WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Browser Forensics to Detect Online Spread of Terrorism

Dr. S. Mohandoss[1], E. Durga Nandini[2], Jothi S[3], Sharolin S[3], Alagu Parvathi Athinya A[3]*

[1]Department of Computer Science and Engineering, Dr. M. G. R. Educational and Research Institute, Chennai-600095, Tamilnadu, India.
[2]Centre Of Excellence in Digital Forensics, Chennai- 600095, Tamilnadu, India.
[3]Department Of Cyber Forensics and Information Security, Dr. M. G. R. Educational and Research Institute, Chennai-600095, Tamilnadu, India

A B S T R A C T

In recent times, terrorism has increased in various parts of the world. Cyber terrorism is a modern way for terrorists in the 21st century. Terrorists use modern social networks to execute their attacks. They gather people to take part in terrorist activities by spreading terrorism through videos and messages with the internet as a medium. For accessing the internet, a web browser, or an application program is used. The browser saves the user log file which contains user activities such as downloading files, accessing email accounts, use of social media applications. These log files help to collect information on criminals. The purpose of this project is to detect and respond to the spread of terrorism in web browsers. Browser forensics is used to investigate a browser's history, settings, and cookies. This also refers to monitoring traffic on the webpage and analysis of log files from the server to find keywords related to terrorism.

Keywords: Browser forensics, Web data mining, Terrorism, Machine learning, Search word analysis

1. Introduction

Since the late 1980s, the Internet has proven to be an excellent means of communication. The benefits of Internet technology are many; It has unique usability primarily for sharing information and ideas. It is important to realize that the technologies that facilitate this communication can also be used for terrorist purposes. Terrorism has gone deep in some parts of the world. The purpose of the internet is articles, videos, URLs, etc. is to publish.

Allow young people and spread hatred towards them. The use of the Internet for terrorism creates challenges and opportunities in the fight against terrorism. Data analysis, artificial intelligence, machine learning, etc. help protect people from threats. Crime Online Ad Search works on Twitter data to find crime-related tweets. The software uses sentiment analysis to classify terrorist tweets into threat tweets and informational tweets. The purpose of the detector is to help authorities find threatening tweets so that they can be removed from social media.

1.1 Problem Statement

Our system uses web mining and data mining to identify patterns, key terms, and related information in non-textual web pages. Our system will mine web pages using web mining algorithms to investigate information on web pages and detect malicious pages. Data mining and web mining are sometimes used together to get good results.

1.2 Proposed System

We propose a system whose main purpose is to create a website where users can check a web page or a website if there is a crime. To this end, our website will allow users to enter the URL of the web page they wish to browse. After you enter the URL, our system counts words from all web pages and compares them with words already in the file. Each word we will store in the file will have a score. Our system will get a score from our database for each word found on the user's website and finally calculate the overall ranking of the site.

This ranking will determine whether the user's website is criminal or not. Our system will only use the website to identify patterns, content and other information (including information search) in inappropriate text on web pages. Our system will use web mining algorithms to investigate information on the web page and check for crime related pages. Data mining and web mining are sometimes used together to get good results.

1.3 Machine Learning Algorithms

- Random forest algorithm, as the name suggests, consists of a large decision tree working together. Each tree in the random forest produces a class prediction, and the class with the most votes becomes our model's prediction.

- A decision tree is a tree diagram that represents nodes where we select attributes and ask query statements; edge, question. area; simple represents the actual output or text category. They are used for non-linear decisions with simple linear decision surfaces.
- Naive Bayes classifier is a collection of classification algorithms based on Bayes theorem. It's not just one algorithm, it's a set of algorithms that all show a single path, meaning each separate pair of features are independent of each other.
- Logistic regression is a type of predictive analysis. Logistic regression is used to describe data and explain the relationship between a binary variable and one or more nominal, ordinal, interval, or ratio independent variables.
- K Nearest Neighbors is a simple algorithm that stores all existing and classifies new events based on common quality metrics (such as remote work). KNN has been used as a nonparametric technique for prediction and pattern recognition since the early 1970s.

1.4 Implementation

We implemented a set of machine learning algorithms using WEKA (Waikato Environment for Information Analysis), free software licensed under the GNU General Public License, and the book companion software "Data Analysis: Using Technology Learning Tools and Methods"

We compare all algorithms in terms of accuracy and precision (words and scores stored in statistical data and a word on the web page that users want to test) data by applying their algorithms we choose our cluster and the most accurate algorithm: Random Forest. The table above shows each algorithm implementation and its accuracy.

Once you log in, it will redirect you to a page where you can enter the URL of the web page you want to check for signs of malicious activity. When you enter the URL and click Search, the information shows the entire page with the most frequently used words and symbols related to Violence.

1.5 Model of our System

Avoid hyphenation at the end of a line. Symbols denoting vectors and matrices should be indicated in bold type. Scalar variable names should normally be expressed using italics. Weights and measures should be expressed in SI units. All non-standard abbreviations or symbols must be defined when first mentioned, or a glossary provided.

1.5.1 Hardware & Software Requirements

a. HARDWARE REQUIREMENTS:

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It should what the system does and not how it should be implemented.

PROCESSOR : PENTIUM IV 2.6 GHz, Intel Core I3

RAM : 2 GB DD RAM

MONITOR : 15” COLOR

HARD DISK : 40 GB

CDDRIVE : LG 52X

KEYBOARD : STANDARDKEYS

MOUSE : 3 BUTTONS

b. SOFTWARE REQUIREMENTS:

The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the teams and tracking the team’s progress throughout the development activity.

4.1.1 Operating system : - Windows 07/ XP Professional

4.1.2 Front End : - Visual studio 2010, Microsoft .NET Framework

4.1.3 Back End :- Microsoft SQL Server

1.6 Modules and their Description

The system consists of the following 5 modules:

- Access: Here, the administrator or authorized personnel must enter the login information.
- Add keywords After successful login, the administrator can add more keywords selected for the attack.
- View Websites Here the administrator can add multiple URLs to check the website for suspicious messages.
- View / verify that all URLs entered by all site administrators are listed here. Can check suspicious keywords.
- Change Password The system allows administrators or authorized employees to change their passwords.

1.7 Methodology

Web mining algorithms are used to mine information on web pages and determine their relevance to crime. Websites created by various platforms can be followed using this application. The system will check whether web pages promote crime. Web pages will be divided into different categories and analyzed accordingly. This system uses two functions such as data searching and web mining. Data mining is the only technique used to efficiently use the data obtained from big data and use all the results obtained. Web mining also includes text-writing techniques that allow us to crawl and extract valuable content from malicious content. The system is for use only by federal officials responsible for national security. The system will help police easily keep track of communities affected by crime.

The site will have the following features:

- Load measurement: Since the system will only be available when the administrator enters, the load of the server will be limited to the administrator's time.
- Easy Access: Documents and other information can be easily accessed and stored separately.
- User Friendly: The website will provide a convenient way for all users.
- Effective and reliable: keep all data safe on the server and accessible and controllable when needed by the customer and compared to this keep all customer information in spreadsheets or logbooks, its price will be very good.
- Easy to use: Web Data Mining for Terrorism Analysis Websites

2. Illustrations

Fig. 1 - Login Page

Fig. 2 – Copy Paste Suspected URL

Fig. 3 – History of URLs

3. Applications

1. We use web mining algorithms to mine information on web pages and discover links to crime.
2. Websites created by various platforms can be followed using this application.
3. The system will check if the page promotes violence.
4. The system divides web pages into different categories and sorts them accordingly. The system uses two functions, namely data mining and web mining.
5. Data mining is a technique used to discover and exploit all the useful information in big data.
6. Web mining also includes text mining methods that allow us to crawl and extract valuable content from irrelevant content.
7. Data mining and web mining are sometimes used together for effective Development.
8. The system will monitor web pages that are more vulnerable to malicious activity and publish IP addresses to users using the system.

9. This process is used only by government officials working for national security. The system will help police easily monitor communities at risk of crime.

4. Conclusion & Future Scope

It is useful to prevent terrorist threats and destroy the online presence of terrorist groups such as ISIS and other radical websites. We need a system that will detect and remove websites that spread negative content and are used to make young people helpless. We examine the use of online networks (OSNs) in combating terrorists.

We use different metrics such as number of tweets, whether users in countries want to tweet, retweet or reply, population, region, and we set new metrics (reach and impressions) and show tweets. Although developing countries face many limitations in the use of OSNs, such as unreliable electricity and poor Internet connections, it is still difficult to publish findings when disasters occur such as terrorist attacks. We recommend that places around the world leverage OSN for emergency communications to save more lives in emergencies.

References

1. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1383486&queryText%3DWeb+Data+Mining+For+Terrorism+Analysis
2. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4547644&queryText%3DWeb+Data+Mining+For+Terrorism+Analysis
3. Junghoon Oh, Seungbong Lee, Sangjin Lee, “Advanced evidence collection and analysis of web browser activity”, Digital Investigation 8, 2011, 8, PP. 62-70.
4. Apurva Nalawade, Smita Bharne, Vanita Mane,“Forensic Analysis and Evidence Collection for Web Browser Activity”, IEEE 2016.
5. Murilo Tito Pereira, “Forensic analysis of the Firefox 3 internet history and recovery of deleted SQLite records”, ScienceDirect, Digital Investigation 2009, PP. 93-103.
6. Huwida Said, Noor Al Mutawa, Al Awadhi, Mario Guimaraes,”Forensic analysis of private browsing artifacts”, Proceedings of International Conference on Innovations in Information Technology IEEE 2011, pp. 198-202.
7. Erhan Akbal, Fatma Gunes, Ayhan Akbal, “Digital Forensic Analyses of Web Browser Records”, Journal of software,2016,doi: 10.17706/jsw.11.7.631-637 International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 07 | July 2018 www.irjet.net p-ISSN: 2395-0072 © 2018, IRJET | Impact Factor value: 7.211 | ISO 9001:2008 Certified Journal | Page 279
8. Andrew Marrington, Ibrahim Baggili, Talal Al Ismail, Ali Al Kaf, “Portable Web Browser Forensics -A forensic examination of the privacy benefits of portable web browsers”, IEEE 2012.
9. en.wikipedia.org
10. Microsoft Developer Network (MSDN): http://msdn2.microsoft.com/en-us/default.aspx: This is a valuable online resource, and is a must for any developer using Microsoft tools.
11. http://www.asp.net/: This is the official Microsoft ASP.NET web site. It has a lot of: tutorials, training videos, and sample projects.