Design of expert system for tool selection in digital forensics investigation

Erika Ramadhani 1, Elyza G. Wahyuni 2, Hanif R. Pratama 3

1,2,3 Department of Informatics, Universitas Islam Indonesia, Yogyakarta, Indonesia

erika@uii.ac.id

Abstract. Along with the diverse types of electronic goods and digital evidence, techniques have also developed in conducting digital forensic investigations. Another effect is the increase in types and types of tools for conducting digital forensic investigations. The problem is the increasing number of digital forensic tools is not always developed in conjunction with the manual book. So this is what makes the investigator to check the tool to make it suitable for investigations. Choosing the right tool is very important in the investigation process, because each stage of the forensic investigation process has a different treatment for each of its staging. Seeing these conditions, then in this paper we make a system design to facilitate investigators in choosing the right digital forensic tool and in accordance with using an expert system. The design system shows that it’s still needed to confirm to an expert to get a match output if there is a new rule. Later this system will be implemented with website technology.

1. Introduction
Digital forensics is a part of forensic science that aims to find and investigate the data contained in digital devices such as computers, mobile phones, tablets, PDAs, network devices, storage media and the like. During the investigation, the investigator uses an application tool to assist in collecting data and analyzing data, then used to look for evidence of digital data that is commonly used in the realm of courts. The stages carried out in the forensic investigation process include acquisition, identification, evaluation and admission. Each of these stages has certain procedures within handling it. When the investigator or investigator first chooses what tool is suitable for used in the digital forensic investigation process, it requires a fast ability to assess and choose the tool used. Therefore, the selection of tools in the forensic investigation process is very important urgent. The stages carried out in the digital forensic investigation process include acquisition, identification, evaluation, and admission. each stage has a certain procedure in handling it. When the investigator first chooses what tools are suitable for use in the investigation process, it requires a quick ability to judge and choose the tools to be used. Therefore, the selection of tools in the investigation process is very important.

Now more and more tools are available in the field of digital forensics, these tools are not always introduced along with the manual. So an investigator must find out and try independently to get the tools which are deemed suitable for working on a problem. Due to the proliferation of digital storage
and communication devices, it significantly increases the workload of digital forensic testers who need access to appropriate forensic tools. The rapid expansion in the number and variety of digital forensics requires a more efficient method of selecting equipment than is currently available. Another problem is, one of the shortcomings of organizations that develop digital forensic tools can lie with the developers themselves. Many tool developers feel their creation is an answer, but many tools are undefined and well categorized. For the analysis of the flow of values organization is needed to enable practitioners to choose the appropriate tool. Similar problems were observed in the development of digital forensic tools. Some vendors often market their own tools as a complete solution, without considering the broad digital forensic landscape. Forensic examiners are practitioners who need a method to identify and choose the right tool. Indeed, the researchers agreed that computer forensics experts need to make significant decisions regarding the selection of appropriate tools for digital evidence investigation [1], [2].

This research utilizes an expert system that functions as a substitute experts in their fields. Expert systems are systems that seek to adopt knowledge humans to computers, so computers can solve problems like an expert [3]–[7]. Base of the selection of tools in the field of digital forensics is based on a rule-based expert system with use forward chaining whose characteristics refer to research. Forward chaining is one of the methods of the expert system used to attract conclusion based on the facts, the search begins with the facts and then moves through the premise to lead to conclusions. Nowadays more and more tools are available in the fields digital forensics, the tool is not always immediately introduced along with the manual. The result, an investigator must find out and try independently to get which tool deemed suitable for working on a problem. Utilization of expert systems with the forward method chaining in this study to facilitate investigators in choosing the right tool in the process forensic investigation, then an application is made that can facilitate the investigator to retrieve the right decision.

2. Current system and research
The existing system is a system that displays several digital forensics tools by showing their categories and functionality. the form of the system is web based, for example like www.forensikwiki.org. This system includes a short description of the digital forensic tool, but there is no search feature for each description, so it is more useful if the system allows searching and selecting tools for a particular task.

There are also tools in the form of the SENS Investigative Forensic Toolkit (SIFT) Workstation which is a collection of open source forensic tools distributed by the SANS Institute. SANS lists tool characteristics such as file system support, image proof support, and partition table support. When it comes to understanding device functions, this category of features can be very useful. However, this category does not answer questions about the type of output, the skills needed to use the tool, or how well the tool generates reports for court purposes, which is one of the main tasks of digital forensic professionals. Other digital forensic tool selection system can be found in the online Computer Forensic Tool Catalog, managed by the National Institute of Standards and Technology. This online https://toolcatalog.nist.gov/ search tool is supported by several filterable fields. This "technical parameter" for example, the "disk imaging" function has parameters that can be searched from "digest hash algorithms" while the "hash analysis" functionality has the same parameters called "supported hash algorithms," but with different values. The description by NIST calls “Forensic Tool Taxonomy"[8].

According to [2], making an ontology is possible to form general information structures that are useful for reusing knowledge, for taking assumptions in a domain, and for analyzing each piece of knowledge. This is important for the cyber forensics field, because the knowledge gained is about the computer domain of forensic experts and the specifications of the tools that can be used as primary evidence in main investigations that are still being developed.

3. Proposed Model
The model that was proposed to create a digital forensics investigation tool selection system is divided into two parts, namely: a model for expert systems and a model for web based applications. The manufacturing structure consists of three parts: data, services, and reports. The main component in the development of this system is data that contains the data name of the digital forensics tool, characteristic data, and recommended tool results. Then the service in this system is as a medium to assist investigators at the acquisition stage in the digital forensic investigation process. The output form of this system is a recommendation tool that can be used.

3.1. Expert System

3.1.1. Concept of expert system. The most important component in an expert system is the knowledge base and the interference engine. The knowledge base is a place for storing knowledge in computers where the knowledge base uses certain rules. Interference engines are at the core of expert system applications. In this section is a user input in the form of facts, so that a conclusion can be obtained. These two components will be equipped with user interfaces, which make it easier for users to choose tools in the digital forensic investigation process. Figure 1 is a concept of the digital forensic tool selection system using an expert system using the inference forward chaining method.

3.1.2. Forward Chaining. Forward chaining is a method that guides a collection of data or facts that aim to get conclusions. Swara (2018) said that forward chaining is a method to draw conclusion by using the facts [9]. In this method, the facts come from the discussion result with the expert. The technique used in decision making is to use the rule as a representation of knowledge by matching the facts of the IF part of the IF-THEN rule[10], [11]. In this paper, we use a decision tree to get the conclusion according to the fact. Figure 2 concluded the method of forward chaining, as we can see, there are two or more rule that related to the the facts and at the end we can gain the conclusion. The note is, sometime from the forward chaining process, there are a new rule that arise according to the fact. So from that condition, we can discuss again the new rule with an expert for further result. Figure 3 shows the flowchart on how the user use the system. A user/investigator will use a tool in the form of a web interface where the user will answer the question criteria of the tool in the form of "Y" for yes and "N" for no from all of the questions given. The application will provide a response in the form of tool recommendations that can be used based on the fields already filled in by users that match the
tool criteria. Each question is taken based on the literature in accordance with the characteristics of the tool, then the question will be adjusted according to the characteristics of the tool. The answer to this question is based on the number of answers "Y".

![Figure 2. Forward Chaining](image)

3.1.3. Decision table analysis. The representation used is a table representation because the factual statement does not have to be sequential. This table representation is used as a reference to draw conclusions. Table 1 and table 2 are data tables for reference in making decisions. Characteristic data for simulations are taken from research conducted by Goalkeeper, J. Richard Rick [1]. In this paper, the tools used as simulation are the Autopsy, EnCase, Foremost, FTK / LAB, F-Response, FTK

![Figure 3. Flowchart system](image)
Imager, IEF, Redline, RegRipper, SkypeAlyzer, Volatility, and so on tools shown in Table 1. While the characteristics used are Memory requirements, processing speed, output format, required skills, costs, and exam focus.

| Kode Tool | Nama Tool | Kode Tool | Nama Tool |
|-----------|-----------|-----------|-----------|
| T001      | Autopsy   | T013      | Santoku   |
| T002      | Encase    | T014      | DEFT      |
| T003      | Foremost  | T015      | PALADIN   |
| T004      | FTK/LAB   | T016      | SANS      |
| T005      | F-Response| T017      | AFLogical OSE |
| T006      | FTK Imager| T018      | Laron     |
| T007      | IEF       | T019      | Andriler  |
| T008      | Redline   | T020      | UFED      |
| T009      | RegRipper | T021      | XRY       |
| T010      | SkypeAlyzer| T022    | MOBILedit |
| T011      | Volatility|           |           |
| T012      | X-Ways    |           |           |

Table 2. List of characteristic

| No | Karakteristik  | Range | Kode |
|----|----------------|-------|------|
| 1  | Memory Requirement | 1-2 Low K1 |      |
|    |                 | 3 Average K2 |      |
|    |                 | 4-5 High K3  |      |
| 2  | Processing Speed  | 1-2 Low K4  |      |
|    |                 | 3 Average K5 |      |
|    |                 | 4-5 High K6  |      |
| 3  | Output Format    | 1-2 Low K7  |      |
|    |                 | 3 Average K8 |      |
|    |                 | 4-5 High K9  |      |
| 4  | Required Skill   | 1-2 Low K10 |      |
|    |                 | 3 Average K11|     |
|    |                 | 4-5 High K12 |      |
| 5  | Cost             | 1-2 Low K13 |      |
|    |                 | 3 Average K14|     |
|    |                 | 4-5 High K15 |      |
| 6  | Exam Focus       | 1-2 Low K16 |      |
|    |                 | 3 Average K17|     |
|    |                 | 4-5 High K18 |      |

3.1.4. Decision table. At this stage a simulation of the decision making process will be carried out based on table 3. Table 3 is a conclusion from the results of the data collection in the form of a questionnaire filled out by two expert practitioners. In table 3, there are three applications that are not filled by the experts, it is Redline, Regriper, and SkypeAlyzer. So that these three applications are deactivated first in the application that will be developed. Table 4 shows the rule formation table based on table 3. Based on table 3, 36 rules will be formed. The description of the rule is only written down to the third rule which can be seen in table 4.
3.1.5. Rule. In this section, the forward chaining system rules are made. The rules were made based on the results of questionnaire data collection from two expert practitioners. Making rules is based on table 3. So that some rules can be made which are shown in the table 5 below. The output of the rule is a recommendation tool.

### Table 4. Table for rule

| Rule | THEN |
|------|------|
| R1   | T001 |
| R2   | T001 |
| R3   | T002 |

Table 5. IF THEN Rule of selection digital forensics tool

| Rule | IF | THEN |
|------|----|------|
| Rule 1 | IF Memory Requirement $High$ AND Processing Speed $High$ AND Output Format $Low$ AND Required Skill $Average$ AND Cost $Low$ AND Exam Focus $High$ | THEN Autopsy |
### Rule 2

| IF Memory Requirement High AND Processing Speed High AND Output Format High AND Required Skill High AND Cost High AND Exam Focus Average THEN Encase |

### Rule 3

| IF Memory Requirement Average AND Processing Speed Average AND Output Format Low AND Required Skill High AND Cost Low AND Exam Focus Low THEN Foremost |

3.1.6. **Mockup.** Interface design is used to provide how the system will be created later. The interface that is used by user is a web based application. Figure 4. shows the web based display when user using the application. Figure 6. is the use case diagram of system, it shows that the user has three processes, i.e. functionality selection, characteristics selection, and user can get the recommendation of tool.

![Mockup web based expert system](image1)

**Figure 4.** Mockup web based expert system

![Expert system for tool selection use case diagram](image2)

**Figure 5.** Expert system for tool selection use case diagram

3.1.7. **Simulation.** At this stage a system simulation will be conducted on the expert system using forward chaining. Simulation at this stage by providing examples of cases that want to be resolved by an expert investigator. Example: An investigator is given a digital evidence that has been neatly wrapped, to explore the contents of data on digital evidence, what tools might be used by an investigator? For the record, investigators want a tool that is easy to use, open source, and provides output that is reader friendly, requires high memory during the investigation process, a process with an average speed. Based on these characteristics, the suitable characteristics are K3, K5, K8, K11, K13, and K16 with file analysis functionality. According to Figure 4., the recommended output tool is the FTK Imager tool.

3.1.8. **Findings.** Another result appear according to another simulation. The results will differ based on the functionality selection made at the beginning of the system. The functionality for Foremost and
Volatility are Document Meta data Extraction and File Analysis, respectively. From table 5, there are some similarities in rules, namely Rule 3 and Rule 23 for the Foremost and Volatility output tools. In addition there are other rules that have the same results:
- Rule 6, 7, and 9
- Rule 10, 11, and 29
- Rule 24, 27, and 30
- Rule 25 and 26

4. Discussion
The discussion section of this system is to explain if there are new facts that do not suitable to the rules that have been made. If it is occur, then the new rule will be evaluated by an expert. The new results of the new rule will be re-entered into the database rule.

For example, in the example of simulation in the section 3.1.7 we get suitable characteristics i.e. K3, K5, K8, K11, K13, and K16. The question is, what is the result if the characteristics entered by an investigator not suit to the system such as he input K1, K5, K8, K11, K13, and K16. If we check with table 3 and table 4 then the rule is not available in the rule database. This condition is said to be a condition of the existence of new facts in the system. Furthermore, this new fact will be entered into the database rule that has been created.

5. Summary and Suggestion
According to the explanation of the problems that have been discussed in the introduction, then to answer these problems is to create a system based on experts to facilitate investigators in making decisions in the selection of tools. Knowledge gathering is taken from two experts who act as practitioners. The knowledge collected is in the form of digital forensic tools that have three functionality and six characteristics which are categorized into three low-moderate-high choice parts. The selection of functionality is taken based on a questionnaire to practitioners and concluded based on tools that have uniform functionality. The results of this study are design systems which are focused on the rule based forward chaining method.

From these method we find that there are several similarity result in the recomendation of tool. To overcome this similarity we suggest to use a priority question of each characteristic. Another suggestion is to utilizes the tree decision method than table decision method. Due to the lack of experts filling out the questionnaire, the results of the rule are not as complex if there are conditions there are new facts. To anticipate this, new facts are re-entered into the knowledge database to enrich existing rules.

References

[1] J. R. Kiper, “InfoSec Reading Room Pick a Tool , the Right Tool: Developing a Practical Topology for Selecting Digital Forensics Tools,” 2018.
[2] S. Grigalunas, J. Toldinas, and A. Venckauskas, “An Ontology-Based Transformation Model for the Digital Forensics Domain,” no. June, pp. 1–6, 2017.
[3] H. Hibashi, T. Vidas, and L. Cranor, “Usability of Forensics Tools: A User Study,” 2003.
[4] L. S. Riza, M. Pradini, E. F. Rahman, and Rasim, “An Expert System for Diagnosis of Sleep Disorder Using Fuzzy Rule-Based Classification Systems,” in IOP Conference Series: Materials Science and Engineering, 2017, vol. 185, no. 1.
[5] J. Gudu, D. Gichoya, P. Nyongesa, and A. Muumbo, “Development of a Medical Expert System as an ExpertKnowledge Sharing Tool on Diagnosis and Treatment of Hypertension in Pregnancy,” Int. J. Biosci. Biochem. Bioinforma., pp. 297–300, 2012.
[6] L. R. Kashapova, D. L. Pankratov, and P. P. Utyaganov, “An expert system for ensuring the reliability of the technological process of cold sheet metal forming,” in IOP Conference Series: Materials Science and Engineering, 2016, vol. 134, no. 1.
[7] C. P. C. Munaiseche, D. R. Kaparang, and P. T. D. Rompas, “An Expert System for Diagnosing Eye Diseases using Forward Chaining Method,” in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 306, no. 1.

[8] “Computer Forensics Tools & Techniques Catalog - Home.” [Online]. Available: https://toolcatalog.nist.gov/. [Accessed: 12-Oct-2019].

[9] G. Y. Swara and Anisya, “Forward chaining method on selecting low cost green car with rule based,” in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 420, no. 1.

[10] W. Supartini and H. Hindarto, “Sistem Pakar Berbasis Web Dengan Metode Forward Chaining Dalam Mendiagnosa Dini Penyakit Tuberkulosis Di Jawa Timur,” *Kinetik*, vol. 1, no. 3, p. 147, 2016.