Semantic Data Mining for Security Informatics: Opportunities and Challenges

Syed Ahsan, Saleh Alshomrani, Ali Hassan

FCIT, King Abdulaziz University (North Jeddah)
saehsan@kau.edu.sa, sshomrani@kau.edu.sa

Abstract: Terrorism has emerged a most serious and dangerous threat to the world. This unparallel threat seriously threatens the infrastructure, economy and the people. It is important to investigate into such systems that may help to prevent terrorist attacks by reducing its vulnerability and also minimizing the damage and recovery from attacks that occur. In order to achieve these objectives there is a need for new approaches and methods to gather, store, analyze and share intelligence and information through the use of information technology. Also, the adoption of WWW and modern Information Technologies by the terrorist organizations has necessitated the need of an IT-based counter terrorism infrastructure. In this paper we attempt to identify, i) the areas where IT can contribute in countering terrorism ii) the unique IT problems and challenges in counter terrorism applications where such applications are being used and developed, and, iii) lessons learned for developing countries so that an IT counter terrorism infra structure can be established with minimum cost in terms of time and money. We have also tried to highlight the role Information Technology and Internet have to play in counter terrorism by developing and using advanced information technologies such as Data Mining and Semantic Web. The challenges posed are discussed and initiatives to be taken are also suggested.

[Ahsan S, Alshomrani S, Hassan A. **Information Technology and the World Wide Web: New Tools, New Frontiers in War against Terrorism**. *Life Sci J* 2013;10(12s):417-422] (ISSN:1097-8135). http://www.lifesciencesite.com. 71

Keywords: Counter terrorism, disaster mitigation, homeland security, semantic web, data mining, heterogeneous databases, interoperability; Cognitive Distraction; Drivers; inattention; Features Extraction; Fatigue

Introduction

Two decades ago, strict physical security at vulnerable facilities, intelligence gathering by government intelligence agencies, vigilant patrolling on international borders and other conventional methods were used to deter terrorism and terrorist activities [9] This also included natural and man-made disasters which posed an equally challenging set of problems such as an optimal response plan, resource allocation and distribution, material and social support [2]. In the last few decades, technology has played a revolutionary role in improving the economy with dramatic productivity improvements and myriad of new communications and information processing tools [1]. Whereas modern terrorist organizations have been quick to appreciate and adapt the capabilities of modern computing and communication devices, the security and intelligence apparatus is just beginning to realize the potential of IT and communication technologies in crucial areas such as domestic defense and emergency preparedness [2].

Modern terrorism as gradually evolved in to an organized crime [3] . Similar to other organized crimes such as drug mafias, terrorist organizations and groups need to communicate and coordinate to plan terrorist attacks and conspiracies [10]. For this purpose, these terrorist organizations use the World Wide Web as their medium of choice. Also Information Technologies such as the Internet are misused by terrorists and criminals for identity deception, identity theft, recruitment, and as a propaganda medium to promote hatred and violence [13]. This Internet, when used in this criminal context is referred to as the Dark Web [11, 12].

Amongst all the technologies, Information technology has been cited as the most important tool in making a country safer from terrorism [1]. IT can support intelligence and knowledge discovery, pattern extraction and probabilistic analysis by collecting, processing, and analyzing, terrorism-and crime-related data [5, 11]. We feel that countries across the globe must invest resources and effort in this domain to gain maximum security and public confidence with minimum investment [1, 2]. In our opinion, the modern information technology has tools and if supported by infrastructure, and commercial capabilities, it can make domestic defense easier, less expensive, and more effective.

In this paper, we have highlighted the steps and initiatives taken by the developed countries and argued the need of a similar IT based security structure for developing countries. In Section 2, we have described the IT-enabled counter-terrorism efforts of developed countries and the counter- terrorism infrastructure that is in-place there. The challenges that must be overcome to enable a functional IT enabled counter terrorism infrastructure are discussed in Section 3. The role of IT technologies such as Semantic Web is discussed in Section 4. We have argued the need of a similar

structure for developing countries in Section 5. In Section 6, we conclude the chapter by drawing conclusions and listing out future directions.

2. It-Enabled Counter Terrorism Infrastructure

Countering modern terrorism is a complex activity warranting multi and interdisciplinary effort. That is why a concerted effort by experts in the natural, computational, and social sciences as well as engineering, medicine, and many other fields is needed to enhance the capabilities in fighting the new counterterrorism war [9]. Developed countries have already taken significant steps in this direction [4]. Amongst all the technologies, Information technology has been cited as the most important tool in making a country safer from terrorism [1]. IT can support intelligence and knowledge discovery, pattern extraction and probabilistic analysis by collecting, processing, analyzing, and developing applications for terrorism-and crime-related data [7]. Six critical areas have been identified where IT can contribute to accomplishing strategic national security objectives [2].

i) Pre emptive measures through Intelligence and warning

Terrorists elaborately plan and prepare before executing an attack. They may employ false identities, virtual networks and other means to avoid their detection. By looking for activity patterns using data mining techniques employed over heterogeneous data resources, IT can detect deceptive identities. By employing other surveillance and monitoring techniques aided by biometric technologies, real time and critical alerts and warnings can be generated [2, 9].

ii) Smart and vigilant Borders

Machine readable documents such as passports can help in interactively sharing travel information on borders. Information such as traveler identities, images, fingerprints, vehicles used, and other characteristics can greatly improve counterterrorism and crime-fighting capabilities. Information sharing and integration through interoperability and data integration, collaboration and communication, and biometrics such as image and speech recognition can help achieve this [2,9].

iii) Global approach to counter-terrorism

Modern terrorism is an organized crime where terrorists might participate in local crimes to generate funds for international or domestic terrorism. IT can help find this relationship between domestic and international terrorism. Similarly terrorist from one country may receive training from another country [2, 7 and 9]. Also, criminal use of email and web pages can help law enforcement agencies and policy makers.

iv) Protecting national infrastructure and critical assets

Roads, bridges, water supplies, and many other physical service systems constitute critical infrastructure and key national assets [2]. Their importance, public nature and hence vulnerabilities make them potential targets of terrorist attacks. Virtual (cyber) infrastructures such as the Internet are also vulnerable to terrorist attacks such as hacking, intrusions and other threats [6]. Cyber terrorism is a modern day reality.

To safeguard these assets, we need not only physical devices, such as sensors and detectors, but also advanced information technologies that can detect abnormal use and behaviors by distinguishing it from normal behaviors. .

v) Contingencies and Emergency preparedness

Prompt and effective responses reduce the damage in national emergencies. In addition to helping defend against catastrophic terrorism, IT can help design and test optimized response plans through simulations, identify experts, train response professionals, and manage the consequences of an attack. [2].

The investments in such an IT structure should be into a whole range of supporting technologies [3] such as those needed for identity verification. An automotive identity validation system can be used as a tool to prevent future terrorist events and also in post-event investigation. Investments in technologies such as Smart ID cards can result in implementing the following so that real time and prompt decisions can be taken to deter terrorism [9]:

i) Data integration and sharing

Providing real time, remote access to federated data sources and criminal records and intelligence information from a variety of federal, provincial, and local agencies can help identify wanted criminals and suspected terrorists when they encounter law enforcement or attempt to enter secure facilities [7,20].

ii) Smart ID cards and visas with biometric identifier

Driver's licenses and other identification documents can be made forgery and fraud proof by adding biometric signatures such as thumbprint scans and other standardized security features [1].

iii) Digital surveillance

Extending conventional principles of law enforcement and surveillance to the Internet by permitting surveillance of email and other electronic data will help prevent cyber crimes and cyber terrorist attacks.

v) Face recognition technology that can detect known terrorists amongst crowds at potential targets such as the Olympics and other national festivals or social, cultural or religious events [1, 10].

vi) Sentiment Analysis ability

The WWW has become the medium of choice for terrorist organizations to communicate and coordinate

their activities [14, 15 and 16]. Web forums are extensively used for information interchange and for promotion of terrorist agenda [13]. Efforts are underway to develop techniques to enable effective analysis of forum conversation sentiments. [14, 16 and 18]. Techniques using support vector machines (SVM) already indicate a high level of classification accuracy for classifying and analyzing sentiments in extremist forums [17].

vii) Terrorist Network Detection

As discussed in Section 1, modern terrorism is an organized crime that is planned and executed by terrorist networks spread globally. A clear understanding of network structures, operations, and individual roles is essential for law enforcement and intelligence agencies to develop, plan and execute effective control strategies to prevent terrorist activities from taking place [18].

Hague Program includes the following measures[24, 25]:

- European evidence warrant;
- European arrest warrant;
- strengthening of Schengen and visa information systems;
- biometric details on passports;
- combating terrorist financing (see below);
- prevention of recruitment and radicalization;
- greater controls over trade, storage and transport of explosives.

The successful implementation of this measure relies heavily on a reliable and efficient IT infrastructure. SIS is central to the strategic goals of EU's Counter-Terrorism Strategy. These strategic goals are: Prevent, Protect, Pursue, and Respond [24, 25].

i) Prevent

The goal focuses on countering radicalization and recruitment to terrorist groups. It emphasizes on ways impede terrorist recruitment through the internet using sentiment detection and pattern extraction techniques and technologies.

ii) Protect

The commitment is to reduce the vulnerability of key targets to attack and mitigating the resulting impact of an attack. This requires EU's collective action involving collaboration of border security, transport and other cross-border infrastructures. The establishment of the Visa Information System and second generation SIS ensures that the relevant authorities and agencies can share and access and share the necessary information.

iii) Pursue

Detection and arrest of potential and convicted terrorists, and individual deported for terrorism related offences necessitates the exchange of information and intelligence between national security and intelligence agencies and Europol, FRONTEX, the European Borders Agency European Arrest Warrant and European Evidence Warrant. The development of new IT systems such as the Visa Information System and the next generation SIS has provided improved access to the authorities responsible for internal security thereby widening the base of information at their disposal.

iv) Respond

An act of terrorism in EU is likely to have cross border warranting a need for rapid sharing of operational and policy information. Schengen Information System (SIS II) enables sharing of information, coordination and collective decision making in emergencies. SIS has also facilitated an increased coordination between FRONTEX, the European Borders Agency, EUROPOL and EUROJUST.

3. Counter-Terrorism Domain Challenges

Government agencies collect a large amount of public and related data. This data is in addition to the data collected by law enforcement agencies. As a result, an information and data bottleneck is being created because our ability to record data far outpaces our ability to interpret it. Also this data poses some unique IT problems and challenges owning to its volume and peculiar characteristics.[1].

i) Temporal /spatial distribution and inadequate cyber laws

With the global availability of internet and global technologies, terrorists are able to work and plan transparently across national boundaries. They can correspond, communicate, plan and execute their plans while located in safe heavens across the globe. This makes it difficult to track them or prevent them from coordinating. [9]. With the laws and policies governing cyber world and the transactions taking place through it, still in nascent stage, criminals are able to commit various types of cyber crimes with impunity [9].

ii) Heterogeneous sources and formats

Large data volumes and diverse sources and formats create significant challenges owning to their large volumes, distribution and independent development [4]. The sources of data for law enforcement and intelligence agencies are mostly distributed, non standardized and unstructured. [3].

Moreover, as there is no agreed format for these data sources and most of the people involved in data gathering are not trained in IT, data formats range from structured database records to unstructured text, image, audio, and video files [1, 18].

iii) Identifier Conflicts

There are no standards governing the usage of identifier amongst various government agencies. This leads to identifying two different concepts/entities with some common identifier or worst still identifying two different concepts or entities with some common identifier [1, 10].

iv) Sentiment Analysis

Sentiment analysis attempts to identify and analyze opinions and emotions. Although analysis of terrorist group's web content has attracted the attention of researchers, there is a need for increased focus on analysis and evaluation of forum and chat room discussions [13, 17]. Usage of roman text to write different languages adds another dimension to this complexity.

The following questions need to be worked upon and answered.

1) How to identify and analyze conversation of interest in Web forum discussions where more than one languages are used?

(2) Possibility of stylistic feature based sentiment classification .

(3) How feature selection can improve 1) and 2) above.

v) Criminal Network Detection and Analysis

Modern terrorists operate through global terrorist networks. Understanding the structure and organization of these terrorist networks is important to preempt any planned terrorist activity and also for investigation of any previous event of terrorism [16, 18]. Based on these investigations and understandings, effective strategies to prevent crimes can be devised. However, criminal network analysis is still a non-automotive process and extensive research is being done to automate terrorist network analysis. [12].

4. Role of Data Mining and Semantic Web

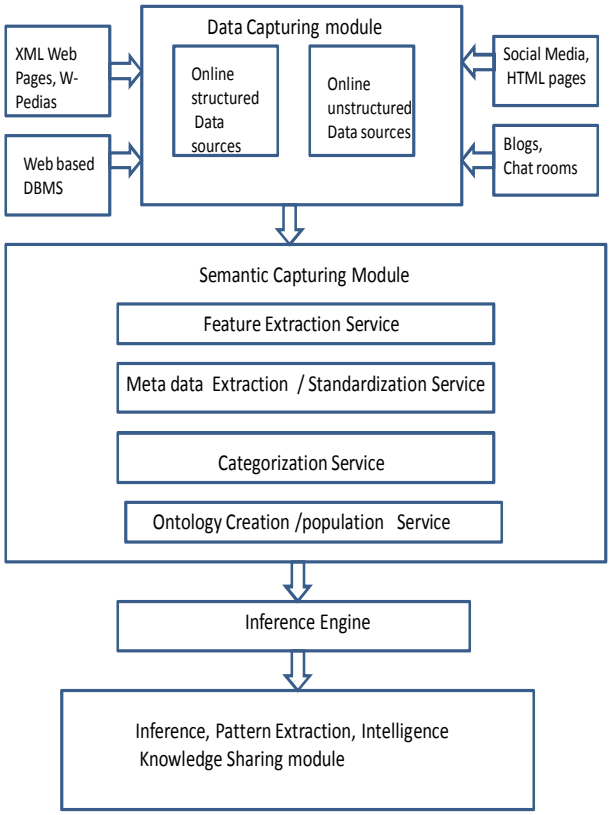
Research in data integration, data analysis, text mining, image and video processing, and evidence combination have resulted in valuable tools for pattern extraction and intelligence analysis [2]. With the employment of AI artificial intelligence and related technologies such as data mining, text mining, Web mining, natural language processing, planning, reasoning, conflict resolution, link analysis, and search algorithms, it is possible to intelligently decipher this data for effective and timely decision making.[5].

However employing these technologies is difficult as much of the information exists in disparate databases scattered among federal, provincial, and local entities and also because of the reasons mentioned earlier such as identifier conflicts and spatial distribution [6]. In many cases, these computer systems cannot share information-either "horizontally" (across the same level of government) or "vertically" (between federal, provincial, and local governments). This problem is further compounded because there is no proper and well defined distribution of responsibilities and scope for various government organizations. Databases used by security apparatus and public health services are not connected transparently enough to allow for interoperability [3]. This makes it difficult for government agencies to share information of interest with other agencies [4, 9]. Vast amounts of knowledge resident within each agency should be shareable at all levels of government with appropriate security, privacy and secrecy measures. There are two major problems that are endemic to the government agencies and are a hurdle in building an efficient government-wide information system. First, government procurement of information systems is not coordinated and standardized. Over time, hundreds of new systems may be acquired to address specific, ad hoc and immediate agency requirements. These may not be and are not in most cases compatible and interoperable with other IT resources of federal or provincial organizations. Organizations thus become compartmentalized and evolve into islands of technology [6]. Second, legal, linguistic, cultural and bureaucratic barriers often prevent agencies from exchanging and integrating information. Information- sharing capabilities are similarly deficient at the provincial and local/council levels. Terrorism, gang, and drug databases maintained by one province cannot be accessed by others. An added difficulty is usage of incompatible compatible communications equipment which in turn is the result of uncoordinated and unplanned policies. Wireless and communication technology used by various organizations within a country should be compatible [9].

Another application of data mining is to analyze paper documents that routinely used in general litigation and criminal and terrorist investigations. Currently OCR technique is used as state-of-the-art to process these documents. This limits the search strictly to the text. This ignores all handwriting, signatures, logos, images, watermarks, and any other non-text artifacts in a document. This limitation in document search can be improved by annotating documents with more meta-data. In case such meta-data is not present, techniques can be applied to extract key metadata from paper documents such as logos and signatures and match these against a set of known logos and signatures.

With the web content becoming richly-structured and heterogeneous, link mining is increasingly being used these kinds of data sets. Link mining employs link analysis, hypertext and web mining, relational learning and inductive logic programming, and graph mining. Link mining techniques are also being used to help construct semantic web. Also semantic and ontological information is helping in link mining endeavors as in case of methods for discovering interesting sub-graphs based on semantic information associated with the edges.

Semantic Web also offers promising solution to the problems mentioned above my facilitating semantic search across different domains. For realizing these benefits, the following issues must be addressed.



Security Informatics Framework

• Integrate information sharing across federal, provincial and local governments, private industry, and citizens
• Adopt common "meta-data" standards for electronic information relevant to security and law enforcement agencies. Identifier usage in context of cultural, technical, procedural, textual and semantic differences should be standardized.
• Improve emergency communications focusing on compatibility and interoperability.

5. Conclusion

Modern terrorist groups and other kinds of organized crime are the most potent threat to the world and are often geographically and temporally dispersed. We feel that the developed countries advantage in science and technology, especially in the IT domain, is a key to providing an effective response to major terrorist incidents and natural disasters. However, to use IT effectively for counter-terrorism, domain challenges, such as of data distribution, diverse data formats, data integration and interoperability need to be overcome and significant research is needed in disciplines of Artificial Intelligence, Knowledge Engineering, Data Mining and Semantic Web. In doing so, we can also expect many significant side benefits, as many of the investments to develop and deploy technologies for modernizing and upgrading our counter terrorism, security and law enforcement capabilities will result in significant economic, health, and public safety benefits. We feel that for an effective usage of these technologies, technologically advanced stake holders should help develop the related IT infrastructure in developing countries.

References

[1] D. Zeng, H. Chen, L. Tseng, C. Larron, M. Eidson, 1. Gotham. C. Lynch, and M. Ascher, "West Nile Virus and Botulism Pan & A case study in infectious disease informatics," in Intelligence and Security Informatics, ISI2004, ser. Lecture Notes in Computer Science, No. 3073. Springer, 2004.
[2] Office of the Homeland Security, The White House, National Strategy fro Homeland Security, July 2002.
[3] H. Chen, R. Moore, D. Zeng, and J. J. Leavitt, Eds., "Intelligence and Security Informatics for National And Homeland Security. Proceedings of the Second Symposium on Intelligence and Security Informatics (ISI'04)".. Lecture Notes in Computer

Science (LNCS) 3073. New York: Springer-Verlag, 2004.
[4] H. Garcia-Molina, J. D. Ullman, and J. Widom, "Database Systems, the Complete Book", Englewood Cliffs, NJ: Prentice-Hall, 2002.
[5] DeRosa M.," Data Mining and Data Analysis for Counter Terrorism", Center for Strategic and International Studies, CSIS Press, 2004.
[6] Hawthorne S., "Knowledge Related to a Purpose: Data- Mining to detect terrorism "Bulletin of the American Society for Information Science and Technology", Vol. 29 No. 1 October / November 2002
[7] Daily" The News", Pakistan, Printed October 12, 2005
[8] Jeffrey W. Seifert. CRS Report for Congress "Congressional Research Service ˜ The Library of Congress, Updated December 16, 2004
[9] National Research Council, "Making the Nation Safer: The Role of Science and Technology in Countering Terrorism. Washington", DC: Committee on Science and Technology for Countering Terrorism, U.S. National Research Council, The National Academies Press, 2002.
[10] Abbasi, A., Chen, H., and Salem, A. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. ACM Trans. Inform. Syst. 26, 3, Article 12 (June 2008), 34 pages. DOI = 10.1145/1361684.1361685 http://doi.acm.org/10.1145/1361684.1361685
[11] Chen, H. 2006. Intelligence and security informatics for international security: information Sharing and data mining. Springer, London.
[12]Glaser, J., Dixit, J., and Green, D. P. 2002. Studying hate crime with the internet: what makes racists advocate racial violence? J. Social issues 58, 1, 177–193.
[13]Burris, V., Smith, E., and Strahm, A. 2000. White supremacist networks on the internet. Sociol. Focus 33, 2, 215–235.
[14]Hearst, M. A. 1992. Direction-based text interpretation as an information access refinement in text-based intelligent systems: current research and practice in information extraction and Retrieval, p. Jacobs, ed. Lawrence Erlbaum associates, Mahwah, NJ.
[15]Wiebe, J., Wilson, T., and Cardie, c. 2005. Annotating expressions of opinions and emotions in Language. Lang. Resources eval. 1, 2,165–210
[16] Wiebe, J. 1994. Tracking point of view in narrative. Comput. Linguist. 20, 2, 233–287.
[17]Schafer, J. 2002. Spinning the web of hate: web-based hate propagation by extremist organizations. J. Criminal just. Popular culture 9, 2, 69–88.
[18]Crilley, K. 2001. Information warfare: new battle fields, terrorists, propaganda, and the internet. Aslib proc. 53, 7, 250–264.
[19] Zhou, Y., Reid, E., Qin, J., Chen, H., and Lai, G. 2005. U.S. extremist groups on the web: link and Content analysis. IEEE Intel. Syst. 20, 5, 44–51.
[20] Jennifer J. Xu and Hsinchun Chen crime net explorer: a framework for criminal network knowledge discovery. ACM transactions on information systems, vol. 23, No. 2, April 2005, pages 201–226.
[21]Chen, H., Zeng, D., Atabakhsh, H., Wyzga, W., and Schroeder,J. 2003. Coplink: managing law Enforcement data and knowledge. Commun. ACM 46, 28–34.
[22]Saether, M. and Canter, D. V. 2001. A structural analysis of fraud and armed robbery networks In Norway. In proceedings of the 6th international investigative psychology conference (Liverpool, UK, Jan.).
[23] Gang Wang, Hsinchun Chen, and Homa Atabakhsh automatically detecting deceptive criminal identities communications of the acme march 2004/vol. 47, no. 3 Page 70-76.
[24] Press Release COUNCIL OF THE EUROPEAN UNION The European Union Counter-Terrorism Strategy Justice and Home Affairs Council meeting, Brussels 1 December 2005R u e d e l a L o i 1 7 5 B – 1 0 4 8 B R U S S E L S http://ue.eu.int/Newsroom
[25]COUNCIL OF THE EUROPEAN UNION Implementation of the Action Plan to Combat Terrorism Justice and Home Affairs Council meeting, Brussels 1 December 2005.
[26]Accessed 3-6-2008 http://wps.prenhall.com/bp_turban_ec_2008/79/20296/5195785.cw/index.html.

11/15/2013