Volume: 02, Issue: 06, Pages: 941-949 (2011)

XGraphtics^{CLUS}: Web Mining Hyperlinks and Content of Terrorism websites for Homeland Security

Dr.S.K.Jayanthi
Head & Associate Professor, Computer Science Department, Vellalar College for women, Erode, India
Email: jayanthiskp@gmail.com
Ms.S.Sasikala
Assistant Professor, Computer Science Department,KSR College of Arts and Science, Tiruchengode, India
Email: sasi_sss123@rediff.com

----------**ABSTRACT**----------
World Wide Web has become one of the best and fast communication media and information could be distributed within few seconds to the world day by day. The evolution of social networking media increases it further more to transfer information in a rapid speed to common people. Terrorism organizations utilize these facets of the web in very efficient manner for their destructive plans. Understanding web data is a decisive task to assure the better perceptive of a website. This paper focuses on the content and link structure mining of the website which was suspicious through XGraphtics^{CLUS}. This is done through viewing the web as graph and retrieving the various content of the website. This could help in terms of better understanding the motto and various other web connections in the suspicion. The navigational links offered in the particular website could leave with some informative evidence. This paper puts a step towards the national security and provides the user a good perception.

Keywords - **Hyperlinks, Content, Mining, Terrorism websites.**

Paper submitted: 14 December 2010 Date of Acceptance: 17 March 2011

I. INTRODUCTION

The popularity and development of web technology has made the Internet an important and popular application platform for disseminating and searching information. However, due to the lack of uniform schema for web documents and the huge amount of web information available on the Internet, web users always find it is difficult to obtain the needed information from the Internet accurately and easily. Such demands have posed a lot of challenges to web researchers and engineers.

Web site is a collection of Web pages that are linked to each other and very often to Web pages on other Web sites. In this paper, the structure of a Web site refers to the hyperlink structure of the Web site that is used to organize Web pages in hierarchy. Because this hyperlink structure usually reflects the implicit logical relationship among Web pages, it is directly applied to extracting relationship among the core content that connected Web pages are providing the relevant information's. Since the advent of the Internet, many studies have investigated the possibility of extracting knowledge and patterns from the Web, because it is publicly available and contains a rich set of resources. Many Web mining techniques are adopted from data mining, text mining, and information retrieval research. Most of these studies aimed to discover resources, patterns, and knowledge from the Web and Web-related data (such as Web server logs).

" Web mining is the application of data mining techniques to extract knowledge from Web data, where at least one of structure (hyperlink) or usage (Web log) data is used in the mining process (with or without other types of Web data)" [1].

A. Web usage mining

Web usage mining is a process of extracting useful information from server logs i.e users history. Web usage mining is the process of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data

B. Web content mining

Web content mining is the process to discover useful information from text, image, audio or video data in the web. Web content mining sometimes is called web text mining, because the text content is the most widely researched area. The technologies that are normally used in web content mining are NLP (Natural language processing) and IR (Information retrieval). Although data mining is a relatively new term, the technology is not. Companies have used powerful computers to sift through volumes of supermarket scanner data and analyze market research reports for years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.

C. Web structure mining

Web structure mining is the process of using graph theory to analyze the node and connection structure of a web site. According to the type of web structural data, web structure mining can be divided into two kinds:

1. Extracting patterns from hyperlinks in the web: a hyperlink is a structural component that connects the web page to a different location.
2. Mining the document structure: analysis of the tree-like structure of page structures to describe HTML or XML tag usage.

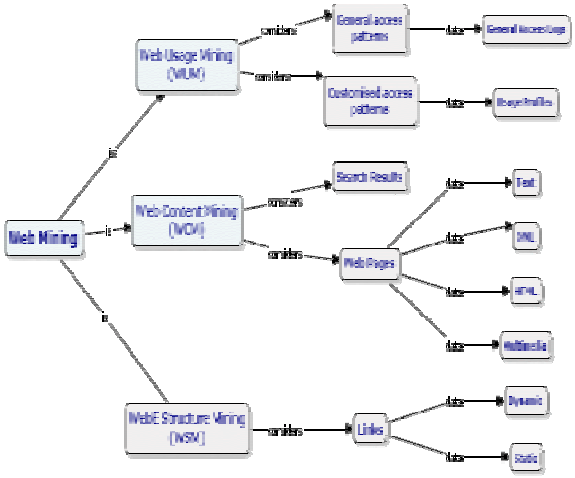

Figure 1 Web mining

II. UNSTRUCTURED DATA MINING

Data Mining is a familiar term today, thanks to the value it has added to the understanding of Structured Data in retail, science, engineering, and even anti-terrorism. However, deriving insights from unstructured data is a completely different ball-game.It is absolutely critical to gain actionable insights from data that pours in from the plethora of sources all around us, such as blogs, forums, user opinions, polls, survey feedback, reviews, portals and almost any other source online, where data exists as text, audio, images and other forms. These unstructured data sources often contain a wealth of critical information and knowledge which can be unlocked by Text Mining and other techniques.

Figure 2 Unstructured Data Mining - Components

Unstructured Data Mining leverages elements of Computational Linguistics, Natural Language Processing (NLP), Text Mining, Audio/Speech Processing, Image Processing & Analysis, Web Mining, Sentiment/Opinion Analysis, and Text Categorization. For any data, irrespective of the source from which it is obtained, the primary objective is to derive actionable intelligence and meaningful reporting.

However, even with all of this data available, very few applications exist that actually help business users with their decision making. While data mining and analytics applications address this business need, they are not widely adopted because of the unique nature of every business. A deep understanding of the business domain as well as expertise in data mining algorithms is required to build applications that are relevant to business users.

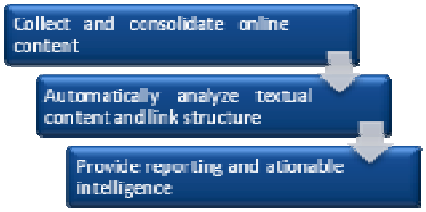

Figure 3 The Actionable Intelligence Workflow

Achieving this relevance is a challenge faced by many, more so when the data collected is unstructured. In addition, it is often time-consuming and expensive when dealing with such unstructured data coming from a variety of sources online.

- Automated Text classification
- Junk removal from online web articles
- Information and entity extraction from text using rules and machine learning based systems
- Discovering domain specific sentiment or opinion words
- Co-reference resolution and deep Natural anguage rocessing (NLP)

III. INTERNET AND TERRORISM

The Internet presents us with a paradox. On the one hand, it facilitates free discussion and the exchange of information. It also provides another venue for individuals to exercise their right to freedom of expression. But by the same mechanism, it allows people to seek each other out anonymously; to reinforce their negative views or plan violent action. The challenge for open societies, therefore, is to maintain the free flow of information and respect for freedom of expression, while discouraging those who would exploit it to harm others.

Geospatial imagery, such as Google Earth or Microsoft Bing, also present challenges. While such geospatial information systems provide amazing opportunities for the general public to view the world, the risk for abuse by bad actors anywhere in the world – whether terrorists or

criminals – in planning and executing attacks is *extremely* high. The argument by companies that provide these programs that these are tools, and they are not responsible for how they are used, is another challenge. How does the government balance freedom of speech and the protection of companies which provide such services with counterterrorism objectives and national security?

Consider the following statement

"From February 2004 through February 2010, FBI data show that individuals on the terrorist watchlist were involved in firearm or explosives background checks 1,228 times; 1,119 (about 91 percent) of these transactions were allowed to proceed because no prohibiting information was found—such as felony convictions, illegal immigrant status, or other disqualifying factors—and 109 of the transactions were denied. " – Source: US Government Accountability Office www.gao.gov.

But stopping criminal activity online is also important for the us. Safe havens in cyberspace and the ability to transfer funds, materiel, and people depend on existing regional underground networks, such as those that exist for narcotics trafficking and piracy. For this reason, fighting cyberterrorism and cyber crime demands a regional response. And Southeast Asia has one of the fastest growth rates of Internet usage in the world. This intends the need for anti-terrorism mechanisms

Effective counterterrorism policy requires strong international partnerships, as terrorism is too big a security threat for any one country to face alone. To be sure, terrorism is a common challenge shared by nations across the globe—one that requires diplomacy.

The net effect has been a strengthening of the international sense of resolve against cyberterrorism and terrorist use of the Internet, a renewed commitment to capacity-building, and development of global norms so that countries can work together toward better security. Our work on these issues is critical to balancing each nation's security in cyberspace and the free flow of information. With the Internet increasingly becoming an important part of our day-to-day activities, the cyber world has attracted attention from all the wrong corners, especially cyber terrorists. Even though terrorists continue to use their conservative methods of bomb blasts and armed attacks, they have also donned a new avatar by becoming tech savvy.

It is alarming to note that terrorists, criminals and other anti-state elements have often been early to adapt the new technology trends and are currently using cyber space to recruit, transfer/launder money, carry out propaganda/spread their ideology, network and co-ordinate their widely dispersed activities and attacks in an innovative way. Terrorists and criminals are using cyber attacks because of the advantages that Internet provides them. In addition to being cost-effective, it is difficult to keep a track and it can be remotely controlled from anywhere around the globe. Identities can be kept secret and can affect a large number of people. The list for such wrong doers is endless.

The use of cyber space by terrorists for communications and propaganda is well known. For example, the 9/11 attacks were preceded by certain innovative means of communications on the cyber space to avoid getting detected. Even the impact of the terrorist operation of 26/11 drew an increased attention to the very existence of the term cyber terrorism. In the 26/11 incident, it was not that a cyber device was used to trigger a bomb, like it once happened in the Tirupati bomb attack on Chandrababu Naidu. But cyber space was used for the preparation of the attack as well as for providing logistic support. After the event, cyber space was used by terrorists to mock the government too. All this brought cyber terrorism into the limelight that ended up in the clause on cyber terrorism being part of ITA 2008 passed on December 23-24, 2008. The way terrorist utilize the web, is quite disguise and unbelievable. It gives an aversion that how they utilize the resources

TABLE I TERRORIST UTILIZATION OF WEB

Web sites - As many as 50,000 terrorists web sites were in existence across the web
Forums - There are about 300 terrorist forums which includes more than 30,000 members with close to 1,000,000 messages posted.
Blogs, social networking sites, and virtual worlds- Blogs and social networking sites, mostly hosted by terrorist sympathizers.
Videos and multimedia content - Terrorist sites are extremely rich in content with heavy usage of multimedia formats containing about 1,000,000 images and 15,000 videos from many terrorist sites and specialty multimedia file-hosting third-party servers.

IV. DEFINING MOMENT

According to a recent survey conducted by Websense, among fifty CIOs, chief risk officers and IT managers, it was found that nearly all respondents (98%) were under pressure to protect their data from any loss. The gravity of the situation would be underlining in each of these crime categories:

- Unauthorized access to computer systems or networks/hacking
- Theft of information contained in an electronic form: This includes information stored in computer hard disks, removable storage media, etc

- Email bombing: This refers to sending large number of emails to the victim, who may be an individual or a company or even mail servers, which would ultimately result in crashing
- Data diddling: It involves altering raw data just before a computer processes it and then changing it back after the processing is completed
- Salami attacks: Such crime is normally prevalent in financial institutions or for the purpose of committing financial crimes. An important feature of this type of offence is that the alteration is so small that it would normally go unnoticed. For example, the Ziegler case wherein a logic bomb was introduced in the banks system, which deducted 10 cents from every account and deposited it in a particular account
- Denial of service attack: The victims computer is flooded with more requests than it can handle, which makes it crash. Distributed denial of service (DDoS) attack is also a type of denial of service attack, in which offenders are large in numbers and are widespread. For example, Amazon, Yahoo!, etc
- Virus/worm attacks: Viruses are programs that attach themselves to a computer or a file and then circulate themselves to other files and to other computers on a network. They usually affect the data in a computer, either by altering or deleting it. Worms, unlike viruses do not need the host to attach themselves. For example, love bug virus, which affected at least 5% of the computers across the globe resulted in losses worth $10 mn. The worlds most famous worm was the Internet worm let loose on the Internet by Robert Morris in 1988, which almost brought development of the Internet to a complete halt
- Logic bombs: These are event dependent programs
- Trojan attacks: It is an unauthorized program which passively gains control over anothers system by representing itself as an authorized program
- Internet time thefts: Normally, in these kinds of thefts the Internet surfing hours of the victim are utilized by another person
- Web jacking: Here, the hacker gains access and takes control over the website of another person and then maybe mutilates or changes the information available on the site
- Copyright infringement, child pornography, etc

V. THREAT POSED BY DARK WEB CYBER

The count of potential targets is enormous and is growing by the day. Everything becomes a potential target, from private networks, institutions, academia, the military, the banking system, the government, the private business systems, public utilities, airlines , and many others

* The amount, the variety and the complexity of potential targets ensures that terrorists will find weaknesses and vulnerabilities to exploit

* The cyber space provides anonymity which is of course welcomed by the terrorists using it. methods. Operation wise, terrorist attackers, lake any other user use screen names and can log on an existing website without the need of identification or a proof of, making it very difficult for the protective agencies and law enforcement forces to track down the terrorists' real identity

* Both scientific and practical studies have shown that the entire critical infrastructures, as detailed in the list released by the DHS, including electric power grids, emergency services, oil and gas pipelines and refineries, the water system, airports and commercial ports, all are vulnerable to a cyber terrorist attack. But not only those systems are exposed; the military and intelligence networks are even more susceptible, although the consensus is that those networks are much better protected. This is a viable threat mainly because the infrastructures and servers/computers systems that run those networks are very complex and connected with each other in many subtle undetectable ways, making it effectively impossible to eliminate all potential weaknesses, not to mention problems that such system encounter routinely

* Terror based cyber attacks are launched remotely, a very appealing characteristic to terrorists. Not only that, but once triggered a cyber terrorism attack needs no more monitoring, supervision or presence on the web, and the results, if successful, can be heard and seen shortly on all media outlets.

* Cyber terror does not requires physical training, only recruiting of well trained IT professionals, which is an easy task. In addition, there is no need of psychological training, there is no physical risk and the chances of being caught are slim anyway, better than other, more dangerous alternatives, and all together easier and more safe.

* There is no need of subsequence investments since everything is virtual, remote and unidentifiable. If the professional knowledge exists, then the goals are set to fit the knowledge that such an IT terrorist can deliver. And the more accumulated knowledge is put in the game, the greater the threats.

* A crucial element and probably the main interim goal set by terrorists is receive as much as possible media coverage for as long as possible. Generating publicity, propaganda campaigns and using the internet as a recruiting tool have

proven to be very effective, which makes the internet Even more attractive as a strategic target.

VI. SOCIAL CONNECTIONS

In data mining, the practice of looking for underlying connections between people is called social network analysis. Phone data is useful because it helps expose relationships and associations among different groups. With social network analysis, contacts are commonly laid out graphically to illustrate connections and find patterns. At the simplest level, this could be shown as links similar to the spokes of a wheel leading to one source, indicating that a person holds a leadership position within a terrorist cell. Looking deeper, it could uncover relationships, such as two suspected terrorists linked only through a third, unknown person.

VII. LINK ANALYSIS

Link analysis could be used to find abnormal patterns in web. There have been many discussions in the literature on link analysis. Link analysis uses various graph theoretic techniques. It is essentially about analyzing graphs. Note that link analysis is also used in web data mining, especially for web structure mining. With web structure mining the idea is to mine the links and extract the patterns and structures about the web. Search engines such as Google use some form of link analysis for displaying the results of a search.

As stated that the challenge in link analysis is to reduce the graphs into manageable chunks. As in the case of market basket analysis, where one needs to carry out intelligent searching by pruning unwanted results , with link analysis one needs to reduce the graphs so that the analysis is manageable and not combinatorial explosive. Therefore results in graph reduction need to be applied for the graphs that are obtained by representing the various associations.

The challenge here is to find the interesting associations and then determine how to reduce the graphs. Various graphs theoreticians are working on graph reduction problems. It is need to determine how to apply the techniques to detect abnormal and suspicious behavior. Another challenge on using link analysis for counter-terrorism is reasoning with partial information. For example, agency A may have a partial graph, agency B another partial graph and agency C a third partial graph. One would argue that it is need a data miner that would reason under uncertainty and be able to figure out the links between the three graphs. This would be the ideal solution and the research challenge is to develop such a data miner.

The other approach is to have an organization above the three agencies that will have access to the three graphs and make the links. One can think of this organization to be the Homeland security agency. In the next section as well as in some of the ensuing sections we will discuss various federated architectures for counter-terrorism. We need to conduct extensive research on link analysis as well as on other data and web data mining techniques to determine how they can be applied effectively for counter-terrorism. For example, by following the various links, one could perhaps trace say the financing of the terrorist operations to the president of say country X. Another challenge with link analysis as well with other data mining techniques is having good data. However for the domain that we are considering much of the data could be classified.

If we are to truly get the benefits of the techniques we need to test with actual data. But not all of the researchers have the clearances to work on classified data. The challenge is to find unclassified data that is a representative sample of the classified data. It is not straightforward to do this, as one has to make sure that all classified information, even through implications, is removed. Another alternative is to find as good data as possible in an unclassified setting for the researchers to work on. However, the researchers have to work not only with counter-terrorism experts but also with data mining specialists who have the clearances to work in classified environments. That is, the research carried out in an unclassified setting has to be transferred to a classified setting later to test the applicability of the data mining algorithms. Only then can we get the true benefits of data mining.

For creating clusters within the suspect websites the following algorithm could be implied.
For each webpage x there exist a coefficient giving the degree of being in the kth cluster $u_k(x)$. Usually, the sum of those coefficients for any given x is defined to be 1:

$$\forall x \left(\sum_{k=1}^{\text{num. clusters}} u_k(x) = 1 \right).$$

With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

$$\text{center}_k = \frac{\sum_x u_k(x)^m x}{\sum_x u_k(x)^m}.$$

The degree of belonging is related to the inverse of the distance to the cluster center:

$$u_k(x) = \frac{1}{d(\text{center}_k, x)},$$

then the coefficients are normalized and fuzzyfied with a real parameter $m > 1$ so that their sum is 1. So

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(\text{center}_k, x)}{d(\text{center}_j, x)} \right)^{2/(m-1)}}.$$

For m equal to 2, this is equivalent to normalizing the coefficient linearly to make their sum 1. When m is close to 1, then cluster center closest to the point is given much more weight than the others. Steps involved in creating the XGRAPHTICS^{CLUS} listed as follows.

- Choose a number of clusters.
- Assign randomly to each webpage coefficients for being in the clusters.
- Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than ε, the given sensitivity threshold) :
    - Compute the centroid for each cluster, using the formula above.
    - For each point, compute its coefficients of being in the clusters, using the formula above.

The algorithm minimizes intra-cluster variance as well.
Consider w, URL of the sample webpage it resides in a domain which was treated as a cluster CLUS(w). Now consider ^{IN}(CLUS(w)), the incoming links to the particular domain of w. Also assume that the link concludes at certain point in the domain and lead to another domain and called as CON^{Temp}. The outgoing links which leads to the different cluster can be considered as ^{OUT}(CLUS(w)). TRA^{LVL} could be set to a fixed value to restrict the iteration. Consider the threshold TV^{DSCLUS}. Now if the web page exceeds the threshold limit then the page will be marked as spam page.

1. For each URL x in ^{IN}(CLUS(w)) perform
2. If CLUS(x) != CLUS(w) and not present in ^{IN}(CLUS(w)), then add it in the Cluster
3. Set w as CON^{Temp} and set current level of traversal TRA^{CUR} to 0.
4. If level, TRA^{CUR}<= TRA^{LVL} then, For each URL y in ^{OUT}(CLUS(w)) perform
    a. If CLUS(y) != CLUS(x) and if it is not found in ^{OUT}(CLUS(w)) then add it to the domain list of outgoing links.
    b. Else if CLUS(y) = = CLUS(x), then set TRA^{CUR}++ and set y as TRA^{LVL} and repeat step 3 and 4.
5. Calculate the intersection of ^{IN}(CLUS(w)) and ^{OUT}(CLUS(w)). If the number of elements in the intersection set is equal to or bigger than the threshold TV^{DSCLUS}, mark x as a bad page.
6. Repeat the steps for every search result URL, x.

Here with this a web graph simulated is shown in figure 4 which may further improvised in order to find the cluster structures based on content and link based parameters. The graph simulated in figure 4 is a type of online crime , click frauds. Likewise the graph could be constructed for other potential suspects.

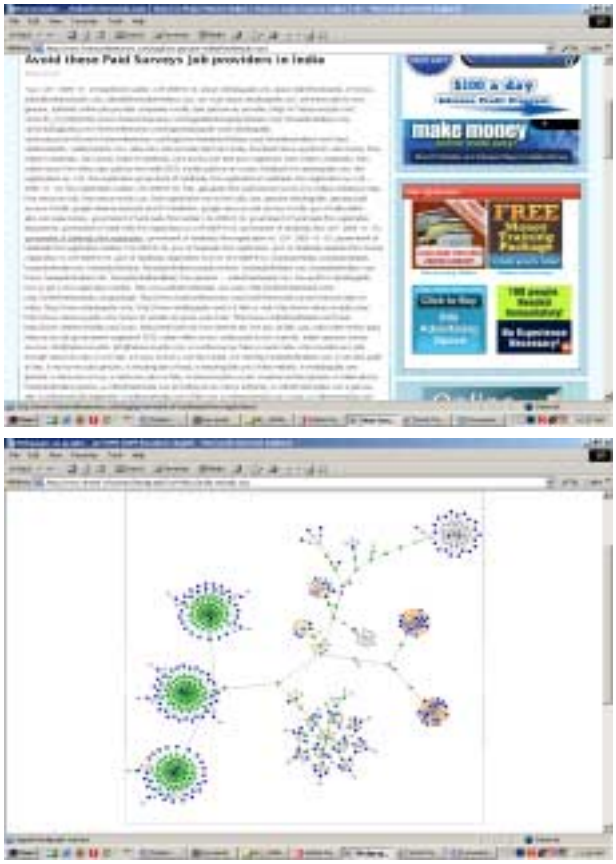

Figure 4 web graph

In addition the following are some potential state of the art techniques implied to mine the suspicious data from the web.

A. Methodology for countering terrorism

The computational tools are grouped in two categories:
i. *Collection*
ii. Analysis and Visualization.
iii. Collection-

1) Visualisation of website navigation

Visualising the navigation of a website is very important in order to understand the users behaviour while visiting a website [4]. In that sense, WET calculates Edmonds algorithm to extract the Maximum Spanning Tree of the webgraph using the frequency of usage of the links on the site. The resulting navigational tree depicts the paths followed by most of the users from a selected root node to any other page on the site.

2) Highlighting System

To provide a maximum insight on the data, and to take advantage of all the visualisations at the same time, a system has been built in order to highlight the same information in all the available visualisations. For instance, this system enables the location of a selected page in the website structure as well as in the navigational tree. Hence, it is easy to compare the fastest path needed to reach a page from the root, and the real path followed by the users.

3) Metrics Mapping System

This system allows the mapping of the calculated web metrics, overlaying them on top of the hierarchical visual metaphors used to visualise the structure and the navigation of the users. Hence, every web metric can be mapped into one visual attribute, enabling the visualisation of several metrics at the same time. By now, the system has implemented four visual attributes: size, colour, shape and border colour.

4) Web site spidering

Spiders/crawlers are based on previous digital library research. Spiders can access password-protected sites and perform randomized (human-like) fetching. Spiders are trained to fetch all html, pdf, and word files, links, PHP, CGI, and ASP files, images, audios, and videos in a web site.

5) Forum spidering

The forum spidering tool recognizes forum hosting software and their formats. by collecting the complete forum including: authors, headings, postings, threads, time-tags, etc., which allow us to re-construct participant interactions and by processing forum contents in Arabic, English, Spanish, French, and Chinese using selected computational linguistics techniques.

6) Sentiment and affect analysis

Not all sites are equally radical or violent. Sentiment (polarity: ositive/negative) and affect (emotion: violence, racism, anger, etc.) analysis allows one to identify radical and violent sites that warrant further study.

7) Video analysis

A significant portion of our videos are lED related. Based on previous terrorism ontology research, we have developed a unique coding scheme to analyze terrorist generated videos based on the contents, production characteristics, and meta data associated with the videos.

8) Dark Web analysis

A smaller number of sites are responsible for distributing a large percentage of lED related web pages, forum postings, training materials, explosive videos, etc and unique signatures can be used for those sites based on their contents, linkages, and multimedia file characteristics. Much of the content needs to be analyzed by military analysts. Training materials also need to be developed for troops before their deployment.

VIII. CONCLUSION

Researchers are still unclear whether the ability to communicate online worldwide has resulted in an increase or a decrease in terrorist acts. It is agreed, however, that online activities substantially improve the ability of such terrorist groups to raise funds, lure new faithful, and reach a mass audience. The most popular terrorist sites draw tens of thousands of visitors each month. Obviously, the Internet is not the only tool that a terrorist group needs to 'succeed.' However, the Net can add new dimensions to existing assets that groups can utilize to achieve their goals as well as providing new and innovative avenues for expression, fundraising, recruitment, etc. At the same time, there are also tradeoffs to be made. High levels of visibility increase levels of vulnerability, both to scrutiny and security breaches. The proliferation of official terrorist sites appears to indicate that the payoffs, in terms of publicity and propaganda value, are understood by many groups to be worth the risks.

Data mining and web data mining technologies will have a significant impact on counter-terrorism. The discussions on counter-terrorism in this paper are based on various newspaper articles and documentaries. The goal is to explore how data mining can be exploited for counter-terrorism. The purpose of the paper is to raise the awareness that data mining could possibly help detect and prevent terrorist attacks. This paper also provided an overview of some of the privacy concerns and discussed the directions in privacy preserving data mining and privacy constraint processing. There are many discussions now on privacy preserving approaches as it is necessary to continue with this research and develop viable solutions that can carry out useful mining and at the same time ensure privacy. As it is seen, one of the major concerns of the nation today is to detect and prevent terrorist attacks. This is also becoming the goal of many nations in the world. It is necessary to examine the various data mining and web mining technologies and see how they can be adapted for counter-terrorism. It is also need to develop special web mining techniques for counter-terrorism.

REFERENCES

[1] M Ozaki, Y. Adachi, Y. Iwahori, and N. Ishii, Application of fuzzy theory to writer recognition of Chinese characters, International Journal of Modelling and Simulation, 18(2), 1998, 112-116.

[2] Weimann, G. (2006). Terror on the Internet: The new arena, the new challenges. WashingtonD.C.: United States Institute of Peace Press.

[3] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," in Proc. 1998 WWW Conf., Brisbane, Australia, 1998.

[4] J. Kleinberg, "Authoritative sources in a hyperlinked environment," in Proc. 9th ACM-SIAM Symp. Discrete Algorithms, San Francisco, CA,1998.

[5] M. Chau and J. Xu, "Mining communities and their relationships in blogs: A study of online hate groups," Int. J. Hum.-Comput. St., vol.65, no. 1, pp. 57–70, 2007.

[6] X. Fang, M. Chau, P. J. Hu, Z. Yang, and O. R. L. Sheng, "Web miningbased objective metrics for measuring Website navigability," in Proc. Int. Conf. Inf. Syst., Milwaukee, WI, 2006.

[7] H. M. Chen and M. D. Cooper, "Using clustering techniques to detect usage patterns in a Web-based information system," J. Amer. Soc. Inf.Sci. Tech., vol. 52, no. 11, pp. 888–904, 2001.

[8] G. Marchionini, "Co-evolution of user and organizational interfaces: A longitudinal case study of WWW dissemination of national statistics," J. Amer.

Soc. Inf. Sci. Tech., vol. 53, no. 14, pp. 1192–1209, 2002.

[9] H. Chen, E. Reid, 1. Sinai, A. Silke, and B. Ganor (Eds.), "Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security," Springer, forthcoming, 2008.

[10] H. Chen, S. Thoms, T. Fu. "Cyber Extremism in Web 2.0: An Exploratory Study of International Jihadist Groups," in Proceedings of the 2008 IEEE Intelligence and Security Informatics Conference, Taiwan, June 17-20,2008.

[11] Reid, E. and H. Chen, "Contemporary Terrorism Researchers' Patterns of Collaboration and Influence," Journal ofthe American Society for Information Science and Technology, forthcoming, 2008.

[12] Measuring Radicalization on the Internet, " in Proceedings of the IEEE International Intelligence and Security Informatics Conference (Taipei, Taiwan, July 17-20, 2008). Springer Lecture Notes in Computer Science.

[13] www.gao.gov - United States Government Accountability Office For Release on Delivery Expected at 10:00 a.m. EDT Wednesday, May 5, 2010

Authors Biography

Dr.S.K.Jayanthi received the M.Sc., M.Phil., PGDCA, Ph.D in Computer Science from Bharathiar University in 1987, 1988, 1996 and 2007 respectively. She is currently working as an Asso. Professor, Head of the Department of Computer Science in Vellalar College for Women. She secured District First Rank in SSLC under Backward Community. Her research interest includes Image Processing, Pattern Recognition and Fuzzy Systems. She has guided 18 M.Phil Scholars and currently 4 M.Phil Scholars and 4 Ph.D Scholars are pursuing their degree under her supervision. She is a member of ISTE, IEEE and Life Member of Indian Science Congress. She has published **5** papers in International Journals and one paper in National Journal and published an article in Reputed Book. She has presented **14** papers in International level Conferences/Seminars (Papers has been published in IEEE Xplore, ACEEE Search Digital library, online digital Repository and with ISBN Numbers) in various places within India and in London (UK), Singapore and Malaysia, **18** papers in National level Conferences/Seminars and participated in around 40 Workshops/Seminars/Conferences/FDP.

S.Sasikala, currently working as an Asst. Prof. in K.S.R. College of Arts & Science has received the B.Sc(CS) from the Bharathiar University, M.Sc(CS) from the Periyar University, M.C.A. from Periyar University , M.Phil from Periyar University, PGDPM & IR from Alagappa university in 2001, 2003, 2006, 2008 and 2009 respectively. And she is currently pursuing her Ph.D in computer science at Bharathiar University. Her area of Doctoral research is Web mining. She secured University First Rank in M.Sc(CS) Programme under Periyar University and received Gold Medal from Tamilnadu State Governor Dr. RamMohanRao and Cabinet Minister Ms. Subbulakshmi Jagadeesan in 2004. She has published 2 papers in International Journals (IJACST and IJNGN) and refereed 2 International Journals IJNGN and JEEER. She has presented 12 papers in International conferences/Seminars (Papers has been published in IEEE Xplore, ACEEE Search Digital library, online digital Repository and with ISBN Numbers) in various places within India and in Singapore and Malaysia, 22 papers in National Conferences/Seminars and participated in 4 National Conferences/Seminars and 2 Workshops and has a total of 36 publications.