Nanda Diaz Arizona[1]
Department of Accounting Information Systems
Universitas Bina Sarana Informatika, Indonesia
email: nanda.ndz@bsi.ac.id

Muhammad Agung Nugroho
Informatics
Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia
email: m.agung.n@utdi.ac.id

Ahmad Rois Syujak
Information Technology
Fakultas Dakwah
Universtas Islam Negeri Salatiga
email: Ahmad.rois.Syujak@uinsalatiga.ac.id

Rizqi Kurniawan Saputra
Department of Informatics,
Telkom University,
Purwokerto, Jawa Tengah, Indonesia
email: nenisna@telkomuniversity.ac.id

Moh. Abdul Kholik
Department of Information System
IDN Boarding School Jonggol
Bogor, Indonesia
email: rizqisaputra597@gmail.com

Istri Sulistyowati
Informatics
Universitas Widya Dharma Klaten, Indonesia
email: istri@unwidha.ac.id

Metadata Forensic Analysis as Support for Digital Investigation Process by Utilizing Metadata-Extractor

The rapid development of technology in the current era, in addition to providing positive impacts, certainly also has negative impacts. In Indonesia, based on data from the Cyber Crime Directorate (Dittipidsiber) website, the crime rate related to the ITE Law (Information and Electronic Transactions) is increasing day by day. This encourages digital forensic investigators to be able to develop a concept or method that can be adjusted to digital cases, for example cases of digital data manipulation such as photos or documents. Metadata is an information structure that describes, explains, places in a place or makes it easier to find something, use or manage and sources of information. Metadata can also be interpreted as data about data or information about information. One method or approach that can be done in cases of digital files (photos, videos or documents) can be done using forensic metadata analysis. This is because metadata stores information related to a file. By developing a library from java (metadata-extractor) based on open source and developed in the Netbeans 8.0 application, it will make it easier for an investigator or forensic investigator to conduct a forensic metadata approach, which is expected from the results can be used as valid evidence in the digital forensic investigation process.

KeyWords*: Metadata, Digital Forensics, Cyber Crime, Metadata Analysis, Digital Evidence*

This Article was:
submited: 22-11-24
accepted: 29-12-24
publish on: 10-11-24

How to Cite:
N. D. Arizona, et al, "Metadata Forensic Analysis as Support for Digital Investigation Process by Utilizing Metadata-Extractor ", Journal of Intelligent Software Systems, Vol.3, No.2, 2024, pp.27–31, 10.26798/jiss.v3i2.1503

1 Introduction

In today's digital era, the development and utilization of information technology is growing so rapidly, where it has a very broad impact in all fields, not only has a positive impact but the development of information technology can also cause negative impacts. The unlimited and rapid spread of information results in technology users being unable to distinguish whether information is appropriate or not to be digested, so that sometimes it causes widespread losses [1],[2]. Examples include the spread of hoax news or fake news, changes to data or file manipulation. Based on statistical data obtained from the Cyber Crime Directorate (Dittipidsiber) website which can be accessed at https://patrolisiber.id, the following report data was obtained from January 2020 - May 2023:

(1) Child Porn, Total Cases: 119
(2) Criminal, Total Cases: 212
(3) Hoax / Fake News, Total Cases: 682
(4) Forgery of Letters / Documents, Total Cases: 586
(5) Illegal drug sales on the internet, social media, or other social networks, Total Cases: 39
(6) Insults / Defamation, Total Cases: 6402
(7) Blasphemy, Total Cases: 242
(8) Gambling, Total Cases: 13629

It can be explained that the trend of crime cases is increasing between 2020-2023. Of course, cybercrime also has a very broad impact in the field of investigation / investigation of electronic information and transaction cases regulated in Law of the Republic of Indonesia Number 11 of 2008 concerning Electronic Information and Transactions. In Article 5 paragraph (1) of the ITE Law, it is explained that Electronic Evidence is Electronic Information and / or Electronic Documents and / or their printouts are valid legal evidence, which meets the formal requirements and material requirements regulated in the ITE Law, for that the role of digital forensics as a method of proof in a digital cyber-crime

[1] Corresponding Author.

case is very important. This is also stated in the Explanation of Law of the Republic of Indonesia Number 11 of 2008 concerning Electronic Information and Transactions: "Evidence is a very important factor, considering that electronic information has not only not been accommodated in the Indonesian legal procedure system comprehensively, but has also turned out to be very vulnerable to being changed, tapped, falsified, and sent to various corners of the world in a matter of seconds. Thus, the resulting impact can be so complex and complicated."

Computer forensics is a science that explains the current state of digital artifacts related to legal evidence found on computers and digital storage media. Meanwhile, according to Digital forensics is the process of collecting, identifying, preserving, and examining digital evidence [3],[4].

According to Digital evidence is defined as data stored or carried out using a computer that supports or refutes a theory of how a violation occurred or that discusses important elements of the violation such as intent or alibi. The data in question is a basic combination of numbers that represent various types of information such as text, images, audio, and video. Meanwhile, according to digital evidence can be in the form of document files, history files, or log files that contain related data that can be used as supporting information for decision makers [5],[6],[7],[8],[9].

Research related to forensic metadata analysis has also been carried out previously by using the Java programming language, in his research he built a forensic metadata system to read general metadata characteristics and search for metadata correlation files with one of the parameters, namely file owner, file size, file date and file type. Further research by explains the overall methodology, where in the research two simple open source tools are introduced which were developed to help provide example commands in demonstrating several general metadata analysis requests [7],[8],[9].

2 Method

2.1 Metadata. According to Metadata is information embedded in a file that contains an explanation of the file. This metadata contains information about the contents of data that is used for file management purposes, or the data will later be in a database. In his research explained that metadata Metadata is structured information that describes, explains, finds or at least makes information easy to find again, use, or manage. Metadata is often referred to as data about data or information about information. This metadata contains information about the contents of data that is used for file management purposes/the data will later be in a database. Metadata is very necessary in the digital identification process related to a file; by digging up metadata information on a file, it will certainly make it easier to draw a red thread in a digital forensics case. And can see a series of events, when it was changed, whether it is original or whether the file has undergone a digital change process [6],[9].

2.2 Metadata Types.

2.2.1 Machine readable cataloging (MARC). MARC is the ANSI/NISO Z39.2 Information Interchange Format and ISO 2709 Information and documentation—Format for information exchange metadata standard first developed by the Library of Congress. MARC is one of the results and also a requirement for writing library catalog collections. The LC MARC format is very beneficial for the distribution of library catalog data to various libraries in the United States. MARC is the most widely used metadata language in the library community long before XML and RDF.

2.2.2 Dublin Core. Dublin Core Metadata Element Set (DCMES) developed from a 1995 meeting in Dublin, Ohio, which focused on metadata for networked electronic information. Dublin Core was presented because there were several parties who felt it was not appropriate to use the MARC form so that an agreement was made to compile a new metadata that was easier and more flexible and had the ability to be developed compared to MARC. Dublin Core data consists of 15 basic elements, namely:

a) Title: title of the information source
b) Creator: creator of the information source
c) Subject: subject of the information source, usually stated in the form of keywords or classification numbers
d) Description: description of the contents of the information source, for example in the form of an abstract, table of contents or description
e) Publisher: person or agency that publishes the information source
f) Contributor: person or agency that helps create the information source
g) Date: date of creation of the information source
h) Type: type of information source, report, map and so on
i) Format: physical form of the information source, format, size, duration, information source
j) Identifier: number or series of numbers and letters that identify the information source. Example URL, website address
k) Source: reference to the original source of an information source
l) Language: intellectual language used by the information source
m) Relation: relationship between one information source and another information source
n) Coverage: coverage of content reviewed in terms of geography or time period
o) Rights: copyright owner of the information source.

2.2.3 MODS. MODS stands for Metadata Object Description Schema. is one of the metadata standards developed by the Library of Congress Network Development in collaboration with the MARC standard office, this scheme was developed in response to complaints that the Dublin Core scheme was too simple for the library environment, while the MARC 21 format was too complex and less friendly for users outside the library system.

2.2.4 EXchangeable Image File Format (Exif). Tag structure for embedded metadata in digital image files. The TIFF and JPEG file formats support embedded Exif, but do not include the JPEG2000, PNG, and GIF file formats. The Exif specification includes metadata elements such as pixel dimensions, date and time taken, ISO settings, aperture, white balance, and information about the lens used.

2.3 Confusion Matrix. Confusion matrix is one way to visualize the results of system learning, the visualization displayed contains two or more categories. Confusion matrix is used in this study with the aim of being able to measure the performance of a classification method.

3 Results

This research produces a forensic metadata application that is able to read and analyze digital file metadata using parameters such as author, file type, file size, date, and owner. This application is designed with the Java programming language and developed in the NetBeans IDE 8.2 environment, using the open-source Metadata Extractor library. Testing shows that the application can effectively and efficiently identify files related to evidence, saving investigation time compared to manual methods. The accuracy of the resulting classification model is 55%, which indicates that there is room for improvement, especially in terms of prediction accuracy

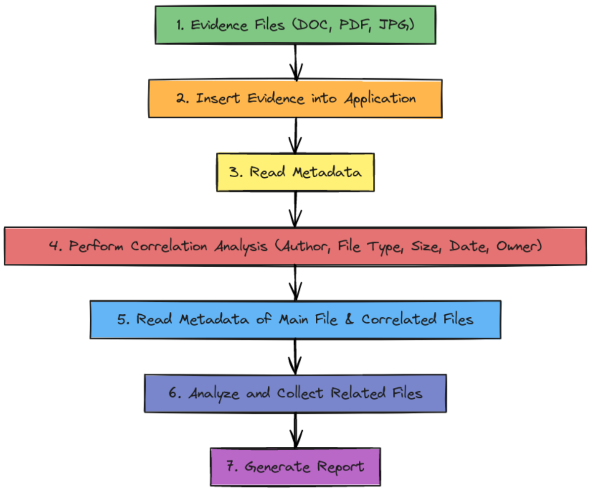

1. Evidence Files (DOC, PDF, JPG)
2. Insert Evidence into Application
3. Read Metadata
4. Perform Correlation Analysis (Author, File Type, Size, Date, Owner)
5. Read Metadata of Main File & Correlated Files
6. Analyze and Collect Related Files
7. Generate Report

Fig. 1 Research Flow

The following is an explanation of the metadata analysis process flow (on Figure 1:

(1) The investigator takes evidence (in the form of a PC, hard disk, SSD or other secondary storage media)
(2) The investigator enters the evidence into the application, the application here is an application that the author designed to analyze forensic metadata.
(3) The evidence file will be read by the application, if read.
(4) The investigator performs a correlation analysis process based on the author, file type, size, date and owner.
(5) Then correlate the evidence by comparing existing parameters.
(6) Next, the investigator collects and groups the evidence from the analysis.
(7) The final process is for the investigator to make a report.

4 Discussions

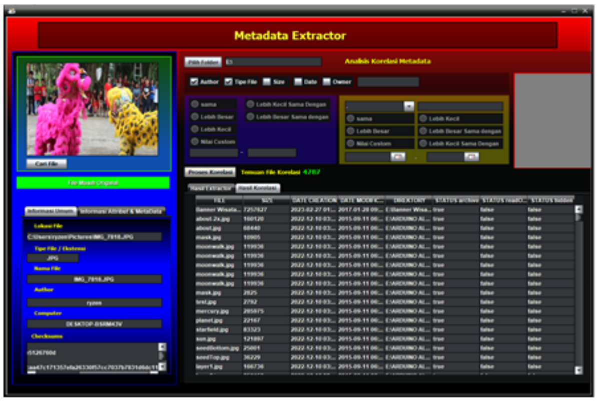
The results of metadata analysis prove that this application facilitates digital forensic investigations by reducing the time required for the process of identifying correlated files. This application utilizes important parameters such as size and modification time to determine correlations between files, which previously required time-consuming manual observation. The use of efficient algorithms in reading metadata allows for rapid detection, which is very useful in time-sensitive investigations. However, the accuracy rate of 55% indicates that this model still needs further development. The cause may be due to the complexity of the various file metadata or the lack of additional more relevant parameters. Therefore, future application development may include testing on different operating systems and adding new parameters that can improve the performance and accuracy of the analysis.

In this process, the evidence obtained by the forensic investigator will undergo a data acquisition stage so that the data is not damaged or lost, this is done so that:

a) Finding evidence can be used for the forensic metadata analysis process.
b) Evidence can be used as evidence for ongoing cases or other cases.
c) Digging up information related to existing evidence.
d) Maintaining evidence so that it is not lost or damaged.

4.1 Forensic Metadata Acquisition and Analysis Stages. The acquisition process is the stage where digital evidence is transferred to other storage, for example a hard disk, or other storage media, this can be done using other supporting applications such as the Encase Forensic application is the most widely known and used forensic tool, which has been produced and launched by Guidance Software Inc. Encase is embedded with various forensic functions that include attributes such as disk imaging and preservation, absolute data recovery in the form of bit streams, in this series of humongous applications, when encase is used to create backups from hard drives, CDs, USB drives, etc.

4.2 Metadata reading stage. In this case it is known that the photo of evidence with the file name "IMG_7818.JPG" was found in the "Pictures" folder, the evidence file will have its metadata read using the Metadata Extractor application by searching for the file that will be processed for reading metadata. Correlation Based on File Type (File Type) analysis results based on file correlation based on File type, to see search results based on file extension correlation, select the file correlation location first, for example on the storage drive location D or E and the results are as shown in the following image on Figure 2:



E:\

Temuan File Korelasi 4787

Lokasi File: C:\Users\ryzen\Pictures\IMG_7818.JPG

Tipe File / Ekstensi: JPG

Nama File: IMG_7818.JPG

Author: ryzen

Computer: DESKTOP-B5RMK3V

Checksums: c5126760d
aa47c171357efa26330f57cc7037b7831d6dc11

FILE	SIZE	DATE CREATION	DATE MODIF...	DIREKTORY	STATUS archive	STATUS readO...	STATUS hidden
Banner Wisata...	7257827	2023-02-27 01:...	2017-01-28 09:...	C:\Banner Wisa...	true	false	false
about-2x.jpg	160120	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
about.jpg	68440	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
mask.jpg	10905	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
moonwalk.jpg	119936	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
moonwalk.jpg	119936	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
moonwalk.jpg	119936	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
moonwalk.jpg	119936	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
moonwalk.jpg	119936	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
mask.jpg	2825	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
text.jpg	2792	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
mercury.jpg	285075	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
planet.jpg	22167	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
starfield.jpg	83323	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
sun.jpg	121897	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
seedBottom.jpg	25001	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
seedTop.jpg	36229	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false
layer1.jpg	166736	2022-12-10 03:...	2015-09-11 06:...	E:\ARDUINO AL...	true	false	false

Fig. 2 Metadata Extractor

A software interface called "Metadata Extractor" designed to analyze metadata from files, perhaps for forensic or investigative purposes. Here are the main components and functions:

a) Main Title (Metadata Extractor): This application is dedicated to extracting and analyzing metadata from files or images.
b) Image Viewer: The top left area displays the image. The image is the file being analyzed. Control Panel: The middle area contains buttons and dropdowns to control the extraction and analysis of metadata. Including options to compare metadata or perhaps analyze differences. Metadata View: The bottom section lists detailed metadata attributes extracted from the file being analyzed. This metadata may include File format and type (e.g., JPEG, PNG). Dimensions (height, width). Timestamp (creation, modification, access).
c) Compression type; Camera or device information. Other technical metadata such as sampling factors and quantization tables.Functionality Tabs or Buttons: There are buttons or tabs labeled with options such as "Load Original File" and "Analyze." This may facilitate loading files for analysis and running metadata extraction routines.
d) Color-Coded Results or Status Indications: Red and yellow interfaces may visually indicate status (errors, warnings, or matching discrepancies in metadata).
e) Purpose: The software is designed for forensic or investigative work, allowing users to: Analyze file metadata for consistency or discrepancies. Compare metadata between the original file and other files (potentially to detect tampering or verify authenticity).
f) Extract technical metadata for further processing or documentation

Table 1 Caption

No	Metadata Information	Evidence File Metadata	Suspect’s Laptop File 1	Analysis Result 1
1	Extention	JPG	JPG	Matching
2	Path :	C:\Users\ryzen\Pictures\IMG_7818.JPG	E:\\$BannerWisata.JPG$	Matching
3	creationTime:	2023-02-07 0	2023-02-0	Not Matc
4	lastAccessTime:	202	2023-	No
5	lastModifiedTime:	2	2017	Matching
6	isDirectory:	FALSE	F	Mat
7	isOther:	FALSE	TRUE	Not Matching
8	isRegularFile:	TRUE	TRUE	Matching
9	isSymbolicLink:	FALSE	FALSE	Matching
10	Size	7257627	7257627	Matching
11	Compression Type	Baseline	Baseline	Matching
12	Data Precisi	8 bits	8	Matchin
13	Image Height	3456 pixels	3456 pixels	Matching
14	Image Width	5184 pixels	5184 pixels	Matching
15	Number of Components	3	3	Matching
16	Component 1	Y component: Quantization table 0, Sampling factors 2 hori/2 vert	Y component: Quantization table 0, Sampling factors 2 hori/2 vert	Matching
17	Component 2	Cb component: Quantization table 1, Sampling factors 1 hori/1 vert	Cb component: Quantization table 1, Sampling factors 1 hori/1 vert	Matching

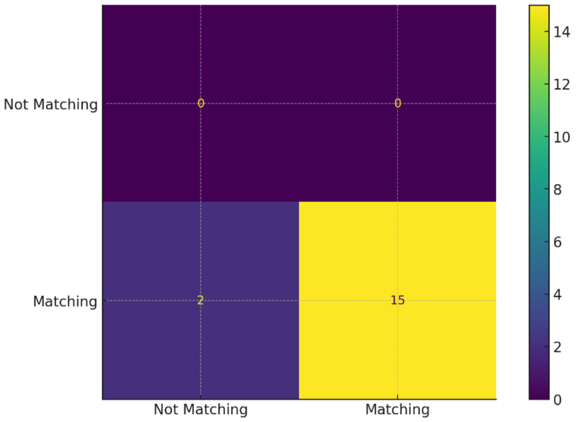


Fig. 3 Confusion Matrix

Based on the results of the analysis of the case of the distribution of photos with pornographic content and a digital forensic analysis related to evidence in the form of photos containing pornographic content where the evidence has been distributed in the community with the file name "IMG_7818.JPG", after an analysis was carried out using forensic metadata file correlation using the file type and size parameters (sub parameters are also used, namely size greater than or equal to) 2 files were found that correlated with the main evidence file, namely files with the names "Banner Wisata.JPG" and "Banner Wisata - Copy.JPG", this can of course be developed by investigators or digital forensic investigators as evidence or references.

The results of this experiment also prove that the application of forensic metadata with existing parameters can find file correlations related to evidence in an easy, effective and efficient time, when compared to the conventional method, namely by searching for files one by one and manually checking the metadata of the files one by one, this will of course take a long time and sometimes the file is missed during the search.

This forensic metadata application is certainly very helpful and makes it easier to speed up the analysis process in a case, besides this application can run on various operating systems (Windows, Linux and Mac), using the Java programming language which is "open source" of course this application can be developed and refined again with methods, parameters or features that can help in the analysis process.

5 Conclusion

The developed Java-based forensic metadata application provides convenience in the process of analyzing digital file metadata. This application is able to read and analyze metadata with parameters such as author, file type, file size, date, and owner, thus accelerating the investigation process compared to manual methods. However, this application is not yet able to determine the authenticity of image files without additional analysis support. Testing using linear regression resulted in a classification accuracy rate of 55%, indicating that there is still room for further development. Overall, the use of parameters such as file size and modification time is quite effective in accelerating the investigation process. However, to improve the performance of the application, it is recommended to add new, more relevant parameters and test the application on various operating systems, such as Linux and Mac OS. Thus, this application can be a more comprehensive and reliable tool in supporting digital forensic investigations.

References

[1] Adhiarso, D. S., Utari, P., and Hastjarjo, S., 2019, “The Impact of Digital Technology to Change People’s Behavior in Using the Media,” Digital Press Social Sciences and Humanities, **2**(2018), p. 00005.

[2] Radicic, D. and Petković, S., 2023, “Impact of Digitalization on Technological Innovations in Small and Medium-Sized Enterprises (SMEs),” Technological Forecasting and Social Change, **191**.

[3] Egho-Promise, E. et al., 2024, “Digital Forensic Investigation Standards in Cloud Computing,” Universal Journal of Computer Sciences and Communications, **3**(1), pp. 23–45.

[4] Ahmed Awan, S. et al., 2022, “Digital Forensics and Cyber Forensics Investigation: Security Challenges, Limitations, Open Issues, and Future Direction,” International Journal of Electronic Security and Digital Forensics, **1**(1), p. 1.

[5] Hershensohn, J., 2000, “I.T. FORENSICS: THE COLLECTION and Presentation of Digital Evidence,” c.

[6] Kessler, G. C., 2010, “Judges’ Awareness, Understanding, and Application of Digital Evidence,” Doctoral dissertation, Nova Southeastern University, Retrieved from NSUWorks, Graduate School of Computer and Information Sciences, https://nsuworks.nova.edu/gscis_etd

[7] Kizza, J. and Migga Kizza, F., 2011, “Digital Evidence and Computer Crime, Securing the Information Infrastructure,” *Digital Evidence and Computer Crime*, IGI Global.

[8] Andreassen, L. E. and Andresen, G., 2019, “Acknowledgments,” .

[9] Iqbal, S. and Abed Alharbi, S., 2020, *Advancing Automation in Digital Forensic Investigations Using Machine Learning Forensics*, Digital Forensic Science.