Jurnal Mantik, 6 (1) (2022) 202-209

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Simulation Signature-Based Carving Raster Image Using Finite State Automata

Hendra Budi Kusnawan[1], Windu Gata[2], Laela Kurniawati[3]

Nusa Mandiri University, Jakarta, Indonesia
[12]Faculty of Information Technology, Computer Science Study Program, Indonesia
[3]Faculty of Information Technology, Information System Study Program, Indonesia
Jl. Kramat Raya No.18, Jakarta, Indonesia

Email: 14210206@nusamandiri.ac.id, windu@nusamandiri.ac.id, laela@nusamandiri.ac.id

ARTICLEINFO

Article history:

Received: Mart 28, 2022
Revised: Apr 05 , 2022
Accepted: Apr 25, 2022

Keywords:

Automata,
FSA,
Raster Image Carving

ABSTRACT

Abstract – Automata is used as the first stage of describing an idea or example to develop a model, be it software or hardware. Theoretically the concept of automata can be used to simulate the signature-based carving process for identification/recovery of raster type image file PNG and GIF. The research and data collection method used is based on a literature study by explaining the Finite State Automata (FSA) concept in simulating the carving process on an image file. This simulation is a simple example of implementing the automata concept. Even though it is only an abstract machine, automata can be used to simulate various other things that will be solved by computation.

1. Introduction

Automata comes from the Greek word which means to self-acting, self-willed, self-moving. Automata theory is the study of mathematical objects called abstract machines and various computational problems it can solve [1]. So it it clear here that automata is an abstract machine and not a real machine. Automata is used as the first stage to describes an idea or frame work to develop a model, software or hardware. Automata is an abstract machine that can recognize, accept, or generate a sentence in a particular language. In the process it accepts onput and produces output, input string is accepted when it reaches the final state and vice versa [2].

Carving is a general term for extracting files from raw data, based on the specific characteristics of the file format present in that data. In addition, carving only uses information from raw data, not from information from the file system [3]. Carving is a technique that is widely used for file recovery. File recovery is the process of recovering deleted or corrupted files from digital storage while their file system metadata is still available. In carving technique, there are three (3) categories, namely Signature-based, Structure-based and Content-based [4]. Nowadays, with the widespread use of various digital devices, sometimes evidence of a crime is found in the digital media of these devices. Therefore, the Carving technique for file recovery has become a method that is widely used in digital forensic investigations.

Image file is a standard specification for encoding information about images into data bits for storage media. In short, an image is stored and encoded into an image format that is known to identify itself as an image and provides useful information such as matrix size, bits and depth for easy interaction with the file. Any program that meets the standard format can open the file and display the image [5]. Currently, there are various types of image files, in this paper we will only discuss as an example of raster image files, PNG and GIF.

Research method and data collection used based on literature study, by explaining the concept of Finite State Automata (FSA) to simulates Carving process in an image file. Currently, modern automata theory focuses on the reproduction of human thought patterns and problem solving abilities using artificial intelligence and other more sophisticated computer science techniques [6].

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Research on image identification and recovery has been done and researched before. The following is a previous study that discusses FSA and identification of Image Recovery: First, the research conducted by Widyasari (2011) in the journal Sisfotenika, vol. 1, no. 1, pp. 59–67. Widyasari researched about Finite State Automata (FSA) theoretically and tested on automatic machines. The similarity between previous studies and this research is that they both describe Finite State Automata (FSA). Meanwhile, the difference between the research carried out lies in the object used.

Second, research conducted by Ardiansyah, Nila Hardi and Windu Gata, (2020) in the journal Komputika J. Sist. Comput., vol. 9, no. 1, pp. 75–83. They carried out identification research on JPEG File Recovery with the Signature-based Carving method and described in automata. The similarities between previous studies and this research both describe the process of Carving Image Files in Finite State Automata (FSA). While the difference in the research conducted lies in the object used.

In this paper, we will simulate how the carving process flow for image files is carried out according to FSA concept. As explained earlier that automatic machines are abstract machines, so here we will only explain if a string as an input can be accepted or rejected by the automatic machine which is depicted in a transition diagram and in the transition table. Furthermore, FSA is also tested to prove if an input is whether accepted or rejected by using the JFLAP v.7.1 application [7].

2. Method

Automata comes from the Greek word which means to self-acting, self-willed, self-moving. Automata theory is the study of mathematical objects called abstract machines and various computational problems it can solve [1]. So it it clear here that automata is an abstract machine and not a real machine. Automata is used as the first stage to describes an idea or frame work to develop a model, software or hardware. Automata is an abstract machine that can recognize, accept, or generate a sentence in a particular language. In the process it accepts onput and produces output, input string is accepted when it reaches the final state and vice versa [2].

Carving is a general term for extracting files from raw data, based on the specific characteristics of the file format present in that data. In addition, carving only uses information from raw data, not from information from the file system [3]. Carving is a technique that is widely used for file recovery. File recovery is the process of recovering deleted or corrupted files from digital storage while their file system metadata is still available. In carving technique, there are three (3) categories, namely Signature-based, Structure-based and Content-based [4]. Nowadays, with the widespread use of various digital devices, sometimes evidence of a crime is found in the digital media of these devices. Therefore, the Carving technique for file recovery has become a method that is widely used in digital forensic investigations.

Image file is a standard specification for encoding information about images into data bits for storage media. In short, an image is stored and encoded into an image format that is known to identify itself as an image and provides useful information such as matrix size, bits and depth for easy interaction with the file. Any program that meets the standard format can open the file and display the image [5]. Currently, there are various types of image files, in this paper we will only discuss as an example of raster image files, PNG and GIF.

Research method and data collection used based on literature study, by explaining the concept of Finite State Automata (FSA) to simulates Carving process in an image file. Currently, modern automata theory focuses on the reproduction of human thought patterns and problem solving abilities using artificial intelligence and other more sophisticated computer science techniques [6].

Research on image identification and recovery has been done and researched before. The following is a previous study that discusses FSA and identification of Image Recovery: First, the research conducted by Widyasari (2011) in the journal Sisfotenika, vol. 1, no. 1, pp. 59–67. Widyasari researched about Finite State Automata (FSA) theoretically and tested on automatic machines. The similarity between previous studies and this research is that they both describe Finite State Automata (FSA). Meanwhile, the difference between the research carried out lies in the object used.

Second, research conducted by Ardiansyah, Nila Hardi and Windu Gata, (2020) in the journal Komputika J. Sist. Comput., vol. 9, no. 1, pp. 75–83. They carried out identification research on JPEG File Recovery with the Signature-based Carving method and described in automata. The similarities between

previous studies and this research both describe the process of Carving Image Files in Finite State Automata (FSA). While the difference in the research conducted lies in the object used.

In this paper, we will simulate how the carving process flow for image files is carried out according to FSA concept. As explained earlier that automatic machines are abstract machines, so here we will only explain if a string as an input can be accepted or rejected by the automatic machine which is depicted in a transition diagram and in the transition table. Furthermore, FSA is also tested to prove if an input is whether accepted or rejected by using the JFLAP v.7.1 application [7].

Figure 1. Process flow FSA simulation and testing

Finite State Automata (FSA), also called Finite State Machine (FSM) or Automata is an abstract machine that consists of a set of states namely the initial state and one or more final states, a set of inputs, a set of outputs, and a transition function. The transition function takes the current state and input and directs the pr FSA is represented by 5 tuples or M = (Q, Σ, δ, S, F) where:

Q = set of states
Σ = set of input symbols
δ = transition function
S = initial state, where S Є Q
F = final state, where F ∩ Qocess flow to the next new state [9].

Transition function shows the FSA process each step. Then we can use it to determine the order of the next steps to be processed by the FSA. If M is an FSA, then the relation for each step in FSA M can be defined as follows:

(q1, cw) |-M (q2, w) iff ((q1, c), q2) ∈ δ

Then the relationship for each step in FSA M can occur as much as 0 or 1 step so that:

C1 |-M* C2

So the computation of FSA M is a finite order of C0, C1, C2, …., Cn where n ≥ 0

Carving is the process of recovering deleted or corrupted files from a digital storage device if the file system metadata is no longer available. The carving method does not require any file system metadata, because it will analyze the storage media on a byte per byte basis [10]. Carving method can be divided into 3 (three) categories sorted by level of complexity, namely Signature-Based, Structure-Based, and Content-Based [4].

Signature-based carving is done by looking for patterns in the dataset that mark the beginning of the file (header) and the end of the file (footer). Then the header and footer values will be entered into the output file. After header and footer obtained, all cluster data between header and footer will also be included in the output file. This technique can also be done using only the header file information, namely by searching the dataset for the header value while the footer will be searched based on the file size. This technique assumes that header and footer data of a file is not lost, damaged or deleted and not fragmented [4]

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

3. Result and Discussion

The process of identifying files from digital media by searching for signature files (headers, footers) and extracting the information that lies between is known as carving. Files on digital media can start in a cluster, sector, or any bytes. To optimize the search process to find the signature header, it is enough to look for the first byte of each cluster or sector [11]. Header-footer signature-based carving is done by looking for the presence of the header to identify the beginning of the file while the footer is to identify the end of the file, so that the block between the header and footer is considered the target file [12].

The signature-based algorithm is the easiest technique for identifying file types. This algorithm utilizes the pattern contained in a file. Many file types start with unique byte patterns called magic numbers that can be used as information for reliable identification [13]. In explanation above, the method used in signature-based carving is to look for information on one or both, the header or SOI (Start of File) and footer or EOI (End of File) of an image file to be identified. The signature file is in hexadecimal form according to the binary value which will then be read byte by byte for each file [10]. As shown in Table 1., this paper will use the signature information as input to be simulated through the FSA model.

Table 1
Signature PNG and GIF

Raster Image	*Signature*	
	Header/*Start of File* (SOI)	*Footer*/*End of File* (EOI)
PNG	89 50 4E 47 0D 0A 1A 0A	49 45 4E 44
GIF	47 49 46 38	03 BB

According to Table. 1, it is known that the header and footer of the PNG file will then be used as input to the FSA as follows:

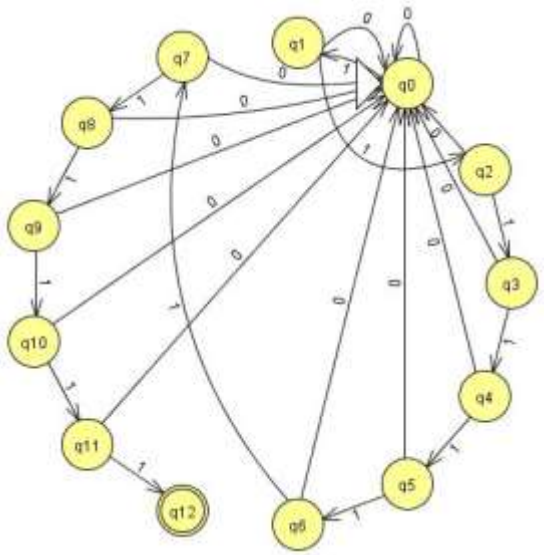


Figure 2. FSA PNG Image

Q = {Q0, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12}
Σ = {0, 1}
S = {I0, I1, I2, I3, I4 … , In}
F = {Q12}
Q = set of states
Σ = set of input symbols, indicated by 0 and 1 where 0 is not appropriate (false) and 1 is appropriate (true)
S = the set of input processes, the start of matching the input whether it matches the header to the footer file
F = final state

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

The FSA above can be explained through the transition function as follows:

S = set of inputs per byte of header and footer PNG image

Q0 = true (1) if input = 89 otherwise it is false (0).

Q0 = 1 → Q1 , Q0 = 0 → Q0

Q1 = true (1) if input = 50 otherwise it is false (0).

Q1 = 1 → Q2 , Q1 = 0 → Q0

Q2 = true (1) if input = 4E otherwise it is false (0).

Q2 = 1 → Q3 , Q2 = 0 → Q0

Q3 = true (1) if input = 47 otherwise it is false (0).

Q3 = 1 → Q4 , Q3 = 0 → Q0

Q4 = true (1) if input = 0D otherwise it is false (0).

Q4 = 1 Q5 , Q4 = 0 Q0 →

Q5 = tru (1) if input = 0A otherwise it is false (0).

Q5 = 1 Q6 , Q5 = 0 Q0

Q6 = true (1) if input = 1A otherwise it is false (0).

Q6 = 1 Q7 , Q6 = 0 Q0

Q7 = true (1) if input = 0A otherwise it is false (0).

Q7 = 1 Q8 , Q7 = 0 Q0

Q8 = true (1) if input = 49 otherwise it is false (0).

Q8 = 1 Q9 , Q8 = 0 Q0

Q9 = true (1) if input = 45 otherwise it is false (0).

Q9 = 1 Q10 , Q9 = 0 Q0

Q10 = true (1) if input = 4E otherwise it is false (0).

Q10 = 1 Q11 , Q10 = 0 Q0

Q11 = true (1) if input = 44 otherwise it is false (0).

Q11 = 1 Q12 , Q11 = 0 Q0

Q12 = Final state.

The following is the transition table from the transition above:

Table 2
PNG Image FSA Transition table

Δ	0	1
Q0	Q0	Q1
Q1	Q0	Q2
Q2	Q0	Q3
Q3	Q0	Q4
Q4	Q0	Q5
Q5	Q0	Q6
Q6	Q0	Q7
Q7	Q0	Q8
Q8	Q0	Q9
Q9	Q0	Q10
Q10	Q0	Q11
Q11	Q0	Q12

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Furthermore, the above FSA test is carried out using the JFLAP application as follows:

Table 3
PNG Image FSA Transition table

Input	Output
111111111111	accepted
101010101010	rejected
111111010101	rejected

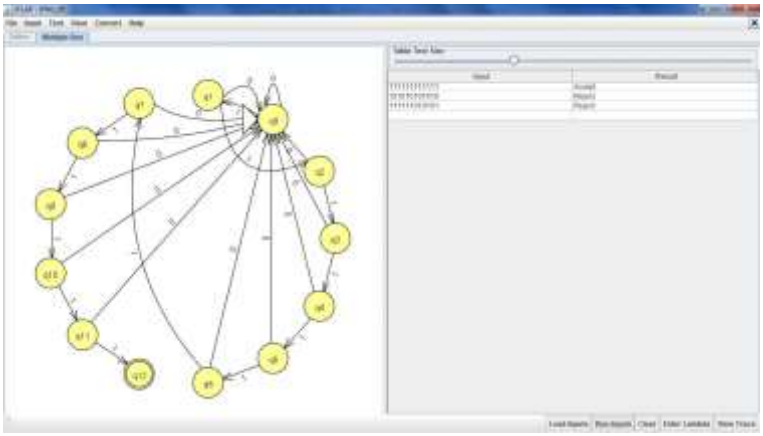
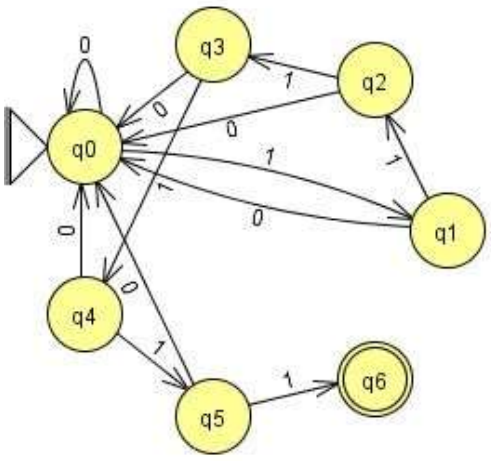


Figure 3. Screenshot of FSA test with JFLAP

From Figure 3 it can be seen that the FSA will accept the input string 111111111111 so that the process ends in the final state. While the input 101010101010 and 111111010101 the process did not succeed in reaching the final state so that the input was rejected by the FSA.

According to Table. 1, it is known that the header and footer of the GIF file will then be used as input to the FSA as follows:



Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Q = {Q0, Q1, Q2, Q3, Q4, Q5, Q6}
Σ = {0, 1}
S = {I0, I1, I2, … , In}
F = {Q6}
Q = set of states
Σ = set of input symbols, indicated by 0 and 1 where 0 is not appropriate (false) and 1 is appropriate (true)
S = the set of input processes, the start of matching the input whether it matches the header to the footer file
F = final state

The FSA above can be explained through the transition function as follows:

S = set of inputs per byte of header and footer PNG image
Q0 = true (1) if input = 47 otherwise it is false (0).
Q0 = 1 $\rightarrow$ Q1 , Q0 = 0 $\rightarrow$ Q0
Q1 = true (1) if input = 49 otherwise it is false (0).
Q1 = 1 $\rightarrow$ Q2 , Q1 = 0 $\rightarrow$ Q0
Q2 = true (1) if input = 46 otherwise it is false (0).
Q2 = 1 $\rightarrow$ Q3 , Q2 = 0 $\rightarrow$ Q0
Q3 = true (1) if input = 38 otherwise it is false (0).
Q3 = 1 $\rightarrow$ Q4 , Q3 = 0 $\rightarrow$ Q0
Q4 = true (1) if input = 03 otherwise it is false (0).
Q4 = 1 $\rightarrow$ Q5 , Q4 = 0 $\rightarrow$ Q0
Q5 = true (1) if input = BB otherwise it is false (0).
Q5 = 1 $\rightarrow$ Q6 , Q5 = 0 $\rightarrow$ Q0
Q6 = Final State.

The following is the transition table from the transition above:

Table 4
GIF Image FSA Transition table

Δ	0	1
Q0	Q0	Q1
Q1	Q0	Q2
Q2	Q0	Q3
Q3	Q0	Q4
Q4	Q0	Q5
Q5	Q0	Q6

Test using the JFLAP application as follows:

Table 5.
GIF Image FSA Transition table

Input	Output
111111	accepted
101010	rejected
101101	rejected

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

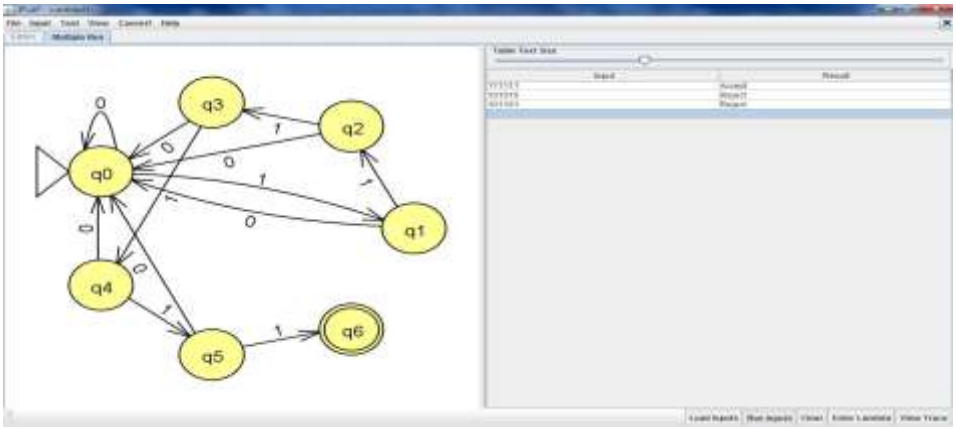


Figure 5. Screenshot of FSA test with JFLAP

From Figure 4 it can be seen that the FSA will accept the input string 111111 so that the process ends in the final state. While the input 101010 and 101101 the process did not succeed in reaching the final state so that the input was rejected by the FSA.

4. Conclusion

As can be seen from the discussion and testing using JFLAP, it can be concluded that automata or FSA can be used for simulations on the signature-based carving method for raster image file types, PNG and GIF. The state transition in the FSA accepts input that matches the header and footer of a raster image file to the final state, while for input that does not match header and footer, the FSA will reject the input so that it does not reach the final state. This simulation is a simple example of implementing the automata concept. Even though it is only an abstract machine, automata can be used to simulate various other things that will be solved by computation.

References

[1] U. Sehgal and S. K. Gill, Discrete Structure and Automata Theory for Learners. BPB Publications, 2020.
[2] Widyasari, "Telaah Teoritis Finite State Automata Dengan Pengujian Hasil Pada Mesin Otomata," Sisfotenika, vol. 1, no. 1, pp. 59–67, 2011, [Online]. Available: https://media.neliti.com/media/publications/.
[3] N. Alherbawi, Z. Shukur, and R. Sulaiman, "Systematic Literature Review on Data Carving in Digital Forensic," Procedia Technol., vol. 11, no. Iceei, pp. 86–92, 2013, doi: 10.1016/j.protcy.2013.12.165.
[4] R. R. Ali, K. M. Mohamad, S. Jamel, and S. K. A. Khalid, "A review of digital forensics methods for JPEG file carving," J. Theor. Appl. Inf. Technol., vol. 96, no. 17, pp. 5841–5856, 2018.
[5] L. K. Tan, "Image file formats," Biomed. Imaging Interv. J., vol. 2, no. 1, pp. 1–7, 2006, doi: 10.2349/biij.2.1.e6.
[6] A. Languages, "Theory of computation, automata and languages," Ife J. Sci., vol. 10, no. 1, pp. 199-206–206, 2008.
[7] "https://www.jflap.org/." .
[8] M. Zed, Metode Penelitian Kepustakaan. Yayasan Pustaka Obor Indonesia, 2014.
[9] A. Adil, Pengantar Teori Bahasa Formal, Otomata dan Komputasi. DEEPUBLISH, 2018.
[10] A. Ardiansyah, N. Hardi, and W. Gata, "Identifikasi dan Recovery File JPEG dengan Metode Signature-Based Carving dalam Model Automata," Komputika J. Sist. Komput., vol. 9, no. 1, pp. 75–83, 2020, doi: 10.34010/komputika.v9i1.2733.
[11] D. Povar and V. K. Bhadran, "Forensic Data Carving," 2011.
[12] E. Alshammary and A. Hadi, "Reviewing and evaluating existing file carving techniques for JPEG files," Proc. - 2016 Cybersecurity Cyberforensics Conf. CCC 2016, no. October 2017, pp. 55–59, 2016, doi: 10.1109/CCC.2016.21.
[13] B. Schildendorfer, "Carving fragmented JPEG images," vol. 2012, no. 2742, 2012.