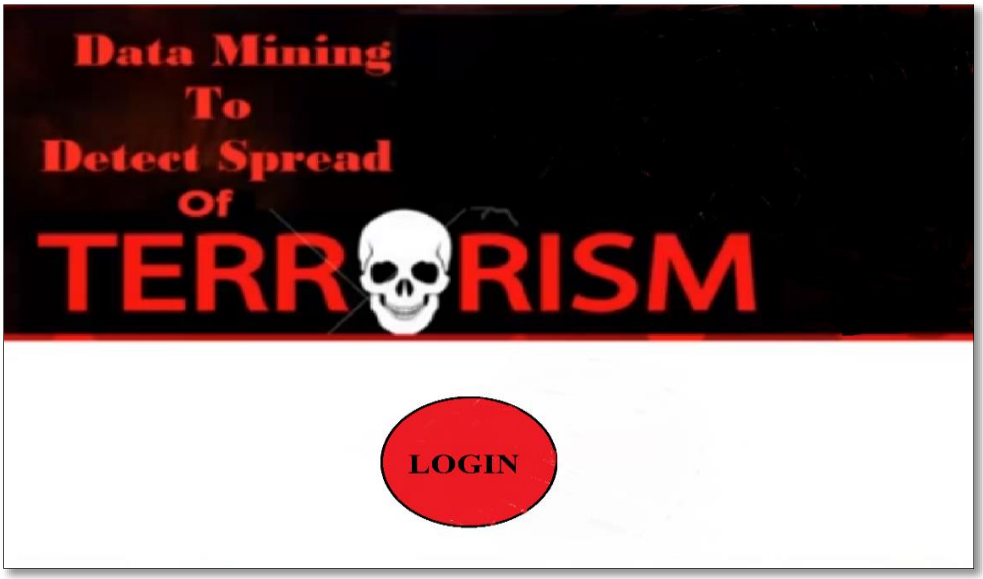
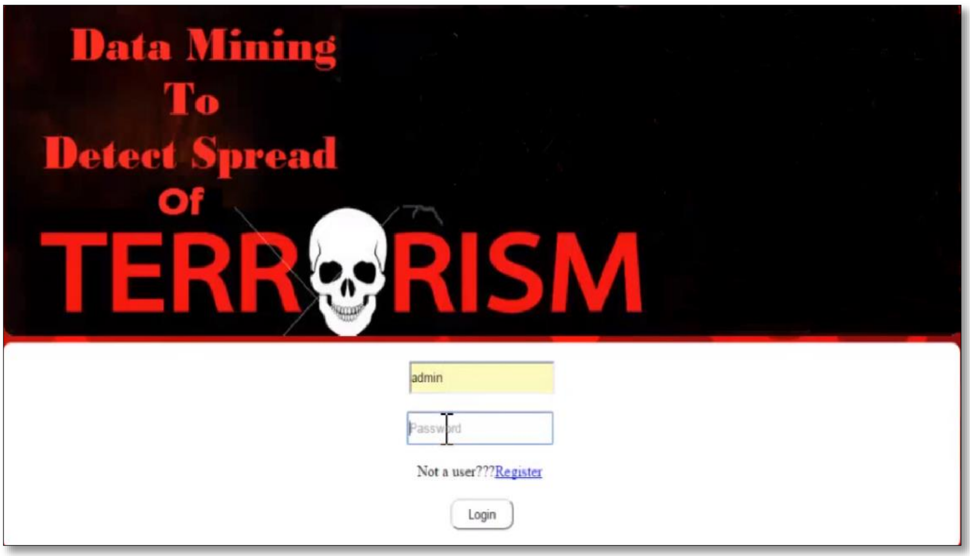


Fig. 2: Login page

Figure 2 depicts the login page where the cybersecurity officials can select it to enter their details.

Figure 3 depicts the user details page. Here a new user can register and create an account provided the user must have authentication to do so. The user can directly login if he/she has already registered. This page consists of two things. One is the user id and the other is the password. These details are kept highly confidential for information security.



admin

Fig. 3: User details page

In figure 4, there are five sections. They add keywords, check the website, and check all websites, update password, view feedback and log out. Add keywords is used to add keywords to the already existing database. Check all website is used to scan the browser history for any suspicious websites. Check website is to add suspicious to the database. View feedback is to get feedback from the authorities. Logout is to log off one's account.

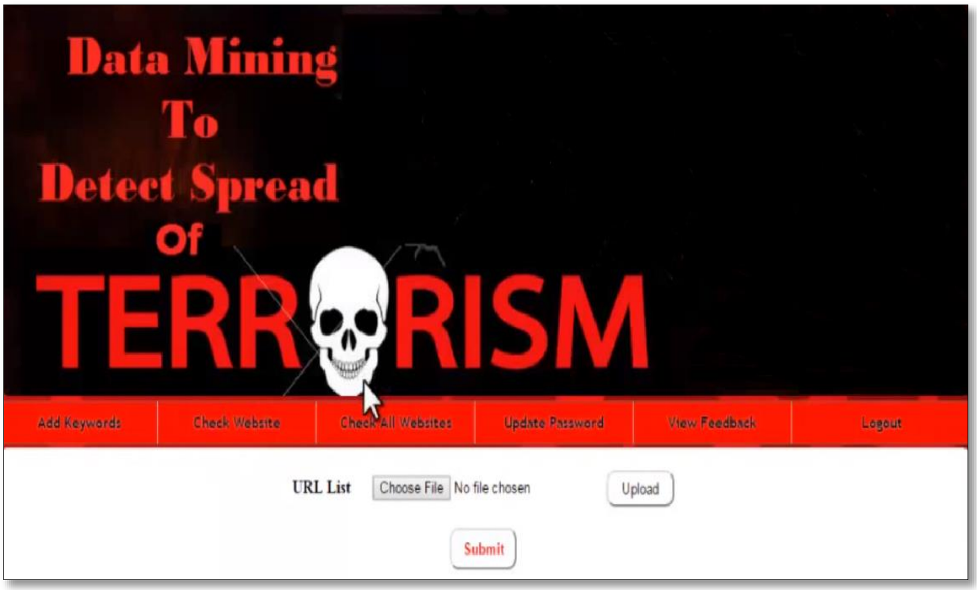


Fig. 4: Sections

5. CONCLUSION AND FUTURE ENHANCEMENT

To check the hazard of terrorist and to crush the online nearness of unsafe psychological militant associations like ISIS and other radicalization sites, we need a legitimate framework to identify and end sites which are spreading unsafe substance used to radicalizing youth and defenceless individuals. In this way, web mining method can be utilized for recognizing and staying away from fear dangers brought about by fear mongers everywhere throughout the world. Information mining furthermore, web information mining advancements will have a noteworthy sway on counter-psychological oppression. As we are seeing, one of the significant worries of our country today is to distinguish and counteract psychological militant assaults. This is additionally turning into the objective of numerous countries on the planet. We have to look at the different information mining and web mining advances and perceive how they can be adjusted for counter-psychological warfare.

In this paper, we have taken only the suspicious words and websites. But in future, we can consider suspicious videos, advertisements and so on related to terrorism. This will enable the user irrespective of the age to safeguard himself/herself from getting influenced to terrorism.

6. REFERENCES

[1] Dongjin Choi, Byeongkyu Ko, Heesun Kim, Pankoo Kim,” Text analysis for detecting terrorism-related articles on the web”, Journal of Network and Computer Applications 38 (2014) 1621.

[2] Mohammad Javad Hosseinpour, Mohammad Nabi Omidvar,” Detecting Terror-Related Activities on the Web with Using Data Mining Techniques”, 2009 Second International Conference on Computer and Electrical Engineering.