Cyber Forensics Ontology for Cyber Criminal Investigation

Heum Park, SunHo Cho, and Hyuk-Chul Kwon

AI Lab. Dept. of Computer Science, Pusan National University, Busan, Korea
parkheum2@empal.com, sean@pusan.ac.kr, hckwon@pusan.ac.kr

Abstract. We developed Cyber Forensics Ontology for the criminal investigation in cyber space. Cyber crime is classified into cyber terror and general cyber crime, and those two classes are connected with each other. The investigation of cyber terror requires high technology, system environment and experts, and general cyber crime is connected with general crime by evidence from digital data and cyber space. Accordingly, it is difficult to determine relational crime types and collect evidence. Therefore, we considered the classifications of cyber crime, the collection of evidence in cyber space and the application of laws to cyber crime. In order to efficiently investigate cyber crime, it is necessary to integrate those concepts for each cyber crime-case. Thus, we constructed a cyber forensics domain ontology for criminal investigation in cyber space, according to the categories of cyber crime, laws, evidence and information of criminals. This ontology can be used in the process of investigating of cyber crime-cases, and for data mining of cyber crime; classification, clustering, association and detection of crime types, crime cases, evidences and criminals.

Keywords: ontology, cyber crime, digital evidence, criminal investigation, cyber forensics.

1 Introduction

Crime has increased amid the explosion of information technology, Internet services and digital equipments, as criminals have used those tools and environments on the cyber space as well as in the real word. Typical cyber crimes are 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; cyber piracy and cyber terrorism; child pornography distribution; identity theft. The Internet's pervasiveness likewise makes identity theft, network intrusion, cyber piracy, and other illicit computer-mediated activities a challenge for many law-enforcement agencies. [1],[6] It is difficult to collect evidence in cyber space, to investigate crime in cyber space, and to connect cyber evidence together with general evidence. Thus, professional engineers are required for collection of cyber evidence as well as crime analysis. Recently, cyber crime and digital data forensics have been studied for the purposes of cyber criminal investigation and data mining of cyber crime. In investigating cyber

M. Sorell (Ed.): e-Forensics 2009, LNICST 8, pp. 160 – 165, 2009.

crime, it is necessary to collect cyber evidence and digital equipment as evidence, to investigate connections with general crime, to classify documents, and to apply relevant relational laws. In addition, data mining technology is requisite. Detecting cyber crime can likewise be difficult because heavy network traffic and frequent online transactions generate huge amounts of data. Thus, data mining of crime (including cyber crime) can be a powerful tool enabling criminal investigators who may lack extensive training as data analysts to explore large databases quickly and efficiently [1]. For the efficient investigation of cyber crime, it is necessary to integrate the various cyber crime concepts.

Recently, knowledge representation systems and information systems using cyber forensics ontology have been studied for criminal investigations and evidences investigation process. However, those studies have concerned only general criminal cases and digital evidence, and have focused exclusively on database-based retrieval systems. In 2007, the cyber crime forensics ontology was presented by A. Brinson, A. Robinson and M. Rogers. The purpose of this ontology was to find specialization, certification, and education within the cyber forensics domain [2]. However, this ontology was designed only for specialization, certification, and education within the cyber forensics domain, thus it is necessary to apply real-world investigation of cyber crime and data mining of crime cases.

Therefore, we focused on building a cyber crime forensics ontology for investigation of real-world cyber crime that can be applied to data mining of cyber crime. We constructed the ontology by including the categories cyber crime, evidence, laws, information on criminals and crime cases, based on the cyber crime in *the cyber division of the Korean National Police Agency (KNPA: http://www.netan.go. kr)*. In the following Section 2, we discuss related studies concerning ontology, the existing cyber forensics ontology, and data mining technology. In Section 3, we introduce the cyber crime forensics ontology for cyber criminal investigation, along with the areas in which the ontology can be applied. In Section 4, we draw conclusions.

2 Related Studies

Ontology for the present context was originally proposed in 1992 by Tom Gruber who defined it as “a specification of a conceptualization.” The word "ontology" seems to generate a lot of controversy in discussions about AI. It has a long history in philosophy, in which it refers to the subject of existence. That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. The definition is consistent with the usage of ontology as set of concept definitions, but is more general. The representational primitives are typically classes, attributes, and relationships. The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application. [3] [4].

The most recent development in standard ontology languages is OWL from the World Wide Web Consortium (W3C). OWL is used for formal description of conceptual meanings and relations. Like Protégé OWL makes it possible to describe concepts but it also provides new facilities. Also, many ontology building tools are introduced and used. OWL ontologies may be categorized into three species or sub-languages: OWL-Lite, OWL-DL and OWL-Full. A defining

feature of each sub-language is its expressiveness [5]. We created the ontology using OWL-DL as a description language and Protégé as a tool.

D. Dzemydiene presented Knowledge Representation in Advisory Information System of Crime Investigation Domain in 2002 and also developed a helpful criminalistics information system. He used an ontology by concepts and relations of crimes; serious crime, theft, illegal keeping of firearm, and others, all of which suggests a forensic intelligence process for criminal investigation [9]. In addition, D. Dzemydiene and E Kazemikaitiene proposed an Ontology-Based Decision Support System for Crime Investigation Processes in 2005, an ontology helps to create the framework and thus to ensure the collection, accumulation, storage, treatment, and transmission, in proper form, of important investigation information, which establishes the conditions to make optimal decision in criminal investigation [8].

C. M. Donalds and K. Osei-Bryson proposed the Criminal Investigation Knowledge System (CRIKS) in 2006, in order to assist security forces in gathering, storing and easily retrieving of information/intelligence/knowledge and reports on criminal activities obtained from members of the public and other local and overseas security forces. They created a domain ontology OntoCRIKS for the criminal investigation, by defining and identification of related concepts and relationships of the ontology can disambiguate idea of concepts and relationships in an organization. The concepts and relationships of this ontology were derived from document base, case base, operations base, discussion base, cognitive base and scenario base, and they applied the text mining technique [7].

The cyber forensics ontology of Ashley Brinson et al was created for the purpose of finding the correct layers for specialization, certification, and education within the cyber forensics domain in 2007. This ontology consisted of tow subtopic; technology and profession, and has five layers; Hardware, Software, Law, Academia, and Military and Private Sector. The hardware section of this model should be broken into Large Scale Digital Devices, Small Scale Digital Devices, computers, and others. The software section of this model contains three categories: analysis tools, operating systems, and file systems. The law section focuses on law enforcement and the involvement of the court and legal system within a cyber forensic investigation. The academia section focuses on curriculum development track within the ontology. The military category focuses on what cyber forensic duties military personnel perform. The military has many needs including data protection, data acquisition, imaging, extraction, interrogation, normalization, analysis, and reporting. The private sector was broken down into consulting and industry [2]. However, it was designed only for specialization, certification, and education within the cyber forensics domain.

H. Chen et al. introduced data mining for general crime applicable to entity extraction, clustering, association, deviation detection, classification, string comparison and social network analysis. In addition, they introduced a general framework for crime data mining that draws on experience and various proven techniques to analyze different types of crimes. Significantly, understanding the relationship between analysis capability and crime type characteristics can help investigators to more effectively use those techniques to identify trends and patterns, address problem areas, and even predict crimes [1]. We can apply the ontology to the data mining of cyber crime, classification, association, detection, clustering and retrieval systems.

3 Cyber Forensics Ontology for Cyber Criminal Investigation

In criminal investigations, first, the process by which the investigation of the crime will be conducted is planned, the crime scene and environment are secured, and evidence is collected by processes. Next, the evidence is examined and analyzed, and the outlines of the crime-case and the exact crime type are discerned. Then, write documents and the pertinent relational laws are determined, and the criminal is arrested [8][9]. Based on the criminal investigation process guidelines of the KNPA (Korea), in the process of investigating cyber crime, first, the scene is secured and evidences (volatile evidence and nonvolatile evidence in both cyber space and the real word) are collected according to the evidence-collection processes (or rules). Second, that evidence is examined and analyzed, after which it is securely stored. Third, the relational laws are applied to the crime for prosecution, and reports of the crime case are written. In cyber forensics, the crime type, the relational laws and the cyber crime type are closely related to each other, so relational crime types and laws can be found easily in the process of collecting evidences.

Accordingly, we defined the concepts and relations among crime types, evidence collection, criminals, and crime case and law, based on the processes of the cyber criminal investigations. We constructed a cyber crime forensics ontology for the cyber criminal investigation, which ontology consists of five concepts; Law, Crime_Case, Criminal, Crime_Type and Evidence. Figure 1 shows a concept diagram of these subclasses. The class 'Cyber_Crime' has the subclass 'Crime_Case' with the property 'hasCrimeCase', along with the subclass 'Law' with the property 'hasBaseLaw'. 'Law' is a collection of relational laws, and the class 'Crime_Case' has the subclasses 'Crime_Type', 'Evidence' and 'Criminal'. In addition, the domain 'Crime_Case' is linked to the range 'Criminal' with the property 'hasCriminal' and conversely with the property 'hasCrimeCase'. Also it linked to the range 'Evidence' with the property 'hasEvidence' and to 'Crime_Type' with the property 'hasCrimeType'. 'Crime_Type' has the subclasses 'CyberTerror' and 'GeneralCrime'.

Law. The Class 'Law' has subclasses 'Criminal Act', 'Act on Promotion of Information and Communications Utilization and Information Protection, etc', 'Information Communication Infrastructure Protection Act', Protection of Communications Secrets Act', 'Framework Act on Telecommunications', 'Telecommunications Business Act', 'Location Information Protection Act', 'Copyright Act', 'Radio Waves Act',

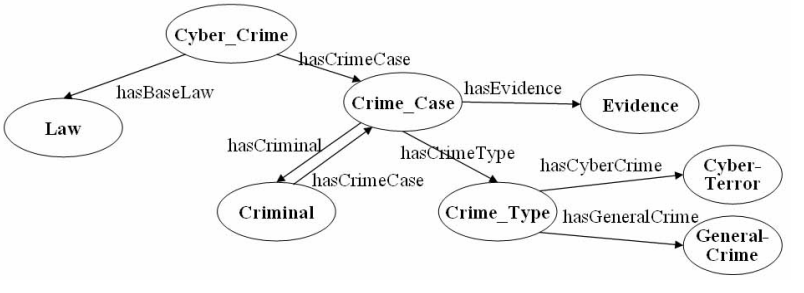

Fig. 1. Concept diagram of cyber forensics ontology for investigation of cyber crime

'Digital Signature Act' and 'Computer Programs Protection Act'. And each subclass has individuals of provisions and items.

Crime_Case. The 'Crime_Case' class includes subclasses 'Crime_Type', 'Evidence', 'Criminal' and 'Law' from above concepts diagram. This class is connected with all concepts; Cyber_Type, Law, Evidence and Criminal.

Criminal. The 'Criminal' class includes Evidence, Crime type and relational Laws as crime information through the Crime_case class. This class is connected all concepts through another concept, and has individuals of criminal.

Crime_Type. The Crime_Type class is a compendium of crime types, it consist of subclasses; Cyber Terror and General Cyber Crime. The Cyber Terror is broken down into three sub classes; Hacking, Distribution Virus and Distribution Spam. Hacking includes the subclasses; Intrusion, Dos, Information Theft, Logical bomb, and others. Distribution Virus has the subclasses; Worm, Virus, Spyware and Adware. Distribution Spam has the subclasses; Mail, Message and Call. As for General Cyber Crime, it is broken down into nine classes; Fraud, Illegal Site, Illegal Reproduction, Defamation, Infringement of Private, Stalking, Sexual Violence, Threatment and CopyRights. Each subclass has individuals.

Evidence. The 'Evidence' class is a collection of processes and evidence types, and has subclasses 'Collection' and 'EvidenceType'. Evidence type has subclasses 'Nonvolatile Evidence' and 'Volatile Evidence'. Nonvolatile evidence includes H/W; computer, scanner, Disk, CD, memory stick, and others, and S/W; illegal S/W, application program, hacking tool et al. Volatile evidence includes memory dump, CPU log, routing history, Disk imaging and blog board et al. Cyber evidence is a compendium of data and history in the cyber space and digital equipments (parts).

Figure 2 shows the subclasses and relations of evidence class. The 'Collection' class has subclass 'OrderedProcessSet' and 'CollectionInfo', and 'CollectionInfo' has subclasses 'Specialist', 'Investigator', 'Time' and 'Location', and 'OrderedProcessSet' has subclass 'Process'; process of collection evidence. 'Process' has subclasses 'takePicture', 'recordVideo', 'drawSketch', 'getCache', 'getRouting', 'getmemoryDump', and others. Each subclass has order of activities and has individuals. 'EvidenceType' has subclasses

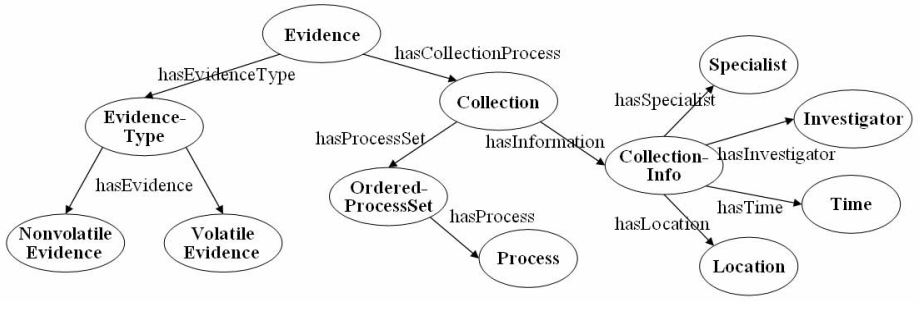

Fig. 2. Subclasses and Relations among Evidence and Collection of process

'VolatileEvidence' and 'NonvolatileEvidence'. We applied to store and retrieve documents and evidences of cyber crime using this ontology for experimental test. In addition, we can apply this ontology to data mining of cyber crime, specifically, to entity extraction, clustering, association, deviation detection, classification, string comparison and social network analysis. Significantly, when we begin a criminal investigation, we can obtain the relational crime type, suspects, similar crime cases and many documents as well as collection evidence and legal-admissibility verification of evidences using the ontology. After completing a crime-case, we can also classify documents, evidence and crime types using this ontology. And also, we can classify (cluster) existing crime documents, and retrieve relational documents of related crime cases and crime types from those existing documents using the data mining techniques.

4 Conclusions

We introduce a cyber forensics ontology for representing concepts of cyber crime and the relations among those concepts in the criminal investigations. Also, we suggest an ontological method for investigation of cyber crimes and a means of applying the ontology to the data mining of cyber crimes. This ontology approach, thereby, can provide comprehensive objective information in the process of conducting forensics. In order to semantically integrate such information, we connected with the concepts and relations of crime case, crime type, criminals, relational law and evidence. This ontology, additionally, can be applied to data mining of cyber crime, extraction of similar crime cases and criminals, clustering (classification, association) of crime types and crime cases, and detection of similar methods of criminal investigation. In the future, we will extend the cyber forensics ontology to make it fully applicable to general purposes. Also we will study automatic concept mapping among ontology and refine the ontology for generalizing.

References

1. Chen, H., Chung, W., Xu, J.J., Qin, G., Chau, M.: Crime Data Mining: A General Framework and Some Examples. Computer 37(4), 50–56 (2004)
2. Brinson, A., Robinson, A., Rogers, M.: A cyber forensics ontology: Creating a new approach to studying cyber forensics. Digital Investigation 3S, S37–S43 (2006)
3. Gruber, T.R.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5(2), 199–220 (1993)
4. Gruber, T. : http://tomgruber.org/writing/ontology-definition-2007.htm
5. Horridge, M., Knublauch, H., Rector, A., Wroe, C.: A Practical Guide To Building OWL Ontologies Using The Prot´eg´e-OWL Plugin and CO-ODE Tools. Univ. Manchester (2007)
6. The Cyber Terror Response Center (CTRC) of the Korean National Police Agency (KNPA), http://www.netan.go.kr/eng/index.jsp
7. Donalds, C.M., Osei-Bryson, K.: Criminal Investigation Knowledge System: CRIKS. In: The 39th Annual Hawaii International Conference on System Sciences, vol. 7, pp. 152–160 (2006)
8. Dzemydiene, D., Kazemikaitiene, E.: Ontology-Based Decision Support System for Crime Investigation Processes. In: Information Systems Development, pp. 427–438. Springer, Heidelberg (2005)
9. Dzemydiene, D.: Knowledge Representation in Advisory Information System of Crime Investigation Domain. In: Databases and Information Systems II, pp. 135–146. Springer, Heidelberg (2002)