IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Survey On Web Mining Techniques

Hrishikesh Wadge, Rohit Pitekar, Saurabh Warade,Shrikant A.Shinde

Department Of Comuter Technology
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 overthe world. In the recent times, terrorism has grown in an exponential mannerin 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, inter-net 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. Datamining algorithms are used to manage organized data sets and web mining algorithms can be helpful in mining.*

***Key Words*: Data mining System Architectures, Data mining Applications,Crime,Terrorism,Warehouse**

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.

II. Related Work

1. Aakash Negandhi 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 at 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.

I.

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 users web page from our database, and in the end it will calculate a total rank of the website. This rank will determine if the users webpage contains any trace of terrorism or not.

This chapter includes the details of the related papers with this system and the respective author's work. These papers are close to the objectives of this system and the observations of these research papers are analyzed in the proposed work. Various existing techniques and algorithms are described in following table.

Sr. No	Paper Name/Year	Author Name	Strengths	Limitations
1	Study on Classification Detect Online Spread of Terrorism Using Data Mining(2020)	Aakash Negandhi, Soham Gawas	If terrorists communicate in codewords not fed in system, the system will not be so efficient.	If terrorists communicate in codewords not fed in system, the system will not be so efficient.
2	T.Anand Terror Tracking Using Advanced Web Mining(2019)	T.Anand	System helps agencies to detect suspicious web pages and track them from their sources.	If terrorists communicate in codewords not fed in system, the system will not be so efficient.
3	Counter Terrorism on Online Social Net- works Using Web Mining Technique(2020)	Farhan Hassan Khan	This may result in finding the words of an attack using web application by terrorism.	if code word is used in this system then it is not efficient
4	if code word is used in this system then it is not efficient (2018)	Naseema Begum A.	This system alerts authorities to block those web pages on time.	Performs a well defined cleaning of data and data storage

5	Performs a well defined cleaning of data and data storage(2020)	Zinab Abdullah	extracting browser history information from log files to detect ISIS terror users.	used different metrics like number of tweets, whether users in developing countries tended to tweet, re-tweet or reply, demographics
6	A Frame-work for Online Counter Terrorism (2019)	Kiyana Zolfaghar, Arash Barfar	describes the major barriers to achieve effective online counter terrorism, and in section.	Lack of co-operation between organizations and secret services of different countries
7	Countering Terrorism Online withArtificial Intelligence (2020)	Trupti Kaule	provides guid-ance to law enforcement and counter-terrorism agencies	If terrorists communicate in code wordsnot fed in system, the system will notbe so efficient.
8	Report on-line material promoting by terrorist (2020)	Mark latham	Report illegal or harmful information, pictures or videos you've found on the internet	if data is huge this system willnot work.
9	Unauthorized Terror At- tack Tracking Using Usage Mining (2014)	Ramesh Yevale	It provides marketing intelligence. Web logs provide an exciting new way of collecting information.	This web mining technique can be used for detecting and avoiding terror threats caused by terrorists all over world.
10	Detection of online spread terrorism using web datamining (2020)	Aniket Subhash Dhanawade	Know how detection of terrorist activities are important for people.	People think that the internet is not safe because of online terrorist activities.

2. Conclusion

We need a proper system to detect and terminate websites which are spreading harmful content used to radicalizing youth and helpless people. We analyse the usage of Online Social Networks 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 Online Social network such as unreliable power and poor Internet connection, still the study finding challenges the traditional media of reporting during disasters like terrorists attacks.

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] 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.

[3] C. Freeman Linton, "Centrality in social networks: Conceptual clarification", Social Networks Web Mining To Detect Spread of Terrorism1, pp. 215-239, 2019.

[4] 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.

[5] Drazen Penzar and Armano Srbljinovic, "About modelling of complex net- works with applications to terrorist group modelling[J]", Interdisciplinary Description of Complex Systems, vol. 3, no. 1, pp. 27-43, 2019.

[6] 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.

[7] Fawad .Ali ,Farhad .Farhan "Counter Terrorism on Online Social Networks A Mathematical Analysis" ,Vol.78, no.06, pp.1214-1220, 2019.

[8] 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

[9] Arush .Barfar ,Muhammad Tarif "A framework for online counter terror- ism Social network analysis of criminal organization structure"vol. 5,no.01,pp.93-115, IEEE 2009

[10] 7.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.