

# Web Mining To Detect Online Spread of Terrorism Activity

Shrikant A. Shinde<sup>1</sup>, Hrishikesh Wadge<sup>2</sup>, Rohit Pitekar<sup>3</sup>, Saurabh Warade<sup>4</sup>

<sup>1,2,3,4</sup> Dept of Computer Technology

<sup>1,2,3,4</sup> Sinhgad Institute of Technology and Science, Narhe,  
Pune, Maharashtra, India

**Abstract-** In Computer Science, the term Web mining is a rapidly growing research area. It consists of Web usage mining, Web structure mining, and Web content mining. Web usage mining refers to the discovery of user access patterns from Web usage logs. Web structure mining tries to discover useful knowledge from the structure of hyperlinks. Web content mining aims to extract/mine useful information or knowledge from web page contents. Web mining techniques can be used for detecting and avoiding terror threats caused by terrorists all over the world. In the recent times, terrorism has grown in an exponential manner in certain parts of the world. This enormous growth in terrorist activities has made it important to stop terrorism and prevent its spread before it causes damage to human life or property. 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. Web mining and data mining are used simultaneously for the purpose of efficient system development. Web mining even consists of many different text mining methods that can be helpful to scan and extract relevant data from unstructured data. Text mining is very helpful in detecting various patterns, keywords, and significant information in unstructured texts. Data mining and web mining systems are used for mining from text widely. Data mining algorithms are used to manage organized data sets and web mining algorithms can be helpful in mining.

**Keywords-** Data mining System architectures, Data mining application, Dataset, Crime, Terrorism, Warehouse, Knowledge Discovery Database.

## I. INTRODUCTION

Terrorist organizations are using the internet to spread their propaganda and radicalize youth online and encourage them to commit terrorist activities. In order to minimize the online presence of such harmful websites we need to devise a system which detects specific keywords in a particular website. The website should be flagged

inappropriate if the keywords are found for efficient system development. Data mining consists of text mining methods that help us to scan and extract useful content from unstructured data. Text mining helps us to detect keywords, patterns and important information from unstructured texts. Hence, here we plan to implement an efficient web data mining system to detect such web properties and flag them for further human review. Data mining is a technique used to extract patterns of relevant data from large data sets and gain maximum insights to the obtained results. Web mining as well as data mining are used simultaneously for efficient system development. Terrorism is widely spread and grown deeply in some areas all around the world. It is very important to control terrorism and to stop it before spreading before some certain period of time. The major source of spreading out terrorism is internet through media such as images, speeches, videos and url's. Terrorist 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. Thus, the proposed and efficient techniques by using segmentation technique. It flags those activities automatically for human review. It is difficult to study for single algorithm in websites because the websites created in various platforms use different data structures. 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. Web mining and data mining are used simultaneously for the purpose of efficient system development. Web mining even consists of many different text mining methods that can be helpful to scan and extract relevant data from unstructured data.

## II. RELATED WORK

[1.]AakashNegandhi et al. applied various machine learning algorithms in "Detect Online Spread of Terrorism Using Data Mining" to mine textual information on web pages and detect their relevancy to terrorism.

[2.]Chen, H. et al. used the features of sentiment analysis to segregate the words of a web page, classify them and assert a score to each word in "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums."

[3.] Fawad Ali et al. studied various methods by which textual data can be fetched and scanned and executed them to counter Terrorism on Online Social Networks using web mining techniques. [4.] Naseema Begum et al. classified the web pages into various categories and sorted them appropriately. There are two features used in this system that are data mining and web mining.

[5.] T.Anand et al. implemented Data mining as well as web mining are used together at times for efficient system development. System will track web pages that are more susceptible to terrorism and will report IP Address to the user who is using the system. In earlier system, the terrorist used the Infrastructure by Internet to exchange their private conversation as terrorrelated activities and schedule it by using more new groups and supporters

[6.] The techniques used to detect the terrorist by eavesdrop by using Internet protocol which is using by terrorists. But ,it is difficult to manage and monitor the terrorist IP address and url which is used to filter keywords and patterns .And there is a frequency change of Location of web server and there is no fixed IP Address. To overcome this disadvantages, there is a law which using ISP traffic through privacy issues. The integrated issues which is done by earlier using fields of Computer Security and retrieval of Information and mining by texting mechanism[7].

### III. PROPOSED METHODOLOGY

#### 1. Overview

We propose a system with the primary goal of developing a website where users can check any webpage or any website for any trace of terrorist activity. To do so, our website will provide the feature of entering the URL of the webpage the user wants to scan. After entering the URL, our system will tally the words of the whole webpage and tally them with the words that are already present in our database. Each word that we will store in our database will have a certain score to it. Our system will fetch the scores of each word that is present in the user's web page from our database, and in the end it will calculate a total rank of the website. This rank will determine if the user's webpage contains any trace of terrorism or not. Our system will detect patterns, keywords and relevant information in unstructured texts in a webpage

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 results.

Traditionally, there was no such system to keep an eye on various websites or any suspicious words present online. Cops were unable to track the terrorist related website or any person with suspicious information. The ratio of terrorism is high in today's world. There must be a system to track those suspicious word online and bring down the ratio of terrorism. In various arrangements and have images, videos etc. intermixed on a single web page. So we here propose to use smartly designed web mining algorithms to mine textual information on web pages and detect their relevancy to terrorism. In this way we may judge web pages and check if they may be promoting terrorism. This system proves useful in anti-terrorism sectors and even search engines to classify web pages into the category. Their relevance to the field helps classify and sort them appropriately and flag them for human review.

### 2. System Requirements

#### 2.1 Software Requirements Platform choice)

- Windows 7 and above 3.5.3(Server Side)
- Python 3.8
- HTML/CSS/JavaScript \
- Database

#### 2.2 Hardware Requirements

- Processor: Above 1.5GHZ
- Hard Disk: 80GB
- RAM: 2GB/4GB \
- CPU: Intel core i3

### IV. PROPOSED SYSTEM

#### a) Algorithm :

*Algorithm1:ProposedAlgorithm*

1. Select Website
2. Scan the website for words
3. Select proper database
4. Compare words from database
5. Word detection

- |    |                                       |
|----|---------------------------------------|
| 6. | Check energy consumption of everynode |
| 7. | Updating score                        |
| 8. | Display the score                     |
| 9. | Display violation alert message       |

First user needs to register to use the features of terrorism detector. For Registration user needs to provide his full name phone number email. After registration is done user receives a mail regarding his login info in which his username and password is generated. Next step is to login by user, once user login to system a command prompt is ran in background through which a directory is created in his personal system. In this directory all the records are stored of terror predictions. After login, User is prompted to provide a url which he/she thinks has terror related activities(Note: This system only works with http related websites), after providing a url system checks for terror related words in the url and it's sub urls. If site contains terror related activity the alert message is displayed that the url entered is not safe and is having some signs of terror, then user can report the site to cyber crime complaint website, the user receives the detailed summary and all terror related work which is found out in url in the same directory which was created at the time of login.

#### b) Web Crawler :

The web crawler is a computer program or automated script that crawls through the World Wide Web in a predefined and methodical manner to collect data. The web crawler tool pulls together details about each page: titles, images, keywords, other linked pages, etc. It automatically maps the web to search documents, websites, RSS feeds, and email addresses. It then stores and indexes this data. Also known as the spider or spider bot, the spider crawl program moves from one website to another, capturing every website. All contents are read and entries are created for a search engine index. The website crawler gets its name from its crawling behavior as it inches through a website, one page at a time, chasing the links to other pages on the site until all the pages have been read. Every search engine uses its own web crawler to collect data from the internet and index search results. For instance, Google Search uses the Googlebot.

Web crawlers visit new websites and sitemaps that have been submitted by their owners and periodically revisit the sites to check for updates. So, if you search for the term "web crawler" on Google, the results you get today may differ from what you got a few weeks ago. This is because a web crawler is continually at work, searching for relevant websites that define or describe a "web crawler" in the best manner, factoring in new websites, web pages, or updated content.

We compared all of the algorithms on the basis of their accuracy and correctness (tallying the words and score stores in the database and the words on the webpage that the user wants to check) by applying these algorithms on our dataset and chose the one which has the highest accuracy: Random Forest. Above table shows each of the implemented algorithms and their accuracy. Once you login, it will redirect you to the page where you can enter the URL of the webpages that you want to check for any trace of terrorism. On entering the URL and clicking on Search, it will show you the complete webpage that its checking along with the words that have the maximum occurrences and that are tagged in the database as related to terrorism.

#### IV. RESULTS AND DISCUSSION

In the proposed method, the aggregated data of url after web crawling is send to the user via mail, then user can take action on the specified url. The performance of proposed system is calculated by using packet delivery, dropping ratio, delay, RSN.

#### V. CONCLUSION AND FUTURE SCOPE

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. We analysed the usage of Online Social Networks (OSNs) in the event of a terrorist attack.

We used different metrics like number of tweets, whether users in developing countries tended to tweet, re-tweet or reply, demographics, geo-location and we defined new metrics (reach and impression of the tweet) and presented their models. While the developing countries are faced by many limitations in using OSNs such as unreliable power and poor Internet connection, still the study finding challenges the traditional media of reporting during disasters like terrorist's attacks. We recommend centres globally to make full use of the OSNs for crisis communication in order to save more lives during such.

#### VI. APPLICATIONS

##### 1. Detecting Malicious Websites:

Terrorist organizations are using the internet to spread their propaganda and radicalize youth online and encourage them to commit terrorist activities. In order to minimise the online presence of such harmful websites we

need to devise a system which detects specific keywords in a particular website. The website should be flagged inappropriate if the keywords are found for efficient system development. The system should help the cops to track communication held between terrorists and should detect web pages developed in different platforms.

## 2. Security Work:

Security Work can be use when terror related activity held in via internet the security team detect it and alert to the specific region.If the terror activity found the Security Team immediately take action.

## 3. Military Team:

When the terror related activity found security team alert to the Military Team and Military Team take immediately take action towards terrorist.

## 4. Cyber Security:

Cyber Security uses device system which detects specific keywords in a particular website.The website should be flagged inappropriate if the keywords are found for efficient system development. The system should help the cops to track communication held between terrorists and should detect web pages developed in different platform

## VII. ACKNOWLEDGEMENT

The authors would like to thank the publishers, researchers for making their resources available and teachers for their guidance. We also thank the college authority for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

## REFERENCES

- [1] H. Chen, W. Chung, J. Qin, E. Reid, M. Sageman and G. Weimann, "Uncovering the Dark Web: A case study of Jihad on the Web," *Journal of the Am. Soc. for Info. Sci. & Tech.*, vol. 59, pp. 1347-1358, 2008.
- [2] H. Chen, C. A. Larson, T. Elhourani, D. Zimbra and D. Ware, "The Geopolitical Web," in *IEEE International Conference on Intelligence and Security Informatics* Beijing, China: IEEE Press, 2011, pp. 60-64.
- [3] Innes, M., Roberts, C., Preece, A., & Rogers, D. (2016). Ten "Rs" of social reaction: Using social media to analyse the "post-event" impacts of the murder of Lee Rigby. *Terrorism and Political Violence*, IEEE Press, 2016
- [4] Spasi, J. (n.d.). FlexiTerm. <http://users.cs.cf.ac.uk/J.Spasi/flexiterm/India> : info & Sci.&Tech., vol.52, pp.1817-1828, 2019
- [5] V. E. Krebs, "Mapping networks of terrorist cells[J]", *Complex networks with their Connections A framework for online counter terrorism Social network analysis*, vol. 24, no. 3, pp. 43-52, 2020.
- [6] C. Freeman Linton, "Centrality in social networks: Conceptual clarification", *Social Networks Web Mining To Detect Spread of Terrorism*, pp. 215-239, 2019.
- [7] David J. Farley, "Breaking Al Qaeda Cells: A Mathematical Analysis of Counterterrorism Operations (A Guide for Risk Assessment and Decision Making) [J]", *Studies in Conflict & Terrorism*, vol. 26, no. 6, pp. 399-411, 2014.
- [8] Drazen Penzar and Armano Srblijinovic, "About modelling of complex networks with applications to terrorist group modelling[J]", *Interdisciplinary Description of Complex Systems*, vol. 3, no. 1, pp. 27-43, 2019.
- [9] Chen Peng and Yuan Hongyong, "Social network analysis of criminal organization structure[J]", *Journal of Tsinghua University (Natural Science Edition)*, vol. 51, no. 08, pp. 1097-1101, 2018.
- [10] L. K. Joshila Grace, V. Maheswari, Dhinaharan Nagamalai, "Analysis of Web Logs and Web User In Web Mining" *International Journal of Network Security & Its Applications (IJNSA)*, Vol.3, No.1, January 2011.
- [11] Rinkel .Goradiya, Shraavan .Mohite "Web Mining To Detect Spread of Terrorism The" ,Vol.4, No.1, March 2020
- [12] Fawad .Ali, Farhad .Farhan "Counter Terrorism on Online Social Networks A Mathematical Analysis" ,Vol.78, no.06, pp.1214-1220, 2019.
- [13] Benjamin Fabian Web Tracking Using social media to analyse the "post-event" impacts- "A Literature Review on the State of Research 2020" vol.52, no.07, pp.1890-1899, 2019
- [14] N Memon, N Harkiolakis and D L. Hicks, Detecting high-value individuals in covert networks: 7/7 London bombing case study[C]// IEEE/ ACS International Conference on Computer Systems & Applications., 2008.
- [15] Li Benxian, Li Mengjun, Sun Duoyong, Chi Yan and Fan Linjun, "The application of social network analysis in anti-terrorism [J]", *Complex Systems and Complexity Science*, vol. 9, no. 02, pp. 84-93, 2012.
- [16] Arush .Barfar, Muhammad Tarif "A framework for online counter terrorism Social network analysis of criminal organization structure" vol. 5, no.01, pp.93-115, IEEE 2009