International Journal of Cyber-Security and Digital Forensics (IJCSDF) 5(2): 75-82
The Society of Digital Information and Wireless Communications, 2016 (ISSN: 2305-0012)

Building Ontologies for Digital Forensic Terminologies

[1]Nickson M. Karie* and [2]Victor R. Kebande †

[1]Department of Computer Science, Kabarak University,
Private Bag - 20157, Kabarak, Kenya
[2]Department of Computer Science, University of Pretoria,
Private Bag X20, Hatfield 0028, Pretoria, South Africa
menza06@hotmail.com*, vickkebande@gmail.com†

ABSTRACT

Digital forensics (DF) is a relatively new discipline with a lot of technical and non-technical terminologies that can be hard to comprehend. During a time-intensive digital forensic investigation process, for example, investigators may at times encounter several new terminologies. In such a scenario, the time required to unearth and analyse the root cause of a potential security incident might be influenced by the complexity involved in resolving the meaning of new terminologies encountered. The difficulty lies in the lack of an approach in DF that can help investigators in resolving the meaning of terminologies or even how these terminologies are perceived by individuals especially when used in their domain of expertise. If existing digital forensic tools, for example, were to be designed in such a way as to allow investigators to automatically resolve or incorporate the meaning of new terminologies used or encountered during investigations, then the time required to unearth and analyse the root cause of a security incident might be reduced extensively. The main problem addressed in this paper therefore, is that, there exists no approaches in DF that have the ability to help investigators in reasoning with regard to the perceived meaning of different digital forensic terminologies encountered during a digital forensics investigation process. Existing tools thus needs to incorporate new approaches that can help in resolving or clarifying the meaning of new terminologies used during investigation processes. For this reason, this paper examines the concept of building ontologies for digital forensic terminologies and proposes an ontological approach to resolve the meaning of different digital forensic terminologies. Besides, ontologies are known to provide a form of knowledge in a given discipline of interest. In the authors' opinion, thus, building ontologies for digital forensic terminologies can support the development of future investigative tools as well as new techniques to a degree of certainty.

KEYWORDS

Ontology; digital forensic; terminologies; digital investigation; tools and techniques

1 INTRODUCTION

The development and evolution of Information and Communication Technologies (ICT) infrastructure in the current society has led to increased usage of electronic devices, which has further transformed the way individuals communicate. More so, this has been realized in the way through which data is transferred among devices. For instance, there have been large data repositories that are able to manage heterogeneous movement of data across diverse locations. Furthermore, this massive explosion of ICT infrastructure has led to the usage of technologies such as: Automation, teleworking, video conferencing, Voice over Internet Protocol (VoIP) and increased online transactions. In fact, the International Data Corporation (IDC) has projected that in the year 2016, ICT spending in the Middle East, Turkey, and Africa (META) will top $260 billion [1]. This shows that a number of businesses organizations across these regions are increasingly embracing digital transformation initiatives in a bid to streamline their costs and bolster their flexibility [1].

Discounting the aforementioned developments in ICT, there has been explosion of numerous information systems that has attracted a number of adversaries who seem to exploit the benefits of these digital transformations. For example, the continued use of ICT has encouraged the emergence of new forms of crime called

cybercrimes which are perceived to be a conduct prohibited by legislation and or jurisprudence that involves the use of digital technologies while committing an offence. Actually according to Anah et al. [15] as well as Halde & Jaishankar [17] cybercrimes can be defined as: "Offences that are committed against individuals or groups of individuals with a criminal motive to intentionally harm the reputation of the victim or cause physical or mental harm to the victim directly or indirectly, using modern telecommunication networks such as Internet (Chat rooms, emails, notice boards and groups) and mobile phones (SMS/MMS)" [15], [17]. According to the National Crime Prevention Council (NCPC), however, cybercrime is also any crime committed or facilitated via the internet [16]. These types of crimes spearheaded the development of digital forensics.

Digital Forensics (DF) provides a platform that can deal with cybercrimes through the use of scientifically proven methods to excavate potential digital evidence that can be admissible in a court of law. Additionally DF provides an investigation mechanism that may enable forensic investigators to provide the link between the suspect and the crime. However, during a Digital Forensic Investigation (DFI) processes, for example, investigators are at times faced by the menace of new technical and non-technical terminologies. In addition, during the DFI process a number of DF tools are usually used to conduct investigations, however, most of these tools lack the ability to clarify the exact meaning of the different DF terminologies encountered. This makes it hard for the stakeholders involved in reasoning with regard to how the terminologies are used and perceived in the domain.

Based on that premise this paper therefore provides the following contributions:

- Analyses the state of the art in building ontologies for digital forensic terminologies.
- Proposes the inclusion of an ontological approach to resolve the meaning of digital forensic terminologies in existing digital forensic tools.

Such an approach should be able to assist digital forensic analysts and forensic practitioners in comprehending the different digital forensic terminologies used during investigation with ease.

As for the remaining part, the paper is structured in the following format: Section 1 has set the scene of the paper through an introduction; Section 2 will introduce the background of the study while section 3 will provide related work. Thereafter, Section 4 will discuss the concept of building ontologies for digital forensic terminologies while Section 5 will provide a discussion of the proposed ontological approach to resolve the meaning of digital forensic terminologies. The paper concludes with Section 6 and makes mention of the future work.

2 BACKGROUND

This section provides a background study on the following aspects: Digital forensics and ontologies. Digital forensics has been discussed to show the scientific process of digital investigation. Ontologies are discussed as one way to bring out the conceptualization of specific pieces of knowledge as well as how they can be used to simplify the understanding of new terminologies when integrated into digital forensics. This also means that, ontologies can further help forensic investigators in comprehending the meaning of terminologies with ease.

2.1 Digital Forensics

Digital forensics is the science of investigation which deals with extracting digital evidence from computing devices that can be used to reconstruct events for purposes of creating a hypothesis that can be used in a court of law. The concepts of digital forensics were first defined in a technical report produced during the first Digital Forensic Research Workshop (DFRWS) in Utica, New York [2].

According to the DFRWS report digital forensics was defined as “The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis,

interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations" [2].

Based on this definition the authors deduce that digital forensics expands simply from the crime scene through the forensic analysts to the courtroom. This involves proper forensic examination of digital evidence by forensic analysts through the Law Enforcement Agencies (LEAs). Therefore the main role of digital forensics is to unearth digital evidence that will assist the LEAs and prosecutorial offices through presentation of digital evidence in a court of law.

As the technological trends in DF keeps changing, new challenges as well as new terminologies are also constantly introduced into the domain. New meanings sometimes are assigned to existing terminologies [3]. Therefore, approaches need to be developed in digital forensics with the ability to effectively assist investigators in comprehending the meaning of new terminologies that may crop up as a result of technological change or domain evolution. Such approaches will further assist in establishing an effective digital investigation process. Furthermore, the requirement for such approaches and in digital forensics is exceptionally important both for the advancement of the field as well as for the effective use of different domain terminologies and the representation of domain information in any court of law [3]. The next subsection now introduces the concept of ontologies.

2.2 Ontologies

According to Uschold & Gruninger [4] ontologies refers to a shared understanding of some domain of interest which is used as a unifying framework in solving problems. In recent years, the development of ontology has become common in many different domains [5]. Moreover, as noted in a paper by by Karie and Venter [6], ontologies are widely used in different domains as a technique for representing and reasoning about domain knowledge. This also implies that, ontologies have inevitably grown to cover unrelated domains across different levels. Despite the widespread ontology applications in different domains, the development of ontologies and their applications in digital forensics is still in want [6]. On the same note, Noy & Musen [7] highlights that ontologies have become a semantic web back-bone and ubiquitous in information systems that facilitates diverse applications. This involves conceptualizations that are aimed at encoding specific pieces of knowledge together so that simple inferences can be made for problem solving purposes. Nevertheless, according to Chandrasekaran, Josephson, & Benjamins [8] ontologies can be used to provide analysis of the structure of knowledge through the formation of the heart of any system of knowledge representation for that domain. In this paper, however, we present ontologies as an approach that can be used to generate a common definition, knowledge, and understanding of domain terminologies in digital forensics as also explained by VanRees [9]. For this reason, to help create a common definition of terminologies that enhances the sharing and reuse of formal represented knowledge in digital forensics (DF) [10], it is important to develop ontologies that define the common terminologies in which the shared knowledge in this field can be represented. It is on these grounds that the authors propose in this paper an ontological approach to resolve the meaning of digital forensic terminologies. The next section presents related research works in this paper.

3 RELATED WORK

Ontologies can be integrated into any discipline either directly into an ontology development process if no other suitable ontology is available. This process allows identification of mismatches across different disciplines. For example, Leung Lau, & Tsang [11] has proposed the following activities that can be used for ontology integration that uses key terms: Develop a set of inspiration scenarios; Conceptualization; Evaluate validity and sufficiency of key terms; Categorize key terms; Identify candidate ontology; Evaluate

concepts of each candidate ontology; Identify source ontology and its knowledge module; Evaluate the quality of knowledge modules and integrate knowledge modules into one ontology. Based on these activities, the authors proposed an approach for ontology integration that allowed provision of a detailed description on how to perform ontology integration through the use of elicitation of key terms, identification of source ontologies and their knowledge modules.

Research by Karie and Venter [3] also proposed a mechanism for measuring semantic similarity between digital forensics terminologies using web search engines. In their paper, the authors argues that semantic similarity between different terminologies is becoming a generic problem that extends across numerous domains, touching applications developed for computational linguistics, artificial intelligence, cognitive science and, in the case of this paper, digital forensics. Although the approach was novel, ontologies were hardly a focus.

On a recent approach by Hoss & Carver, [12] the authors address the issue of weaving ontologies to support digital forensic analysis which has shown that ontologies can be used to model and reason about digital forensics knowledge. The authors presented a technique to model, transform and utilize models containing bi-directional typed links with user-defined semantics to define the semantic relationships between models. Their research showed that users are able to input forensic knowledge, query that knowledge, produce reports, and interconnect with existing forensic tools to upload data and/or perform further analysis [12].

To the best of the authors' knowledge, there currently exists no other approach in digital forensics research utilizing ontologies that specifically focuses on resolving the meaning of terminologies as is the case presented in this current paper. However, there exist successful applications of modelled ontologies in other types of research that gave support toward applying these concepts for use in the proposed ontological approach to resolve the meaning of digital forensic terminologies in this paper. In the next section, the concept of building ontologies for digital forensic terminologies is explained.

4 BUILDING ONTOLOGIES FOR DIGITAL FORENSIC TERMINOLOGIES

Much of the research work done in the digital forensic domain has less focus into issues related to building ontologies for digital forensic terminologies. Researchers have managed to develop different types of ontologies; however, very few have their focus on digital forensics. In this section of the paper, the authors present the concept of building ontologies for digital forensic terminologies by proposing an ontological approach to resolving the meaning of digital forensic terminologies shown in Figure 1. The approach is divided into four parts labelled 1 to 4. Each of the identified parts is explained briefly in the sub-sections to follow.

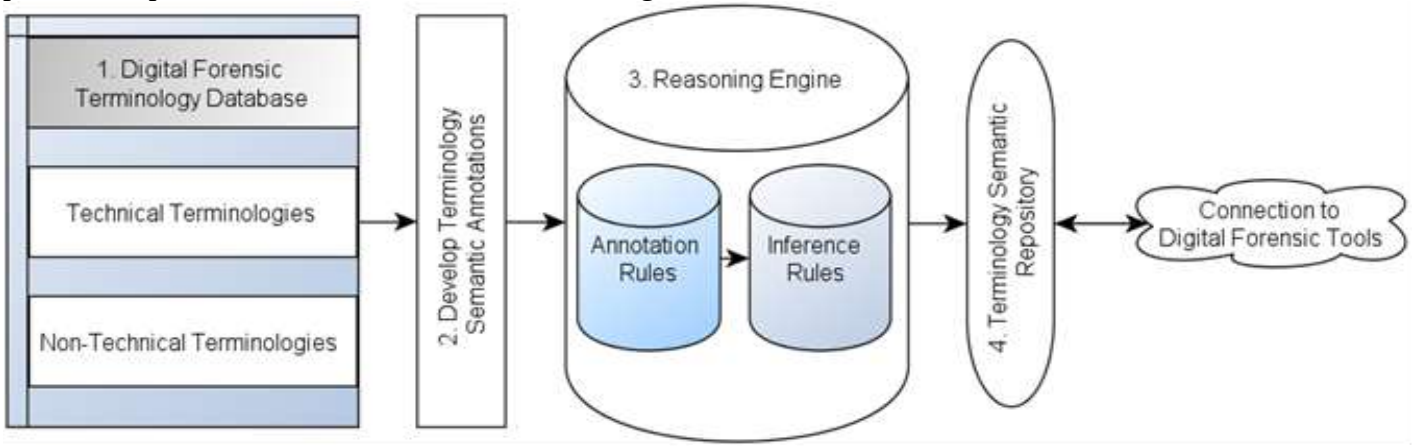

Figure 1: An Ontological Approach to Resolving the Meaning of Digital Forensic Terminologies

4.1 Digital Forensic Terminology Database

According to Wright & Budin [13], a terminology database also known as a term base is a database consisting of concept-oriented terminological entries and related information usually in multilingual format. A term base allows for the systematic management of approved or verified terminologies and is a powerful tool for promoting consistency in terminology use. The terminologies in a term base may include among other information, the definition, the source and context of use as well as the domain area.

In this paper, therefore, the development of an ontological approach to resolve the meaning of digital forensic terminologies starts with a terminology database. As mentioned earlier, with the advancement in technological trends in digital forensics, new terminologies are constantly introduced into the domain and new meanings assigned to existing terms. Developing a terminology database will thus guarantee consistency and accuracy every time a terminology is encountered in a digital forensic investigation process. A customized terminology database specifically designed for digital forensics therefore needs to be developed to support any approach meant to resolve the meaning of digital forensic terminologies. A well designed and clearly defined terminology database can help investigators save time, reduce costs, improve quality and maintain consistency during investigations.

4.2 Develop Terminology Semantic Annotations

The semantic annotation process is an act of expressing knowledge about a particular resource, terminology or phrase. This process involves attaching names, attributes, comments, descriptions, etc., to specific domain terminologies [14]. Semantic annotation is therefore responsible for providing all the information (including additional metadata) about an existing domain terminology or data which in this case will have originated from the terminology database.

A standard semantic annotation exercise has three major building blocks: the ontology, a data instance recognition process and lastly an annotation generation process. Figure 2, below shows the semantic annotation process that makes it possible to assign links to existing semantic descriptions of any domain terminology in question. This makes it possible to relate one domain terminology to another. With this process (see Figure 2) it also becomes possible to annotate different digital forensic terminologies that goes through the reasoning engine and later stored in the semantic repository for use by any digital forensic tool.

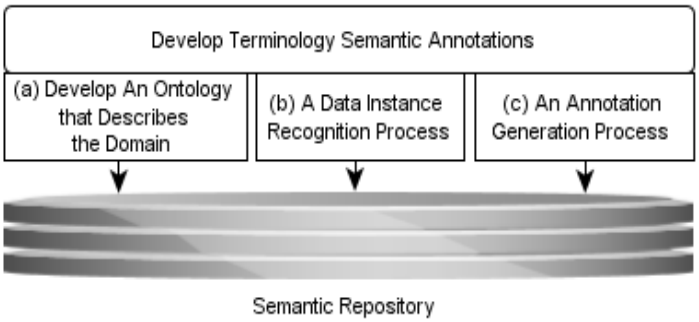

Figure 2: Terminology Semantic Annotation Process

The development of an ontology (labelled (a)) in Figure 2 gives an account of the domain of interest. In this case the domain of interest is digital forensics. The data instance recognition process (labelled (b)) in Figure 2 is meant to discover all the instances of interest in a target document or terminology based on the defined ontology. Finally is an annotation generation process (labelled (c)) in Figure 2. This process is meant to create a semantic meaning disclosure file for each annotated document or terminology [14]. Through the semantic meaning disclosure file, any ontology-aware machine agent can understand the target document or terminology [14].

4.3 Reasoning Engine

A reasoning engine sometimes referred to as rules engine, or a semantic reasoner is a system that is able to infer logical consequences from a set of asserted facts or axioms. The idea behind a reasoning engine is just like that of an inference engine which provides a richer set of mechanisms to work with. The inference rules are commonly

specified by means of an ontology language, and often a description language. In the case of this paper the reasoning engine contains the annotation rules and the inference rules which help in the reasoning process. The result of the reasoning engine is a terminology semantic repository labelled number 4 in Figure 1.

In addition the reasoning engine generates conclusions from the available annotation knowledge using logical techniques such as deduction and induction. The reasoning engine plays an important role in the implementation of artificial intelligence and knowledge-based systems. This also means that any digital forensic tool developed can then be connected to the terminology semantic repository to easy the understanding of terminologies used or encountered during a digital forensic investigation process. The terminology semantic repository is discussed in the sub-section to follow.

4.4 Terminology Semantic Repository

At the heart of the proposed approach lies a terminology semantic repository which is a large and structured set of texts stored in a knowledge base format. The terminology semantic repository is a very useful information source for resolving the meanings of different terminologies. In the case of this paper, such a repository is needed to enable the extraction or retrieval of terminology meanings necessary for a digital forensic investigation process as shown in Figure 1. A standardised digital forensic terminology semantic repository needs to be established for use in implementing newly developed approaches such as the one proposed in this paper, as well as any new forensic tools and techniques. The terminology semantic repository can then be connected to any digital forensic tool for use during digital investigation processes as shown in Figure 1. The next sub-section presents a brief discussion of the proposed approach.

5 DISCUSSION

The proposed ontological approach to resolve the meaning of digital forensic terminologies in this paper is a new contribution in the digital forensics domain. The scope of the approach is defined by the steps shown in Figure 1. The main steps as depicted in Figure 1 include:

- Digital Forensic Terminology Database
- Develop Terminology Semantic Annotations
- Reasoning Engine
- Terminology Semantic Repository

The specific details of the individual steps as identified in the Figure 1 have further been explained in this paper. However, note that the steps as identified in Figure 1 are meant to facilitate this study and primarily focus on the process of resolving the meaning of digital forensic terminologies for use during a digital forensic investigation process. Such proposed steps or guidelines are by no means the final guaranteed steps to successful digital forensic investigation process. Nevertheless, organising the approach into steps was necessary to simplify its understanding.

The Primary reason that led to the development of a new ontological approach in this paper is the lack of approaches in digital forensics that can be used to resolve the meaning of digital forensic terminologies thus simplifying the digital investigation processes to investigators. The proposed approach as demonstrated in this paper can, thus, be used in the digital forensics domain, for example, to help investigators and any other stakeholders including the law enforcement agencies in reasoning with regard to how different terminologies are perceived in the domain.

Academic institutions should also find the approach in this paper constructive, especially when training students on matters related to digital investigation processes. Such an approach can as well be incorporated in curriculums and education materials for different programs of study within the field of digital forensics. Developers of digital forensics tools can also use the proposed approach to develop digital forensic tools with the ability to resolve the meanings of terminologies used during an investigation process before producing the final

report. This also implies that developers might find the approach in this paper useful, especially when considering the development of new digital forensic tools and techniques for addressing different challenges experienced during a digital forensic investigation process. The next section concludes this paper and highlights the possibility of future research work in this study.

6 CONCLUSION

Digital forensics plays a very important role in both incident detection and digital forensic investigations. For this reason, developing ontologies in digital forensics that can help investigators in comprehending the meaning of new terminologies is of utmost importance. Besides, ontologies can for example, be used to build a foundation to solve both present and future digital forensic challenges arising as a result of domain evolution. Such challenges may include those related to the meaning and definition of new terminologies, as well as the reuse and sharing of common domain knowledge.

This paper, thus, presented an analysis of the state of the art on building ontologies for digital forensic terminologies as well as proposed an ontological approach to resolving the meaning of digital forensic terminologies. However, considering the current technological trends, more research needs to be conducted in future in order to expound on the ideas presented in this paper. Further research on how to integrate ontologies into some of the exiting digital forensic tools must also be conducted to help improve on the efficient of the digital investigation process. Finally, the authors recommends the inclusion of ontologies and ontological approaches in existing digital forensic tools as a way to assist digital forensic analysts and forensic practitioners in comprehending the meaning of different digital forensic terminologies used during an investigation.

7 REFERENCES

1. IDC, (2014).”International Data Corporation”, Accessed at http://idc-cema.com/eng/about-idc/press-center/63154-ict-spending-to-top-260-billion-in-2016-as-digital-transformation-initiatives-take-hold-across-the-middle-east-turkey-and-africa.
2. Gary, P., (2001). “A Road Map for Digital Forensic Research”; Technical Report DTR-T001-01, DFRWS, November 2001; Report from the First Digital Forensic Research Workshop (DFRWS). Available online at: http://www.dfrws.org/2001/dfrws-rm-final.pdf [Accessed M
3. Karie, N.M. & Venter, H.S., (2012). Measuring Semantic Similarity Between Digital Forensics Terminologies Using Web Search Engines. In the Proceedings of the 12th Annual Information Security for South Africa Conference. Johannesburg, South Africa. Published online by IEEE Xplore®, (pp. 1-9).
4. Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *Knowledge engineering review*, *11*(2), 93-136.
5. Noy, N.F. & McGuiness, D.L., (2012).Ontology development 101: a guide to creating your first ontology; http://protege.stanford.edu/publications/ontology_development/ontology101.pdf (accessed August 3, 2012).
6. Karie, N.M. & Venter, H.S. (2014). “Towards a General Ontology for Digital Forensic Disciplines”. Journal of Forensic Sciences. Vol. 59, No. 5. Online ISSN: 1556-4029
7. Noy, N.F. & Musen, M. A. (2004). "Ontology versioning in an ontology management framework," in *IEEE Intelligent Systems*, vol. 19, no. 4, pp. 6-13.
8. Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R., (1999). What are ontologies, and why do we need them?. *IEEE Intelligent systems*, (1), 20-26.
9. VanRees, R., (2012). Clarity in the usage of the terms ontology, taxonomy and classification. Digital Library of Construction Informatics and information technology in civil engineering and construction; http://itc.scix.net/cgi-bin/works/Show?_id=w78-2003-432 (accessed June 26, 2012).
10. Gruber, T.R., (1993). A translation approach to portable ontology specification. Knowl Acquis 1993;5 (2):199–220.
11. Leung, N. K., Lau, S. K., & Tsang, N. (2014, May). A new methodology to streamline ontology integration processes. In *Digital Information and Communication Technology and it's Applications (DICTAP), 2014 Fourth International Conference on* (pp. 174-179). IEEE.
12. Hoss, A. M., & Carver, D. L. (2009, June). Weaving ontologies to support digital forensic analysis. In *Proceedings of the 2009 IEEE international conference on Intelligence and security informatics* (pp. 203-205). IEEE Press.
13. Wright, S.E. & Budin, G.(2001) Handbook of terminology management (Vol. 2): Application-oriented terminology management. John Benjamins Publishing Company.
14. Ding, (2006). Semantic Annotation for the Semantic Web. Available at:

http://www.deg.byu.edu/ding/research/SemanticAnnotation.html [Accessed February 19, 2016].

15. Anah, B.H., Funmi D.L., & Julius, M., (2011) Cybercrime in Nigeria: Causes, Effects and the Way Out. ARPN Journal of Science and Technology. VOL. 2, NO. 7,pp. 626-631
16. Unknown (2012). Cybercrimes. Available Online at: http://www.ncpc.org/resources/files/pdf/internet-safety/13020-Cybercrimes-revSPR.pdf [Accessed April 03, 2016].
17. Halder, D., & Jaishankar, K., (2011), Cybercrime and the Victimization of Women: Laws, Rights, and Regulations. Hershey, PA, USA: IGI Global. ISBN 978-1-60960-830-9.