# A Study on Significance of Event Ontology Approach in Web Crime Mining

Megha Mudholkar
*Research Scholar, Thadomal Shahani Engineering College,*
*Mumbai, Maharashtra, India*

Ujwala Bharambe
*Asst. Professor, Thadomal Shahani Engineering College,*
*Mumbai, Maharashtra, India*

**Abstract- As most aspect of our life move to digital networks, crimes comes with them. Our lives increasingly depend on the internet and digital networks, but these create new vulnerabilities and new ways for criminals to exploits the digital environment. Not only can many existing crimes be replicated in online environments, but novel crimes that exploit specific features of digital networks have emerged as well. With new crimes come new forms of policing and new forms of surveillance and with these come new dangers for civil liberties. These kinds of cybercrime information present on web pages are in the form of text. Because a lot of crime information in documents is described through events, event-based semantic technology can be used to study the patterns and trends of web-oriented crimes. So for cyber crime mining, event ontology is constructed to extract the attributes and relations in web pages and reconstruct the scenario for crime mining.**

**Keywords – Digital Networks, Web Crime Mining, Event Ontology.**

## I. INTRODUCTION

The internet in India is growing rapidly. It has given rise to new opportunities in every field we can think of – be it entertainment, business, sports or education. The Internet is useful way for people to conduct business effectively, very quickly. It saves businesses a lot of time, money and resources. Unfortunately, it is also an open invitation to scamsters and fraudsters and online frauds are becoming increasingly rampant. There are two sides to a coin. Internet also has its own disadvantages. One of the major disadvantages is Cybercrime – illegal activity committed on the internet. The internet, along with its advantages, has also exposed us to security risks that come with connecting to a large network. Computers today are being misused for illegal activities like e-mail scams, credit card fraud, spams, software piracy and so on, which invade our privacy and offend our senses. Criminal activities in the cyberspace are on the rise.

A lot of facts have proved that it is not enough to manage the information on the Internet simply through traditional administrative models. In this concern, Web mining is a novel research direction for the information gathering and analyzing on the Internet, which is explosive and unstructured. The focuses of Web mining research are to develop new web mining techniques and to extract the features of texts to represent them [1].

Cyber Crime is an evil having its origin in the growing dependence on computers in modern life. Along with the rapid popularity of the Internet, crime information on the web is becoming increasingly rampant, and the majority of them are in the form of text. Because a lot of crime information in documents is described through events, event-based semantic technology can be used to study the patterns and trends of web-oriented crimes. So for cyber crime mining, event ontology is constructed to extract the attributes and relations in web pages and reconstruct the scenario for crime mining.

A major challenge facing all law-enforcement and intelligence-gathering organizations is accurately and efficiently analyzing the growing volumes of crime data. For example, complex conspiracies are often difficult to unravel because information on suspects can be geographically diffuse and span long periods of time. Detecting cyber crime can likewise be difficult because busy network traffic and frequent online transactions generate large amounts of data, only a small portion of which relates to illegal activities. In this concern, Web mining is a novel research direction for the information gathering and analyzing on the Internet, which is explosive and unstructured. The focuses of Web mining research are to develop new web mining techniques and to extract the features of texts to represent them.

This work explores the cyber crimes in various web pages by event ontology construction. Such a mechanism is designed, implemented and evaluated that allows us to define and mine interested information. In particular, event ontology is defined and demonstrates how it can be used to describe cyber crimes on the level of event, relation and event class. An event ontology based cyber crime mining procedure is proposed that concerns web texts prehandling, feature extraction, feature reconstruction and crime mining. The paper discusses scope of constructing event ontology for web crime mining and it focused on extracting attributes and relations in web pages.

## II. WEB CRIME MINING

Cyber crime is on the rise, not only in terms of the number of perpetrators and the volume of crimes committed, but also of the range of techniques employed to carry them out. In this environment, malicious code offers a valuable resource for those wishing to perform attacks through the Internet and information technology.

Table 1 lists eight crime categories on which local and federal authorities maintain data, ordered by their increasing degree of harm to the general public. We devised these categories, which include numerous offenses classified by different law-enforcement agencies in various ways, in consultation with a local detective with more than 30 years of experience.

Some types of crime, such as traffic violations and arson, primarily concern police at the city, investigated by local law-enforcement units as well as by national and international agencies. county, and state levels. Other crime types are For example, a city police department's sex crimes unit may track local pedophiles and prostitutes, while the FBI and the International Criminal Police Organization focus on transnational trafficking in children and women for sexual exploitation.

Many crimes, such as the theft of nuclear weapons data, can have profound implications for both national and global security. Transnational fraud and trafficking in stolen property or contraband can severely impact trade, business, and government revenue. Local gangs as well as foreign-based drug cartels and criminal organizations exact a large financial cost as well as threaten public health and safety. Although most types of violent crime—such as murder, robbery, forcible rape, and aggravated assault—are local police matters, terrorism is a global problem that relies on cooperation at all levels of government. The Internet's pervasiveness likewise makes identity theft, network intrusion, cyberpiracy, and other illicit computer-mediated activities a challenge for many law-enforcement bodies.

| **Crime type** | **Local law enforcement** | **National and international security** |
|---|---|---|
| Traffic violations | Speeding, reckless driving, causing property damage or personal injury in a collision, driving under the influence of drugs or alcohol, hit-and-run, "road rage" | -- |
| Sex crime | Sexual abuse, rape, sexual assault, child molestation, child pornography, prostitution | Trafficking in women and children for sexual exploitation, including prostitution and pornography |

| | | |
|---|---|---|
| Theft | Robbery, burglary, larceny, motor vehicle theft | Theft of national secrets or weapon information, illicit trafficking in stolen art and vehicles |
| Fraud | Money laundering, counterfeiting, insurance fraud, corruption and bribery, misappropriation of assets | Transnational money laundering, fraud, and corruption; trafficking in stolen software, music, movies, and other intellectual property |
| Arson | Intentionally setting fires to damage property, such as a warehouse or apartment building | -- |
| Gang/drug offenses | Possessing, distributing, and selling illegal drugs | Transnational drug trafficking, organized racketeering and<br>extortion, people smuggling |
| Violent crime | Murder, aggravated assault, armed robbery, forcible rape, hate crime | Terrorism, air and maritime piracy, bombings hate crime |
| Cyber crime | Internet fraud, such as credit card and advance fee fraud, fraudulent Web sites, and illegal online gambling and trading; network intrusion and hacking; virus spreading; cyberpiracy and cyberterrorism; distributing child pornography; identity theft | |

Some preliminary results on crime analysis have been made through using data mining techniques. Traditional data mining techniques such as association analysis, classification and prediction, cluster analysis, and outlier analysis identify patterns in structured data. 3 Newer techniques identify patterns from both structured and unstructured data. As with other forms of data mining, crime data mining raises privacy concerns. 4 Nevertheless, researchers have developed various automated data mining techniques for both local law enforcement and national security applications.

Entity extraction identifies particular patterns from data such as text, images, or audio materials. It has been used to automatically identify persons, addresses, vehicles, and personal characteristics from police narrative reports. 5 In computer forensics, the extraction of software metrics 6 —which includes the data structure, program flow, organization and quantity of comments, and use of variable names—can facilitate further investigation by, for example, grouping similar programs written by hackers and tracing their behavior. Entity extraction provides basic information for crime analysis, but its performance depends greatly on the availability of extensive amounts of clean input data.

## III. EVENT ONTOLOGY

Ontology has a long history in philosophy. It was introduced to computer science 10 years ago as a method of knowledge representation. For event information, many researchers put forward the idea of event-oriented ontology expansion. An event usually relates to many-sided elements such as time, locations and objects, which is a knowledge unit with a higher granularity compared with the concept. An event is a changing fact with the different times, and there are inherent connections between events. An event ontology is a shared, objective and formal specification of an event class system model.

An ontology is a 'formal, explicit specification of a shared conceptualization', where a 'conceptualization' is an abstract model of some phenomenon of the world which identifies the relevant concepts (or entities) and relations between the concepts of that phenomenon [7]. The conventional ontology only reflects the existing law of objects, especially classified and non-classified relations between objects, but it has obvious defects:

(1) The storing and understanding unit of human beings is ‘event’ for the real world, and the event relates to many-sided concepts, which is a knowledge unit with a higher granularity compared with the concept. The concept model used by the conventional ontology can not reflect the higher and more complex semantics that an event possesses. Therefore, the conventional ontology lacks higher level structures.

(2) The conventional ontology focuses on describing concepts and their relations. The concept is the set of objects which possess the same attributes. The object is represented by attributes, and the conventional ontology is suitable for representing static concepts. But the event is dynamic, and its state is changing. In the conventional ontology, the events are regarded as concepts or relations. Applying the method of describing concepts to denote event classes ignores the dynamic characteristics of event classes. And employing the method of describing relations to represent event classes not only ignores the dynamic characteristics of event classes, but neglects the other elements.

Event features have bigger granularity than terms or characters, it can sharply reduce the features vector space‘s dimensionality in automatic classifier. Events have much more semantic information than terms, and events are more practical and suitable for computers to simulate human cerebral, so it can improve the classification in terms of accuracy.

## IV. Constructing Event Oriented Web Crime Ontology

### *A. Categories of cyber crimes*

According to the decision adopted by ‘The National People's Congress Standing Committee's decision on safeguarding Internet security’ on December 28, 2000, the crimes taking the Internet as tools can enerally be divided into the following categories: (1) fraud; (2) Internet pornography; (3) illegal rade; (4) false advertising; (5) violations of privacy; (6)abetting, inciting all kinds of crime; (7 ) network gambling; (8) damage to reputation, credibility; (9) intimidation, blackmail; (10) tort; (11) insulting, slandering; and (12) others, such as forgery, illegal organizations, etc. However, the above cyber crimes are always reflected on the web by different styles.

### *B. Constructing event-oriented cyber crime ontology*

#### ***1****) Events and their relations*

The basis for constructing event ontology is events and their relations. What elements does an event include? Different applications have different definitions. For cyber crimes, the analyst is most concerned about types of events (denoted by actions), participants, time, location, instruments, involving goods, etc. Therefore, we use the event structure of a six-tuple.

Definition 1. An event refers to one thing happening at a specific time and location, involving a number of actors and goods, and using some instruments. An event can be defined as a six-tuple: e = (A, P, T, L, F, G).

The elements of an event six-tuple are called event elements, which represent action (A), participant (P), time (T), location (L), instrument (F) and good (G) respectively. And the participant involves subjects and objects.

Definition 2. Event relations are divided into two categories: classified relations and non-classified relations. Furthermore, non-classified relations are divided into the following categories: component relation, cause-effect relation, follow relation, and concurrence relation [14].

Definition 3. An event class (EC) refers to the set of events which have the same characteristics.

#### ***2)*** *An Example of Event Ontology*

In project, Event-oriented cyber crime ontology containing 185 event classes is constructed. Figure1 is a fragment of the event ontology containing five such event classes and their relations.

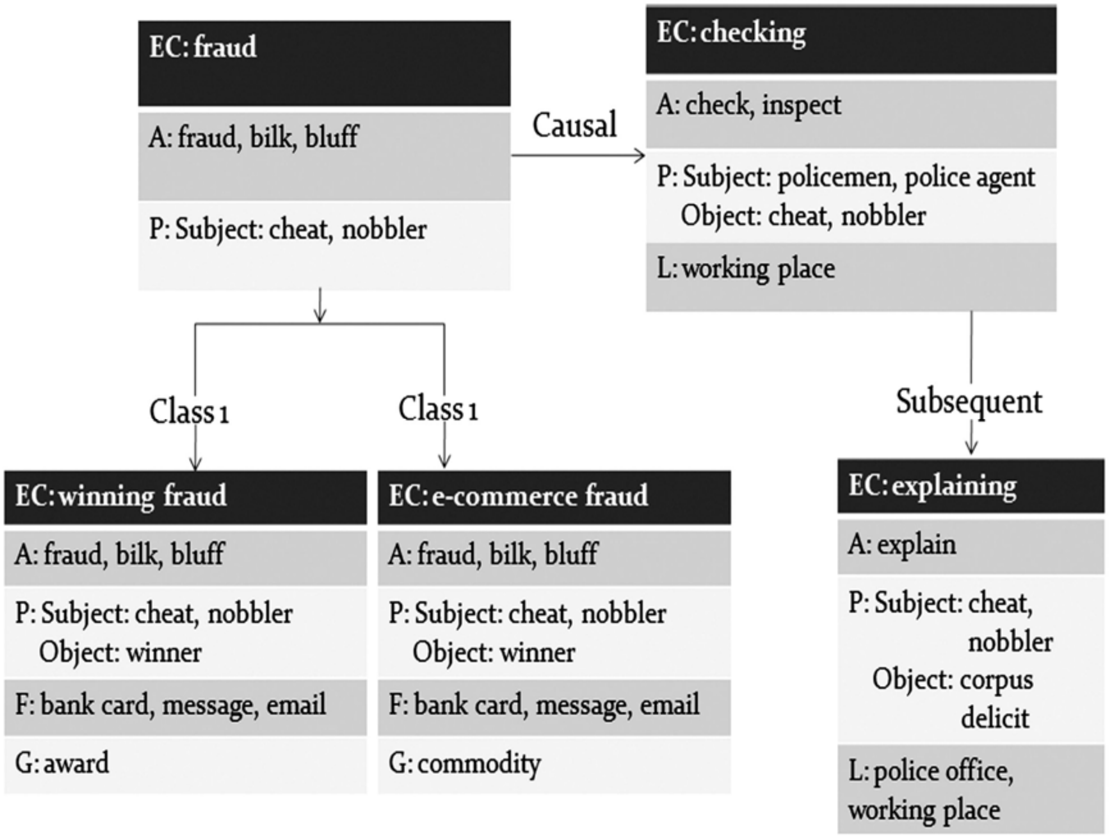

Figure 1. A fragment of Event Ontology for Cyber Crimes

In Figure 1, five domain event classes are shown: 'fraud', 'winning fraud', 'ecommerce fraud', 'checking', and 'explaining'. Usually, 'fraud' will result in 'checking', and 'explaining' follows 'checking'. Because 'winning fraud' and 'e-commerce fraud' are the sub classes of 'fraud', they will inherit the attributes of 'fraud'.

Protégé 4.1 is used as a tool to construct cyber crime event ontology. Protégé OWL is used as ontology language. Fig. 2 is a fragment of cyber crime event ontology hierarchy that is constructed using OWLViz Option in Protégé 4.1.

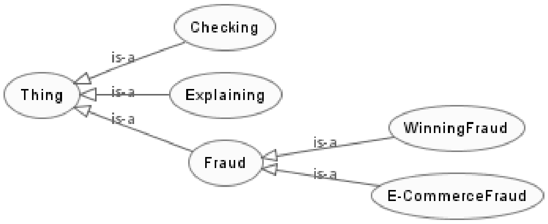

Figure 2. OWLViz Displaying the Hierarchy for cyber crime

For implementation of cyber crime mining, event ontology is constructed to extract the attributes and relations in web pages and also reconstructed the scenario for crime mining. Event ontology and Support Vector Machine

(SVM) as well as Naïve bayes classifications are used to identify different types of cyber crimes. Figure 3 presents the overall process of web crime mining based on event ontology.

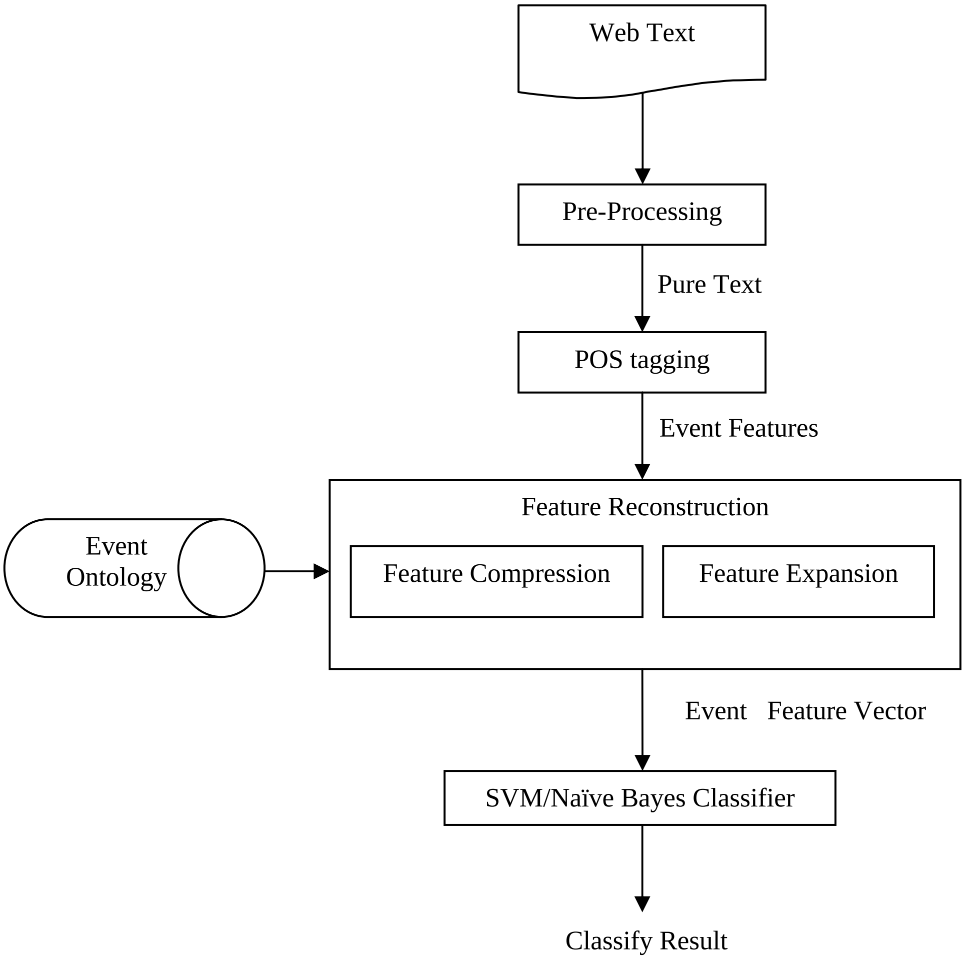

Figure 3. The overall process of web crime mining based on event ontology

The process in Figure 3 mainly includes the following stages.

**Stage 1: Pre-processing web texts (extracting the web pages content)**

Most of the web texts are Html format. After pre-processing, filtering URL, advertising and other information, the pure texts are reserved.

**Stage 2: Obtaining candidate features (extracting event features from texts)**

After POS tagging texts and filtering stop-words, the remaining words are candidate features of pure texts.

**Stage 3: Feature reconstruction (representing the text by event features)**

Feature reconstruction consists of feature compression and feature expansion. Feature compression refers to merge some features using ontology or thesaurus. And feature expansion refers to append a number of features using ontology.

**Stage 4: Data mining**

It contains many techniques such as clustering, classification, retrieving, associative discovery, and so on.

## V. Importance Of Using Event Ontology

In the process of web crime mining based on event ontology, event ontology is used for feature reconstruction. Feature reconstruction comprises of feature compression and feature expansion.

### *A. Feature Compression-*

The features of a document are high-dimensional and sparse. Compressing features for documents not only improves the speed, but improves the precision of text processing. For instance, merging 'drug users', 'drug addicts 'and 'drugger' as the concept 'drug users', which not only compresses feature dimension but increases the weight of feature 'drug users'.

The elements of an event such as subject, object, location, have different forms in language, and many are synonymous concepts, such as the action of event class '(fraud)': '(fraud)' and '(cheat)'. Using event ontology, we can compress feature dimension through merging some features.

### *B. Feature Expansion-*

Some documents are easier to be processed only after expanding their features. Such is divided into two categories: (1) Short document, and (2) omit. On one hand, a lot of documents related to web crimes are very short. For example, we have received a message on illicit selling invoice, "Invoice to come forward and contact me, a lot of concessions". As the message is too short, the information described is very weak, and it is hard to mine such information. On the other hand, depicting the events in the document, some words are often omitted because of language habits.

If the document contains an event, even lacking some elements, we can also conclude what the omitted elements are through using event ontology. According to event relations, other events can also be associated. For instance, the sentence 'After his arrest, he accepted the trial', for event class 'inquest', we can know the subject of 'inquest' is 'police'. And for retrieving event 'Internet pornography', we can associate participants such as 'prostitute' and 'customer'.

The method of feature expansion is only used in the step of SVM classification, and only short documents are selected to expand features. After analyzing some documents, we find that it is inaccurate to expand features just according to whether a document contains an event. For example, the event '(checking)' may exist in a variety of documents such as student's '(checking)' and leader's '(checking)'. And meanwhile, if some events '(arrest)', '(checking)' or event participant '(police)' exist in the same document, then the document will relate to crimes. The document expanded should meet two rules: (1) its length is less than 50 words, and (2) there are three or more than three event elements in it. The method of expanding the document is to append other elements of event class contained by both the document and event ontology. The weights of appending elements are the weight of the event in the document.

## VI. Conclusion

The study takes an event as the basic semantic unit for text processing, and also focused on the method of Web crime mining. On the basis of constructing event-oriented cyber crime ontology, the event ontology used as the priori knowledge and propose a method of reconstructing text features based on event ontology.

# VII. REFERENCES

[1] Li Cunhua, Hu Yun, Zhong Zhaoman, "An Event Ontology Construction Approach To Web Crime Mining", 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010),2010 Page No.2441-2445.

[2] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar, "Web Mining - Concepts, Applications & Research Directions", http://www-users.cs.umn.edu/~desikan/publications/wmo.pdf, Page No. 51-71.

[3] Hsinchun Chen, Wingyan Chung, Yi Qin,"Crime Data Mining: An Overview and Case Studies", Proceedings of the 2003 annual national conference on Digital government research, Boston, M.A, 2003, Page No.1-5.

[4] Hsinchun Chen, Wingyan Chung, Jennifer Jie Xu, Gang Wang Yi Qin, Michael Chau,"Crime data mining: A general framework and some examples", IEEE Computer society, April 2004, Page No.50-56.

[5] T. Abraham and O. de Vel. "Investigative profiling with computer forensic log data and association rules", Proc. of the IEEE International Conference on Data Mining (ICDM'02), 2006, Page No.11 - 18.

[6] J.S. de Bruin, T.K. Cocx, W.A. Kosters, J. Laros and J.N. Kok. "Data mining approaches to criminal career analysis", Proc. of the Sixth International Conference on Data Mining (ICDM'06), 2006, Page No.171-177.

[7] S.V. Nath. "Crime pattern detection using data mining", Proc. of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2006, Page No.41- 44.

[8] Noy N, McGuinness D., "Ontology development 101: a guide to creating your first ontology", http://protege.stanford.edu/publications/ontology_development/ ontology101-noy-mcguinness.html, 2001.

[9] Gruber T.,"What is an Ontology?", http://www-ksl.stanford.edu/kst/what-is-an-ontology.html, 2006.

[10] Matthew Horridge,Simon Jupp, Georgina Moulton, Alan Rector, Robert Stevens, Chris Wroe, "A Practical Guide To Building OWL Ontologies Using Protege4.1tools",http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_1.pdf ,2007,Page No.1-102

[11] Dzemydiene, D.," Knowledge Representation in Advisory Information System of Crime Investigation Domain", Databases and Information Systems II, Springer, Heidelberg, 2002, Page No.135–146. (REF PAPER NOT THERE)

[12] Dzemydiene, D., Kazemikaitiene, E.,"Ontology-Based Decision Support System for Crime Investigation Processes", Information Systems Development, Springer, Heidelberg, 2005, Page No. 427–438. (REF PAPER NOT THERE)

[13] Donalds, C.M., Osei-Bryson, K.,"Criminal Investigation Knowledge System: CRIKS",The 39th Annual Hawaii International Conference on System Sciences, vol. 7, 2006, Page No. 152–160

[14] Brinson, A., Robinson, A., Rogers, M., "A cyber forensics ontology: Creating a new approach to studying cyber forensics", Digital Investigation, 2006, Page No.S37–S43.

[15] Heum Park, SunHo Cho, and Hyuk-Chul Kwon, "Cyber Forensics Ontology for Cyber Criminal Investigation", M. Sorell (Ed.): e-Forensics 2009, LNICST 8, Page No. 160 – 165.

[16] Grigoris Antoniou, Frank van Harmelen, "A Semantic Web Primer", The MIT Press Cambridge, Massachusetts London, England, Page No.10-20

[17] G.P.Zarri, "Semantic Web and Knowledge Representation", Proc. of the 13th International Workshop on Database and Expert Systems Applications (DEXA'02), 2002, Page No.1529-4188.

[18] H.F.Lin and J. M. Liang, "Event-based Ontology design for retrieving digital archives on human religious self-help consulting", Proc. of 2005 IEEE International Conference on e-Technology, e-Commerce and e-Service, 2005, Page No. 522-527.

[19] WANG Li, LIU Zong-Tian, WANG YanHua , SUN Rong, LIU HF, "Event Feature and Personality - Event – Ontology Based for Classifying Chinese Web Pages", Second International Workshop on Computer Science and Engineering,2009,Page No.555-557.

[20] Zhaoman Zhong , Zongtian Liu , Cunhua Li Yan Guan, "Event ontology reasoning based on event class influence factors" International Journal Of Machine Learning & Cybernetics, Volume 3, Issue 2, June 2012,Page No. 133-139

[21] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, "A Practical Guide to SupportVectorClassification",http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf,2010.

[22] Simon Haykin,"Neural Networks-A Comprehensive Foundation", LPE Pearson Education Asia, Second Edition, Page No.318-340.

[23] S.N. Sivanandam, S. Sumathi, S. N. Deepa,"Introduction to Neural Networks using MATLAB 6.0",Tata McGraw Hill, Page No.-340-343.

[24] Meijuan Gao, Jingwen Tian, Meijuan Gao, Jingwen Tian,"Research of Web Classification Mining Based on Classify Support Vector Machine",ISECS International Colloquium on Computing, Communication, Control, and Management,2009,Page No.21-24

[25] Khin Phyu Phyu Shein, Thi Thi Soe Nyunt," Sentiment Classification based on Ontology and SVM Classifier", Second International Conference on Communication Software and Networks, 2010, Page No. 169-172.

[26] Vidhya.K.A, G.Aghila, "A Survey of Naïve Bayes Machine Learning approach in Text Document Classification", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 2, 2010,Page No. 206-211.

[27] S.L. Ting, W.H. Ip, Albert H.C. Tsang, "Is Naïve Bayes a Good Classifier for Document Classification?", International Journal of Software Engineering and Its Applications Vol. 5, No. 3, July, 2011,Page No.37-46.

[28] Dino Isa, Lam Hong, Lee, V. P. Kallimani, R. Rajkumar, ”Text Document Preprocessing with the Bayes Formula for Classification Using the Support Vector Machine”, IEEE Transactions on Knowledge and Data Engineering, Vol.20,Issue 9,Sept-2008,Page No. 1264-1272.

[29] Dan Mayer,“Text2SVM”,http://mayerdan.com/2004/03/09/Text2SVM.

[30] Liu Xiaopeng, Xing Changzheng, “Research on Web Content Mining”, Computer and digital engineering, vol. 33, no. 9,2005, Page No.75-79.