An Approach for Managing Knowledge in Digital Forensic Examinations

April L. Tanner alb117@msstate.edu
Department of Computer Science and Engineering
Mississippi State University
Mississippi State, 39762, USA

David A. Dampier dad6@msstate.edu
Associate Professor of Computer Science and Engineering
Mississippi State University
Mississippi State, 39762, USA

Abstract

Computers and digital devices are continuing to evolve in the areas of storage, processing power, memory, and features. Resultantly, digital forensic investigations are becoming more complex due to the increasing size of digital storage reaching gigabytes and terabytes. Due to this growth in disk storage, new approaches for managing the case details of a digital forensics investigation must be developed. In this paper, the importance of managing and reusing knowledge in digital forensic examinations is discussed, a modeling approach for managing knowledge is presented, and experimental results are presented that show how this modeling approach was used by law enforcement to manage the case details of a digital forensic examination.

Keywords: Digital Forensics, Concept Mapping, Case Domain Modeling, Digital Forensic Examinations

1. INTRODUCTION

Of the many issues associated with computer forensics, knowledge management strategies are also important to the future of not only computer forensics, but digital forensics as well. Several models have been developed [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17]. These models are extensions of the DFRWS model which served as the basis for digital forensic modeling approaches. These models focused on the investigative process and the different phases, they addressed the complexity of an investigation and the features and functionality of devices, and the concrete principles of an investigation. Of the models listed, one focused on a specific phase and produced empirical results. Empirical results of actual application and usage of modeling approaches by digital forensic investigators are lacking significantly. Research involving investigators is extremely limited in digital forensic research, especially when focusing on the examination phase of a digital forensic investigation. Reasons for this may be that investigators can not understand the modeling approach, investigators may be hesitant to learn a new method or model and may rely on their own departmental or organizational procedures, and/or investigators may be unaware of the different modeling approaches. In either of the cases, research is lacking to determine if, in fact, modeling approaches are being used at all in digital investigations. Furthermore, research is also needed to address knowledge management strategies in computer forensics. According to [18], "Effective knowledge management maintains the knowledge assests of an organization by identifying and capturing useful information in a usable form, and by supporting refinement and reuse of that information in service of the organization's goals. A particularly important asset is the internal knowledge embodied in the experience of task experts that may be lost with shifts in projects and personnel." There is a need for knowledge management in digital forensics due to the increased usage of the Internet, the increase in digital crimes using different types of digital media, and the constant advances in technology. A simplified method for capturing and reusing digital crime knowledge could prove to be invaluable to the law enforcement community.

Tacit knowledge or expert knowledge is basically an internal knowing of what needs to be done and how it should be done [18]. Computer crimes are increasing, and there is a great need for knowledge sharing amongst the local, state, and federal authorities to further combat these crimes. When computer forensic examiners perform examinations, their specialized skills may not be recorded. These specialized skills could be very useful for external reviews and training. Skilled and experienced personnel know what to look for, where to look, and how to look without compromising the evidence. Externalizing this knowledge could assist novice examiners in investigations and could potentially lead to the creation of a knowledge repository. In most cases, digital forensic examiners must search through large amounts of data to find evidence. With digital storage capacities becoming increasingly larger, this task is becoming even more complex and time consuming. Knowledge management methodologies in the computer forensics domain have been addressed in [19] [20]. Bruschi, Monga, and Martignoni [19] proposed a model that organizes forensic knowledge in a reusable way. This model uses past experiences to train new personnel, to enable knowledge sharing among detective communities, and to allow third parties to assess the quality of collected information. They also suggested that disciplined methodologies should be created that provide the possibility of archiving digital forensic knowledge that would aid in training and best practice guidelines.

A method for effectively reusing and managing knowledge could greatly improve the digital forensic process. According to [20], the practice of digital forensics could be enhanced by developing "knowledge management strategies specific to law enforcement that will operate within the specific context of criminal investigations". In [19], their approach aims to provide a "methodology for archiving, retrieving, and reasoning about forensic knowledge, in order to incrementally improve the skills and the work of a team of detectives." Their proposed software tool and approach will produce reusable forensic knowledge as support during investigations, will organize past experience to encourage knowledge sharing among forensic experts, and will record collected information in a way that eases quality assessment. In order to demonstrate the importance of capturing and reusing knowledge, Kramer utilized concept maps to provide a method for capturing the tacit knowledge of design process experts.

Kramer's [21] research project attempted to collect, understand, and reuse the knowledge of multiple domain experts on design processes that drive initial design decisions associated with translating "Requirements on Orbit" to "Design Requirements." Concept maps were utilized as a knowledge acquisition and representation tool among multiple domain experts in the translation from a statement of requirements to design requirement specifications. Three specific goals for this research were as follows: demonstrating how concept maps can be used for knowledge acquisition among multiple domain experts; developing a prototype knowledge representation model from the concept maps for guiding the development of design requirements from "Statements of Requirements on Orbit"; and assessing the utility of that prototype knowledge acquisition and representation model by examination of a limited problem set. Kramer was able to effectively show the usefulness of concept maps in eliciting and representing expert knowledge; consequently, this paper explores the possibility of utilizing concept maps in the digital forensics domain. A possibility exists for incorporating concept maps into every phase of a digital investigation; however, in this research, concept mapping will be applied only to the examination phase of an investigation.

2. THE CONCEPT MAPPING CASE DOMAIN MODELING APPROACH

Conceptual models are suitable for representing the information domain of a computer forensics examination. Concept maps are a type of conceptual model that organizes and represents knowledge hierarchically by showing the relationships between concepts. Concept maps were first used in 1972 to track and better understand children's knowledge of science [22]. Since then, researchers and practitioners from various fields have used them as evaluation tools, to plan curriculums, to capture and archive expert knowledge, and to map domain information [21] [22][23]. Novak and Cañas stated that "concept mapping has been shown to help learners learn, researchers create new knowledge, administrators to better manage organizations, writers to write, and evaluators assess learning." Furthermore, a concept map can be viewed as a "simple tool [that] facilitates meaningful learning and the creation of powerful knowledge frameworks that not only permit utilization of the knowledge in new contexts, but also the retention of knowledge for long periods of time" [22]. In other words, information that is learned through the use of concept maps allows one to relate this information to previous and potentially new information and retain this information longer. Concept mapping is suitable for modeling the case domain because concept maps are easy to understand, can be used to organize information, has a semi-automated tool available, can be shared, has the ability to create new knowledge and uncover gaps in a person's knowledge.

The concept mapping case domain modeling approach (CMCDMA) was developed from Bogen's [24] case domain model and the concept mapping model used by Novak and Canãs [22]. Bogen's [24] case domain model provided a framework for analyzing case details by filtering important forensic-relevant case information; in addition, it provided a foundation for organizing knowledge and focusing a forensics examination plan, and it utilized established ontology and domain modeling methods to develop the framework of the model, and artificial intelligence and software engineering concepts, such as Unified Modeling Language (UML) conceptual diagrams, were used to represent the model. The concept mapping model provides a way to organize the case details of an examination, which could be useful later for analyzing the evidential findings. Elements of both models were used to create a five phase, non-linear process for modeling the information domain consisting of the following steps: identifying a focus question, identifying the case concepts, identifying the attributes, identifying the relationships, and instantiating the model.

First, the focus question is created. The focus question helps provide the context for the map to aid in searching for evidence and searching for additional evidence. Second, the case concepts or keywords are identified. Nouns and noun phrases or objects or events are generally chosen to represent the case information. General and specific concepts can be created and used in future investigations. Concepts can be reused from previous cases/models; reusing the concepts can save time when developing future cases/models. Figure 1 provides a representation of the concept mapping case domain model for a murder-gambling case. The case scenario for murder-gambling is as follows:

> *May Doe was involved in a fatal car accident at 12:25 pm, Wednesday, February 11, 2009. She was driving a 2001 Honda Accord. Her death was initially labeled an accident. However, May's parents strongly feel that she was murdered by her husband, Jim Doe. According to John, May and Jim's twenty-five year old son, his father proposed to a woman named Pam Dean one week after his mother's death. John also stated that his father received $500,000 from his mother's life insurance policy with AcciLife Insurance Company.*
>
> *Upon further review of the May's totaled vehicle, it was found that the car did not contain any brake fluid (or oil) and several holes were found in the brake line. Six days prior, May had her engine serviced as a result of the appearance of the engine service light coming on in her vehicle. A receipt taken from her purse showed that her brakes were checked, her brake fluid was refilled, and the oil was changed. In addition, a thumb drive was also found in the arm rest of May's car. Family members, friends of the family, and*

> *neighbors were interviewed by the police; however, no one noticed anything out the ordinary between them. Everything seemed fine according to the son, but John told police that his parents had been arguing a lot lately about his father's gambling.*

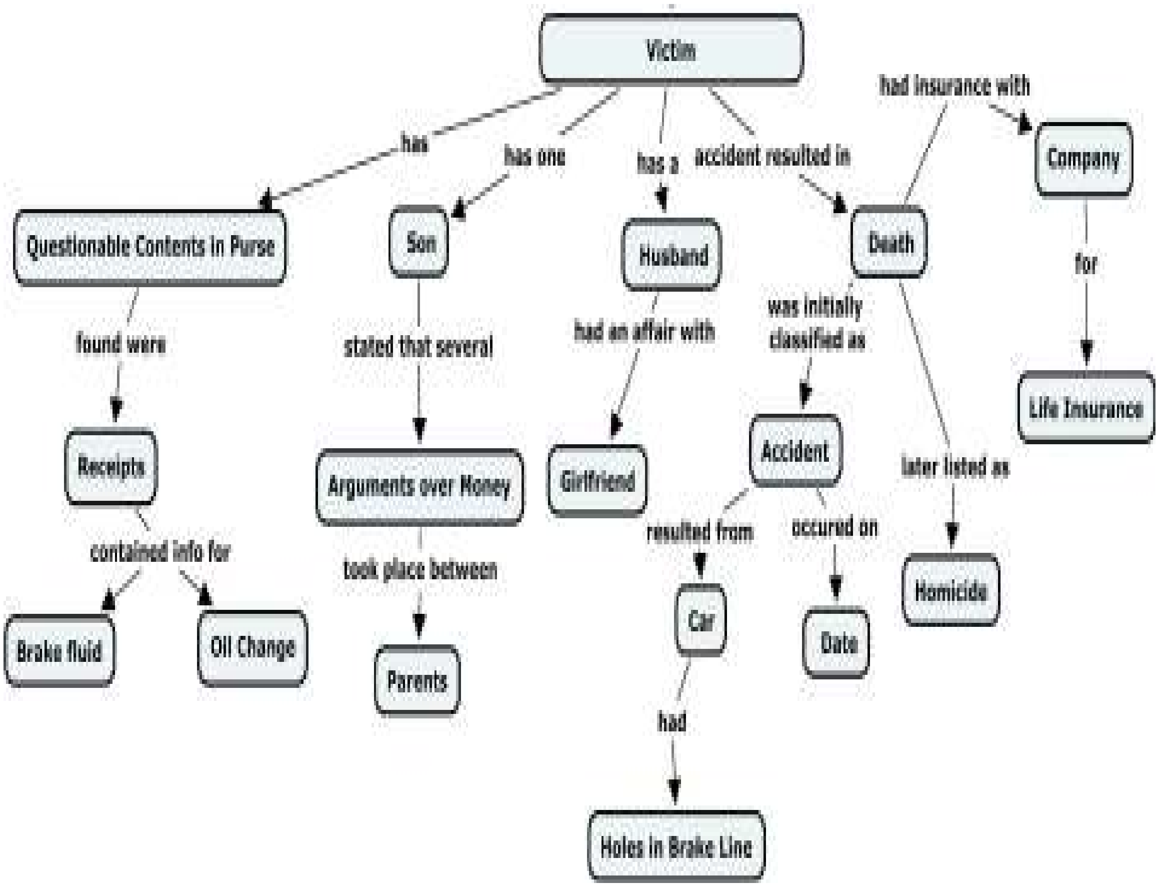

FIGURE 1: Keyword Concept Map for Murder-Gambling Case Scenario

In Figure 1, the general concepts and their relationships are shown. From this concept map, a general, quick overview of the case is shown. After the preliminary map has been created, the attributes from the case scenario should be established. Attributes help clarify the concepts' meanings, represent specific events or objects, and can be used for constructing keyword searches, examining documents, examining network logs, and linking other concepts [24]. Next, the relationships are identified. They show how the concepts are related to one another and consist of verb, verb phrases, numbers, and symbols. In the last phase of the CMCDMA, the model is instantiated by adding the attributes, or the specific information, to the map such as the name of the victim, the type of car driven by the victim, and the date the last oil change was performed as shown in Figure 2. Attributes can also include icons such as photos, documents, video, audio clips, and other digital media. Figure 3 represents an instantiated keyword concept map containing the attributes of the murder-gambling case scenario with icons displayed for the May Doe and Honda Accord Concepts. Each of the figures was created using concept mapping software, CmapTools. The concept mapping case domain model is not reliant on the CmapTools software. This model can be constructed without the use of CmapTools. However, it would be very beneficial in the law enforcement community for including additional resources such as photos, subpoenas, search warrants, and examination search procedures used. Keyword concept maps can provide an examiner with a quick way to view the evidence that was collected based on specific keywords or can be used to store documents associated with the case within the case concept map as well. Additional concept maps can be created to guide an examiner during an examination.

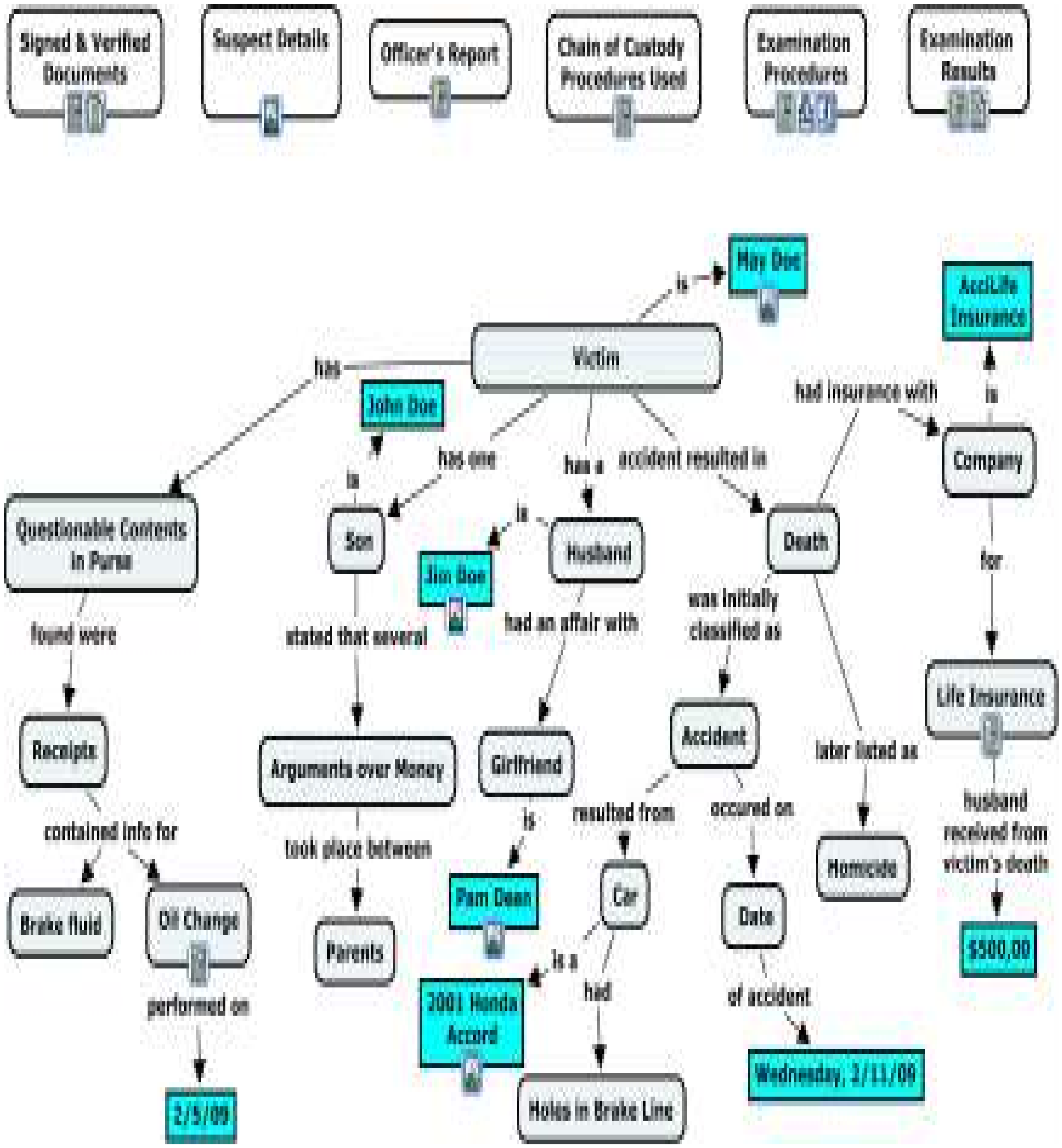

FIGURE 2: Keyword Concept Map for Murder-Gambling Case Scenario with Case Specific Details

Not only can the CMCDMA be used to organize the case details or manage the knowledge of an investigator's report, the approach can be used to structure the examination process also. For instance, Figure 4 provides a general examination concept map that can be used to guide the examiner during an examination. Special techniques suggested by the examiner could easily be added to the map and used in future examinations as well. Given that each case is different, a different set of tasks may be required to search for and identify evidence in an investigation. This map could easily be altered to include additional tasks as needed by following the steps of the CMCDMA. To make the map less cluttered and more readable, it could be broken into two or more concept maps; for instance, one map could include tasks 1-5, and the other map could contain tasks 6-10.

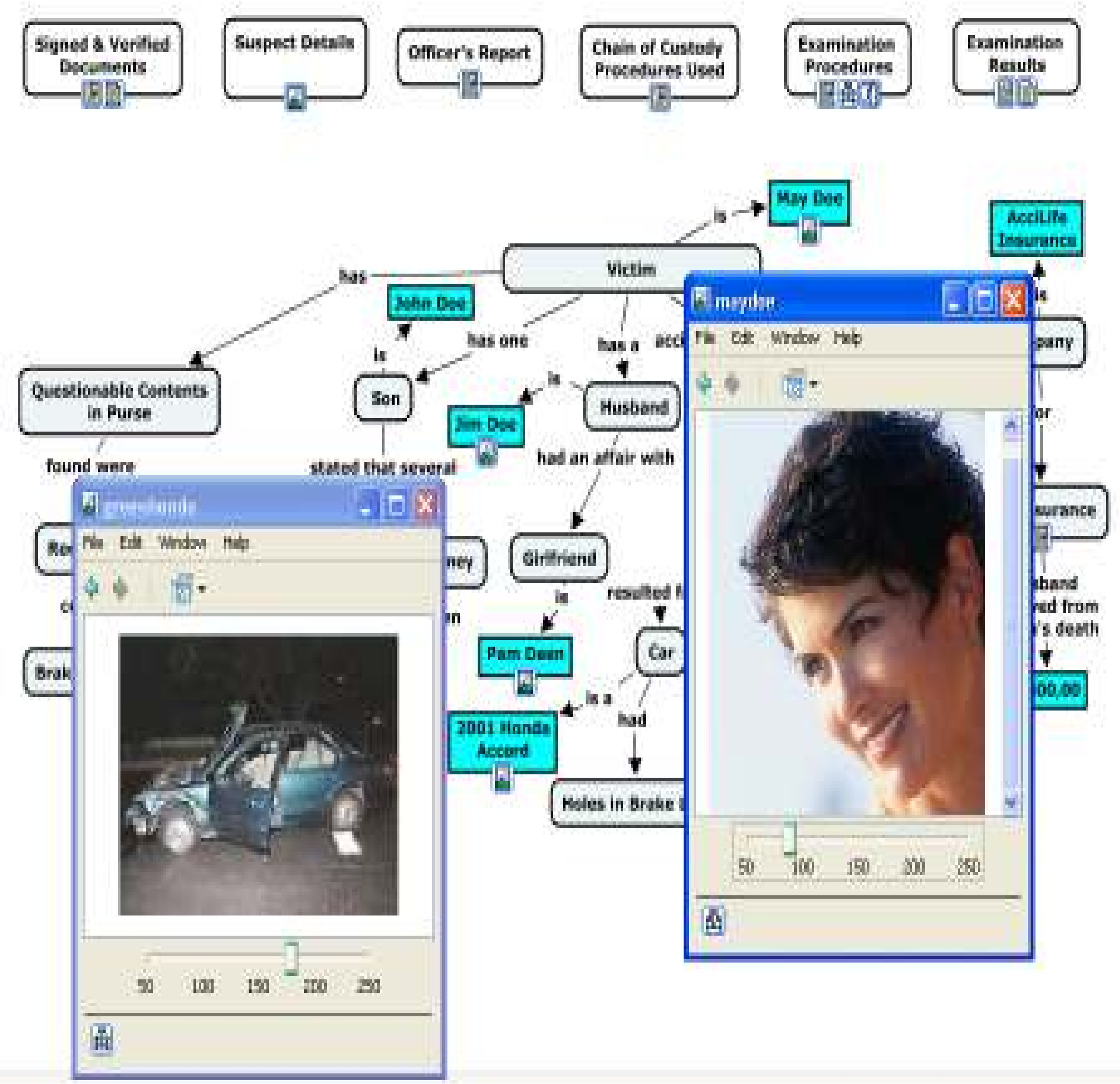

FIGURE 3: Keyword Concept Map for Murder-Gambling Case with Icons Displayed

3. EXPERIMENTAL DESIGN

The subject population consisted of law enforcement officers taking an investigation planning class offered through the National Forensics Training Center. Four experiments were performed. They were divided into a control group and experimental group. The experimental group used the concept mapping case domain modeling approach. The control group did not use the concept mapping case domain modeling approach but used the generally used, ad hoc method. Each group used their respective methods to develop keywords, plan and execute the examination, and record the results. The data in the following tables was collected from the experimental data of the control and experimental groups. The data in the following tables only presents the data provided by the experimental group. This data was categorized based on the experience levels of the subjects. From this data, we were able to determine what affect the subjects' experience with computer forensic examinations had on their abilities to use the concept mapping case domain modeling approach to plan, search, and identify evidence in the digital forensic examination. The overall amount of evidence found and time spent in the phases was compared between those with little or no experience and those with experience.

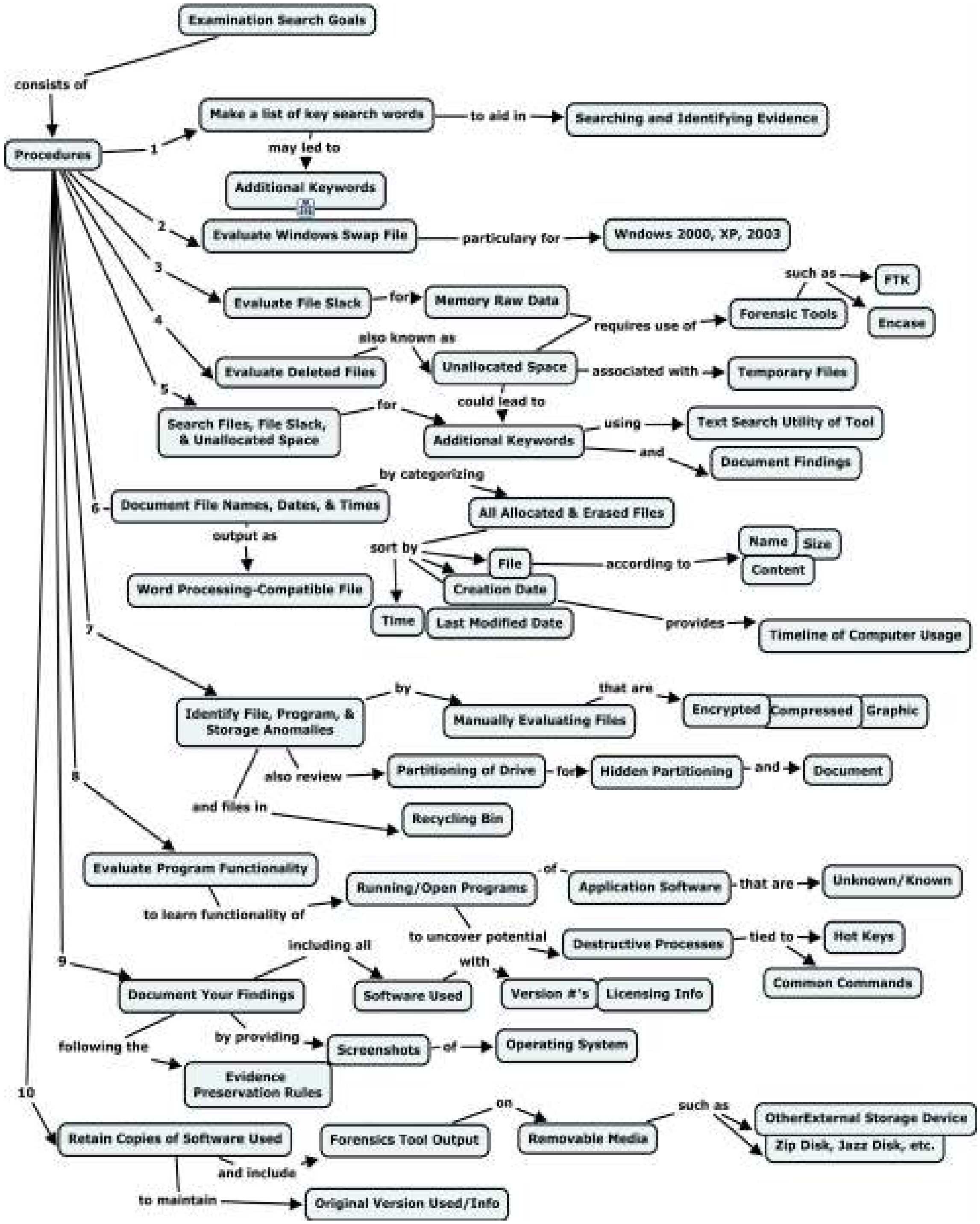

FIGURE 4: A General Examination Concept Map

Survey questions were given in an effort to obtain both qualitative and quantitative data about the concept mapping case domain modeling approach. The responses for the discussion questions are not included; however, the analysis section includes insightful discussion responses given by the groups.

Table 1 provides the level of experience for subjects in the experimental groups for each of the four experiments based on the answers that the subjects provided voluntarily. At the beginning of the seminar course, the subjects were asked to rate their level of expertise with respect to computer forensic examinations. The experience levels were as follows:

- [A] No Experience (0-1 years) consists of knowledge of the computer forensic investigation process.
- [B] Little Experience (1-2 years) which consists of the previous experience level and attended seminars/courses/workshops in computer forensics.
- [C] Some Experience (2-3 years) consists of the previous experience levels, securing the computer/digital evidence, and notifying forensics lab, knowledge of computer forensic software and hardware.
- [D] More experience (3-4 years) consists of the previous experience levels, used digital forensic software and hardware tools to authenticate or copy evidence in an actual digital forensic investigation.
- [E] Expert/Experienced (4-5+ years) consists of the previous experience levels, performed digital forensic examinations, created reports using digital forensics software.

Experiment	[A]	[B]	[C]	[D]	[E]
Experiment 1					
E1E-1			**X**		
E1E-2					**X**
E1E-3			**X**		
Experiment 2					
E2E-1			**X**		
E2E-2				**X**	
E2E-3					**X**
Experiment 3					
E3E-1			**X**		
E3E-2		**X**			
E3E-3		**X**			
Experiment 4					
E4E-1		**X**			
E4E-2		**X**			

TABLE 1: Experience Level of Subjects in Experimental Groups for Experiments 1-4

In order to conduct the examination, forensics software was used to search and identify evidence utilizing the keywords and concept maps created from the concept mapping case domain modeling approach and the examination concept map. This evidence was bookmarked and included in the final report. Computer forensic software, such as FTK, allowed the case examiner to provide additional/important notes about the bookmarked evidence in addition to time and date information and the location of the evidence. For this approach, the bookmarked information was used to indicate what evidence was found and where the evidence was found. Once all the keywords had been searched and the examiner had completed his/her examination of the evidence drive, a report was generated including all of the bookmarked items created by the examiner. After the report had been created, a summary report was filled out. The summary report aided in analyzing the evidence findings and was useful in presenting new information about the case that was unknown by the subject before the examination.

The murder-gambling case scenario discussed previously was used by the subjects in the experiment during the examination. The evidence drive consisted of a 2 gigabyte (GB) thumb drive that contained a total of 2572 files (counts were determined by Forensic Toolkit's count of file items), including 59 evidence files.

4. EXPERIMENTAL ANALYSIS

The data for these statistical analysis tests were taken from the experimental groups of the four experiments. The experimental group data was grouped into two categories: Little or No Experience (LNE) and Experienced (E). The LNE group consisted of four subjects and the E group consisted of seven subjects. The data for experiments 1-4 was combined and analyzed according to the groups. For instance, in Table 2, E3E-2 represents experiment 3 and experiment group subject 2. Table 2 represents the data collected during the planning and examination efforts in Experiments 1-4, where time is expressed in minutes. Time data information was provided for each subject in the experimental groups (concept mapping case domain modeling approach) for each experiment. In this research, subjects with little or no experience had 0-2 years experience in computer forensic examinations; in addition, those subjects with more than 2 years experience in computer forensics examinations were considered experienced.

Little or No Experience	**Planning Time (minutes)**	**Examination Time (minutes)**	**Total Time (minutes)**
E3E-2	38	33	71
E3E-3	5	87	92
E4E-1	7	104	111
E4E-2	55	72	127
AVERAGE	26.25	74.00	100.25
Experience			
E1E-1	10	140	150
E1E-2	44	131	175
E1E-3	30	123	153
E2E-1	13	114	127
E2E-2	40	80	120
E2E-3	40	87	127
E3E-1	27	65	92
AVERAGE	29.14	105.71	134.86

TABLE 2: Planning and Examination Effort for Experimental Groups in Experiments 1-4

Table 3 represents the amount of evidence, which is expressed as percentages, found by each subject in the experimental groups in each experiment. The evidence was classified into seven groups: Emails, May, Jim, Life Insurance, Gambling, Vehicle, and Other. The group names of the evidence represented the types of evidence and the names of the victim and suspect who had files on the evidence drive. In addition, the overall or total percentage of the evidence found by each subject and each group are provided in the last column.

Little or No Experience	% of Emails	% of May	% of Jim	% of Life Insurance	% of Gambling	% of Vehicle	% of Other	Overall %
E3E-2	33.33	100.00	14.29	100.00	91.67	50.00	45.45	54.24
E3E-3	58.33	100.00	14.29	80.00	16.67	70.00	63.64	47.46
E4E-1	58.33	50.00	14.29	60.00	58.33	50.00	81.82	55.93
E4E-2	75.00	100.00	28.57	100.00	83.33	70.00	90.91	76.27
AVERAGE	56.25	87.50	17.86	85.00	62.50	60.00	70.46	58.48
Experience								
E1E-1	91.67	100.00	28.57	80.00	83.33	70.00	63.64	79.66
E1E-2	50.00	100.00	42.86	80.00	33.33	40.00	54.55	52.54
E1E-3	58.33	100.00	28.57	100.00	50.00	40.00	63.64	57.63
E2E-1	58.33	100.00	14.29	100.00	75.00	40.00	45.45	55.93
E2E-2	50.00	100.00	42.86	80.00	83.33	50.00	54.55	62.71
E2E-3	58.33	50.00	14.29	80.00	25.00	50.00	45.45	42.37
E3E-1	58.33	100.00	57.14	80.00	58.33	30.00	36.36	52.54
AVERAGE	60.71	92.86	32.66	85.71	58.33	45.71	51.95	57.63

TABLE 3: Amount of Evidence Found in Experiments 1-4 by Experimental Groups

The chosen method of statistical analysis for testing the hypotheses in the experiment data was the independent, one-sided t-test. The t-test was used to compare the differences or means of the two independent groups. When the t-test's criteria were not met, the non-parametric Kolmogorov-Smirnov (K-S) test was used to evaluate the difference between the means of the two groups. Each of the alternative hypotheses was evaluated based on the 95% confidence interval. The alternative hypotheses were accepted and recognized as having a statistically significant difference when the probability of the null hypothesis was less than or equal to 5% or .05. Otherwise the alternative hypotheses were rejected.

The results of the *t*-tests and K-S tests were appropriately applied to the effort/time data, expressed in minutes, for both the LNE and E groups as shown in Table 4. If t-tests were used to evaluate the data, then the field for t-values contained a value for the test; otherwise, the K-S tests were used and the fields were marked with "- -." Based on the results of the statistical tests, the concept mapping case domain modeling approach resulted in the LNE group spending a significantly less amount of time in the total experimental activity than the E group. Although no significant difference was observed during the planning and examination phases, the LNE group did spend less time in the planning and examination phases than the E group.

Hypothesis	Little or No Experience Mean ($\bar{x}$)	Experienced Mean ($\bar{y}$)	t	p	Result
h_{e1}	$\bar{x}$ = 26.25	$\bar{y}$ = 29.14	-0.258	0.401	Reject h_1
h_{e2}	$\bar{x}$ = 74.00	$\bar{y}$ = 105.71	-1.741	0.058	Reject h_2
h_{e3}	$\bar{x}$ = 100.25	$\bar{y}$ = 134.86	-2.120	0.032	Accept h_3
Hypothesis Legend					
h_{e1} = The group having little or no experience spent a significantly less amount of time in the planning phase/session than the experienced group.					
h_{e2} = The group having little or no experience spent a significantly less amount of time in the examination phase/session than the experienced group.					
h_{e3} = The group having little or no experience spent a significantly less amount of time on the total experimental activity than the experienced group.					

TABLE 4: Statistical Results for Effort Based on Experimental Group Experience Level

Table 5 provides the results of the *t*-tests and K-S tests that evaluated whether the amount of evidence found by the LNE and E groups were statistically significant. The amount of evidence found data is expressed in percentages. Based on the statistical tests, the LNE group found a significantly greater amount of evidence containing Other files than the E group. Although no other significant differences were found between the groups, the LNE group's mean amount of evidence found was slightly higher for Gambling files, Vehicle files, and total overall evidence.

All the subjects from both groups indicated that the model was helpful in understanding the case concepts and examination tasks. The investigators all indicated that they were confident or extremely confident in their abilities to apply the modeling approach during an investigation/examination. The results of the experiment indicated that the concept mapping case domain modeling approach was useful for typical law enforcement involved in computer forensic cases. Furthermore, this experiment showed that subjects with experience or little or no experience in computer forensic examinations were able to properly use the concept mapping case domain modeling approach to plan, search for, and identify evidence. According to the post-experiment discussion survey responses, a majority of the subjects felt that the concept mapping case domain modeling approach and graphical representation would be beneficial to law enforcement during examinations, for training, and for presenting information to jurors. The subjects also stated that the CMCDMA made it easier to organize the details of the case, it offered a graphical representation of what occurred and what was discovered, and it helped them to focus and limited the amount of data to search/analyze/review. On the other hand, the subjects also felt that the CMCDMA was time consuming, the examination map was cluttered and hard to follow, and the concept map duplicated the investigator's notes.

<table>
<tr><th>Hypothesis</th><th>Little or No Experience Mean ($\bar{x}$)</th><th>Experienced Mean ($\bar{y}$)</th><th>t</th><th>p</th><th>Result</th></tr>
<tr><td>h_{e4}</td><td>$\bar{x}$ = 56.25</td><td>$\bar{y}$ = 60.73</td><td>- -</td><td>0.997</td><td>Reject h_4</td></tr>
<tr><td>h_{e5}</td><td>$\bar{x}$ = 17.86</td><td>$\bar{y}$ = 32.66</td><td>- -</td><td>1.000</td><td>Reject h_5</td></tr>
<tr><td>h_{e6}</td><td>$\bar{x}$ = 87.50</td><td>$\bar{y}$ = 92.86</td><td>- -</td><td>0.643</td><td>Reject h_6</td></tr>
<tr><td>h_{e7}</td><td>$\bar{x}$ = 85.00</td><td>$\bar{y}$ = 85.71</td><td>- -</td><td>0.997</td><td>Reject h_7</td></tr>
<tr><td>h_{e8}</td><td>$\bar{x}$ = 62.50</td><td>$\bar{y}$ = 58.33</td><td>0.243</td><td>0.407</td><td>Reject h_8</td></tr>
<tr><td>h_{e9}</td><td>$\bar{x}$ = 60.00</td><td>$\bar{y}$ = 45.71</td><td>- -</td><td>0.377</td><td>Reject h_9</td></tr>
<tr><td>h_{e10}</td><td>$\bar{x}$ = 70.46</td><td>$\bar{y}$ = 51.95</td><td>2.069</td><td>0.035</td><td>Accept h_{10}</td></tr>
<tr><td>h_{e11}</td><td>$\bar{x}$ = 58.48</td><td>$\bar{y}$ = 57.62</td><td>0.946</td><td>0.179</td><td>Reject h_{11}</td></tr>
<tr><td colspan="6">Hypothesis Legend</td></tr>
<tr><td colspan="6">h_{e4} = The group with little or no experience found a significantly different amount of evidence files containing Emails than the experienced group.</td></tr>
<tr><td colspan="6">h_{e5} = The group with little or no experience found a significantly different amount of evidence containing May files than the experienced group.</td></tr>
<tr><td colspan="6">h_{e6} = The group with little or no experience found a significantly different amount of evidence containing Jim files than the experienced group.</td></tr>
<tr><td colspan="6">h_{e7} = The group with little or no experience found a significantly different amount of evidence containing Life Insurance files than the experienced group.</td></tr>
<tr><td colspan="6">h_{e8} = The group with little or no experience found a significantly greater amount of evidence containing Gambling files than the experienced group.</td></tr>
<tr><td colspan="6">h_{e9}= The group with little or no experience found a significantly different amount of evidence containing Vehicle files than the experienced group.</td></tr>
<tr><td colspan="6">h_{e10} = The group with little or no experience found a significantly greater amount of evidence containing Other files than the experienced group.</td></tr>
<tr><td colspan="6">h_{e11} = The group with little or no experience found a significantly greater amount of overall evidence than the experienced group.</td></tr>
</table>

TABLE 5: Statistical Results for Amount of Data Found Based on Experience Level

The concept mapping case domain modeling approach (CMCDMA) was created to improve upon the weaknesses of Bogen's case domain model (CDM). The goals of both models were to create a model that could be used to share and capture knowledge, to create an approach that was domain specific and could be used during the examination phase of a digital forensic investigation, to create an approach that could reduce the time spent planning and examining evidence, and to create a modeling approach that could be used to recover more evidence than when using an ad hoc approach. Section 3 discussed the experimental design and implementation of the CMCDMA; the experimental designs of both models are very similar, and Table 6 provides a brief overview of the evidence disk characteristics and maximum time allotted for the experiment. Although the size of the evidence drive in the CMCDMA was smaller than the sizes of the drives in Bogen's CDM experiments, both modeling approaches utilized similar techniques, such as keyword searches, that saved time and eliminated the needed to search through every file on the evidence drive. This technique allowed the subjects to utilize forensic software that would find specific key terms on the entire evidence drive very quickly. In addition, the subjects in the CDM experiments were given four hours for planning and a hour and a half more time to search for evidence, while subjects in the CMDCMA experiment were given a maximum of two hours for both planning and examination.

Experiment Approach Used	# of Evidence Files	Total # of Files on Evidence Disk	Size of Evidence Disk (GB)	Maximum Time Allowed in Experiment
CMCDMA	59	2572	1	2.5 hours
Bogen (CDM)				
Experiment 1	99	2981	40	4 hours
Experiment 2	29	58,459	40	4 hours
Experiment 3	33	58,894	10	4 hours

TABLE 6: Comparison of CMCDMA and CDM Experimental Design Data

Experiment Approach Used	Mean Planning Time (min)	Mean Examination Time (min)	Mean Total Time (min)	Overall % of Evidence Found
CMCDMA	28.09	94.18	122.28	57.94
Bogen (CDM)				
Experiment 1	162.83	167.00	329.83	49.33
Experiment 2	134.14	137.71	271.86	35.47
Experiment 3	78.67	89.17	167.83	25.50

TABLE 7: Comparison of CMCDMA and CDM Experimental Group Data

In Table 7, the data for the CMCDMA experiment was combined from Table 2 and Table 3 in order to determine the mean planning time, mean examination time, the mean total time spent in the experiment, and the overall percentage of evidence found by the experimental group, which consists of an aggregation of the LNE and E groups data. Table 7 also provides the results of Bogen's CDM experiments as well. The data shows that subjects using the CMCDMA spent less time planning, less total time in the experiment, and found at least 7% more evidence than those subjects using Bogen's CDM method. In all but one of the experiments, the subjects in the CDM experiment spent more time in the examination phase of the experiment than those subjects in the CMCDMA. Reasons for the large amount of time differences in the planning times of the CMCDMA and the CDM method are that the CDM method was more paper intensive and required the subjects to fill out forms and transfer the information to other forms; also, the subjects were required to complete four activities, which consisted of modeling the information domain of the case utilizing UML conceptual diagrams, developing search goals, specifying search methods for each search goal, and finally conducting the examination. Furthermore, each of these activities required additional tasks to be performed. In relation to the CDM method, the CMCDMA consisted of only modeling the information domain of the case utilizing concept maps, which was one process composed of five tasks.

Although the CDM approach was successful in allowing the subjects to recover more evidence than when using an ad hoc approach, several of the CDM subjects indicated that Bogen's method felt more like paperwork; in addition, they indicated that the availability of semi-automated software would have allowed them to model the details of the case and document their findings. The CMCDMA was developed to provide a simpler way represent the case details of the investigation using concept maps that could contain evidence items specific to the case such as photos, documents, video, and other files. All the information related to the case could be accessed in one location, including other important documents such as subpoenas, search warrants, and other critical documents. The concept maps also provided a quick way to review the case details, to locate keywords that could be used to search the evidence drive, and to manage knowledge gained for a particular type of case. General concepts and concept maps created for that particular case could be used in future cases and altered to include specific attributes for each of the different cases. Managing knowledge in this way could greatly reduce the time needed to investigate and examine digital forensic cases in the future.

5. CONCLUSION AND FUTURE WORK

This paper described the need for managing knowledge in a digital forensic investigation. The concept mapping case domain modeling approach was presented as a method for managing knowledge acquired during a digital forensic examination. The approach provided a way to visually represent the knowledge gained during an investigation and also discussed how the approach can be applied in real digital forensic cases, for training and during an examination. Empirical evidence was provided that showed how novice and experienced law enforcement officers used the approach to plan, search, and identify digital evidence in a digital forensic examination. More research is needed to address ways to model the knowledge obtained during a digital forensics investigation due to the ever increasing sizes of digital storage. Domain modeling in digital forensics is still an emerging area. Several modeling approaches have been proposed, however, little or no empirical data is available for comparing the applicability and usability of these approaches by law enforcement and forensic practitioners. From researching several methods, no other experimental data is available in domain modeling other than Bogen's experimental results. Resultantly, there is a substantial need for experimental data produced by modeling approaches in order to determine if these modeling approaches can be applied by law enforcement in real world cases, if they are useful for managing knowledge, and if these approaches can improve digital forensics investigations by reducing the amount of time needed to examine digital forensic evidence. Future research endeavors include automating the concept map creation process after the examination of digital forensic evidence has occurred. Most digital forensic examiners use computer forensic tools to examine digital evidence, such as Encase and AccessData's Forensic Toolkit (FTK). An automated process could be developed that creates and positions the concepts based on the data in the digital forensic examiner's Encase or FTK report and categorize the evidential findings based on their file extensions, date and time, and etc.

6. REFERENCES

1. V. Baryamureeba, F. Tushabe. *"The Enhanced Digital Investigation Process Model".* In Proceedings of the 4th Annual Digital Forensic Research Workshop, Baltimore, MD, 2004

2. N. Beebe and J. Clark. *"A Hierarchical, Objectives-Based Framework for the Digital Investigations Process".* In Proceedings of the 4th Annual Digital Forensic Research Workshop, Baltimore, MD, 2004

3. B. Carrier and E. Spafford. *"An Event-Based Digital Forensic Investigation Framework".* In Proceedings of the Fourth Annual Digital Forensic Research Workshop, Baltimore, MD, 2004

4. S. Ciardhuáin. *"An Extended Model of Cybercrime Investigations".* International Journal of Digital Evidence, 3(1):1-22, 2004

5. M. Reith, C. Carr, G. Gunsch. "*An Examination of Digital Forensic Models".* International Journal of Digital Evidence, 1(3):1-20, 2002

6. G. Ruibin, T. Yun, M. Gaertner. *"Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework".* International Journal of Digital Evidence, 4(1):1-13, 2005

7. J. Venter. *"Process Flow Diagrams for Training and Operations".* Advances in Digital Forensics II, Springer, pp. 331-342 (2006)

8. Tanner and D. Dampier. *"Concept Mapping for Digital Forensics Investigations".* Advances in Digital Forensics V, Springer, pp. 201-300 (2009)

9. Tanner and D. Dampier. *"Improving Digital Forensics Investigations with Concept Mapping".* In Proceedings of the Fifth International Conference on Digital Forensics, Orlando, FL, 2009

10. S. Peisert, M.Bishop, S. Karin and K. Marzullo. *"Toward Models for Forensic Analysis".* In Proceedings of the Second International Workshop on Systematic Approaches to Digital Forensic Engineering. Bell Harbor, WA, 2007

11. M. Khatir, S. M. Hejazi and E. Sneiders. *"Two Dimensional Evidence Reliability Amplification Process Model for Digital Forensics".* In Proceedings of the Third International Workshop on Digital Forensics and Incident Analysis*.* Malaga, Spain, 2008

12. Y. Shin. *"New Digital Forensics Investigation Procedure Model".* In Proceedings of the Fourth International Conference on Networked Computing and Advanced Information Management. Gyeongju, Korea, 2008

13. Carrier, E. Spafford. *"Getting Physical with the Digital Investigation Process".* International Journal of Digital Evidence, 2(2):1-20, 2003

14. National Institute of Justice. Electronic Crime Scene Investigation: A Guide for First Responders 2001 [Online]. Available at: http://www.ncjrs.gov/pdffiles1/nij/187736.pdf*,* 2001

15. M. Pollitt. *"An Ad Hoc Review of Digital Forensic Models".* In Proceedings of the Second International Workshop on Systematic Approaches to Digital Forensic Engineering. Bell Harbor, WA, 2007

16. R. Rowlingson. *"A Ten Step Process for Forensic Readiness".* International Journal of Digital Evidence, 2(3):1-28, 2004

17. P. Stephenson. *"Modeling of Post-Incident Root Cause Analysis".* International Journal of Digital Evidence, 2(2):1-16, 2003

18. Cañas, D. Leake, and D. Wilson. *"Managing, Mapping, and Manipulating Conceptual Knowledge".* IHMC, 2007

19. Bruschi, M. Monga, and L. Martignoni. "How to Reuse Knowledge about Forensic Investigations". In Proceedings of the 4th Annual Digital Forensic Research Workshop. Baltimore, MD, 2004

20. M. Pollitt and A. Whitledge*. "Exploring Big Haystacks: Data Mining and Knowledge Management".* Advances in Digital Forensics II, Springer, pp. 67-76 (2006)

21. M. Kramer. *Using Concept Maps for Knowledge Acquisition in Satellite Design: Translating "Statement of Requirements on Orbit" to "Design Requirements".* PhD Thesis, Nova Southeastern University, 2005

22. J. D. Novak and A. J. Cañas. "*The Theory Underlying Concept Maps and How to Construct Them".* Technical Report IHMC Cmap Tools 2006-01, Florida Institute for Human and Machine Cognition, 2006

23. S.O. Tergan, *"Digital Concept Maps for Managing Knowledge and Information: Searching for Synergies".* Knowledge and Information Visualization, Springer, pp. 185–204 (2005)

24. C. Bogen. *"Selecting Keyword Search Terms in Computer Forensics Examinations using Domain Analysis and Modeling",* PhD Thesis, Department of Computer Science and Engineering, Mississippi State University, 2006