International Journal of Cyber-Security and Digital Forensics (IJCSDF) 8(3): 206-212
The Society of Digital Information and Wireless Communications (SDIWC), 2019 ISSN: 2305-0011

Forensic Authentication of WhatsApp Messenger Using the Information Retrieval Approach

Nuril Anwar [1], Supriyanto[2]

Department of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia[1,2]
nuril.anwar@tif.uad.ac.id, supriyanto@tif.uad.ac.id

ABSTRACT

The development of telecommunications has increased very rapidly since the internet-based instant messaging service has spread rapidly to Indonesia. WhatsApp is the most popular instant messaging application compared to other instant messaging services, according to the statista website users of WhatsApp services in 2018 showed significant growth by gathering 1.5 billion monthly active users or monthly active users (MAU). That number increased 14 percent compared to MAU WhatsApp in July 2017 which amounted to 1.3 billion. Daily active users aka DAU are in the range of one billion. WhatsApp handles more than 60 billion message exchanges between users around the world. This growth is predicted to continue to increase, along with the wider internet penetration. Along with WhatsApp updates with various features embedded in this application including Web-based Whatsapp for computers, this feature makes it easier for users to share data and can be synchronized with their smartphone or user's computer. Besides the positive side found in the application, WhatsApp also provides a security gap for user privacy, one of which is tapping conversations involving both smartphone and computer devices. The handling of crimes involving digital devices needs to be emphasized so that they can help the judicial process of the effects they have caused Mobile Forensics Investigation also took part in suppressing the misuse of WhatsApp's instant messaging service features, including investigating the handling of cases of WhatsApp conversations through a series of standard steps according to digital forensics procedures. Exploration of evidence (digital evidence) WhatsApp conversations will be a reference to the crime of telecommunication tapping which will then be carried out forensic investigation report involving evidence of the smartphone and computer of the victim.

Keywords: Authentication, Mobile Forensics, Instant Messenger, and WhatsApp Messenger.

1 INTRODUCTION

The development of telecommunications has increased very rapidly since the internet-based instant messaging service has spread rapidly to Indonesia. More than 1 billion people in more than 180 countries use WhatsApp to stay connected with friends and family, anytime and anywhere. WhatsApp is free and offers the ability to send messages and make simple, secure, and reliable calls, which are available for phones around the world **[11].** WhatsApp is the most popular instant messaging application compared to other instant messaging services, according to the user statistics website as of January 2017 as many as 1.2 billion people are actively using this application **[5].** Whatsapp is a cross platform application with versions available for the Android, BlackBerry, iPhone and Symbian operating systems. WhatsApp allows to send text messages, send video messages, and audio media messages. This application is available for Android phones, Blackberry, iOS, Symbian (s60), and Windows. WhatsApp Inc. founded in 2009 by Brian Acton and Jan Koum, both veterans of Yahoo! People exchange information such as pictures, videos, activities and events **[6].** However, on the other hand there are many individuals who abuse WhatsApp to commit digital crimes such as fraud, gambling, pornography, corruption, or drug networks. Forensic Android smartphones have evolved over time offering significant opportunities and interesting challenges. In a number of recent crime cases such as the case of Mirna's murder by Jessica and on the case of pornography chat Habib Rizieq Shihab, using WhatsApp conversations as evidence in court. This shows that from the perspective of forensic investigation, the WhatsApp application can store evidence data that can be used in court as evidence, then the smartphone can be analyzed and if the WhatsApp application is installed on the smartphone, WhatsApp forensic analysis can be done to obtain digital artifacts. in the form of conversation history and contact numbers, etc. related to the Whatsapp application **[12].**

Based on some previous studies it can be concluded that the investigation of conversations through instant messenger application conversations found in chat sessions WhatsApp is vulnerable to security because database files that store chat conversations are not partially unencrypted and can be easily accessed through a series of specific ways to obtain details of the entire conversation, by Therefore, in this study digital forensic procedures will be developed using the text mining Information Retrival Approach method as a reference in message clustering involving digital evidence database from the WhatsApp application (msgstore.db) to retrieve conversation sessions even those that have been deleted from the chat option. This method can classify messages in

accordance with the authentication of the message, so that it can assist in investigating messages in exploring or analyzing digital evidence. The text retrival approach text mining method is chosen because it can play a role in techniques that can be used to classify, because text mining is a development of variations of the data mining process that seeks to find interesting patterns from a large number of text data sets. In addition to classification, text mining is also used to handle clustering, information extraction, and information retrieval problems.

Text classification is needed by calculating Similarity by combining two concepts for weight calculation, namely, the frequency of calculating the appearance of a word in a particular document and Cosine similarity is a method that can be used to calculate the level of similarity between two documents or two objects in this case WhatApp database.

2 LITERATURE REVIEW

Previous research has shown that one can get complete access to all information on WhatsApp whether it's WhatsApp Smartphone or WhatsApp Web. Most chat applications follow the same message, contact and user data synchronization patterns when syinc and update conversation data regularly. The approach taken gives a general outline for all similar applications running on Android and Windows platform devices such as Telegram and the like. WhatsApp provides a security gap for user privacy, one of which is tapping conversations involving both smartphone and computer devices. For handling WhatsApp crime, this includes investigating the handling of cases of intercepting WhatsApp conversations through a series of standard steps according to digital forensics procedures **[5]**.

The development of telecommunications has been increasing very rapidly since instant internet-based instant messaging services to Indonesia. However, many people abuse WhatsApp for digital crimes such as fraud, pornography, or drug networks. For more common civil cases, start using evidence from conversations, pictures, videos, and more from the WhatsApp. The applied Forensic NIST mobile method can get picture, video, or text data contained in WhatsApp. So that text messages can be analyzed and provided information about the indications associated with others using text mining methods. Subjects in this research are to build a web system using Python programming language to be able to identify text sent messages between the two actors. The method used in the method is to use the Tf-Idf and Cosine Similarity methods, which prior to the text mining method is the first perpetrator's message with the NIST Forensics method application to obtain text message conversations. The result of this study is a web system that can identify messages from the perpetrator's conversation whether there is an indication of interference by others or not **[1]**.

Other research study explains that the data that can be taken is the main data and supporting data of the application. The main data in the form of a database containing contacts and conversations and artifacts that make up the application. The supporting data of the application is in the form of a backup database and related media files such as images, videos and sounds. After the data is successfully obtained, it will be analyzed using applications and supporting literature to achieve the objectives of the study. As a standard test, digital forensic applications are used which are widely used and available for free, namely Forensic Tool Kit (FTK) Imager and SQLite Browser **[10]**.

The research carried out by **(Kunang et al., 2017)** resulted in a procedure that could be used as a reference in conducting forensic investigations of WhatsApp applications to obtain evidence in the form of conversation sessions, media data such as audio, contact no., Photos and more. The stages of the forensic analysis procedure carried out in this study succeeded in obtaining proof artifacts in the form of chat sessions, avatars, contact numbers on WhatsApp applications, voice notes, profile photos, identity of WhatsApp account holders and also able to get other media files and most importantly backup database files encrypted. The applied WhatsApp database extraction approach successfully extracts chat conversations stored in internal and external memory using the WhatsApp extractor key and decryptor to convert backup databases into text databases that can be seen in the SQLite database browser. This stage can open chat sessions that have been deleted based on the backup data stored either automatically by the WhatsApp application or manual backup.

Other research produced a system that can detect text documents in pdf format. The system can detect the Indonesian language text, and if there is a foreign language in a document the stemming process is not carried out. Detection is done by one-to-many grooves, in order to facilitate the system in detecting and using the Enhanced Confix Stripping (ECS) stemmer algorithm to produce 23-36% similarity values with experiments with 3 journals with the same category, while stemming produces a similarity of 35- 40% **[4]**.

Blackberry Messenger is one of the popularly used instant messaging applications on Android with user's amount that increase significantly each year. The increase off Blackberry Messenger users might lead to application misuse, such as for commiting digital crimes. To conduct investigation involving smartphone devices, the investigators need to use forensic tools. Therefore, a research on current forensic tool's performance in order to handle digital crime cases involving Android smartphones and Blackberry Messenger in particular need to be done. This research focuses on evaluating and comparing three forensic tools to obtain digital evidence from Blackberry Messenger on Android smartphones using parameter from National Institute of Standard Technology and Blackberry Messenger's acquired digital evidences.

The result shows that from comparative analysis conducted, Andriller gives 25% performance value, Oxygen Forensic Suite gives 100% performance value, and Autopsy 4.1.1 gives 0% performance value. Related to National Institute of Standard Technology parameter criterias, Andriller has performance value of 47.61%. Oxygen Forensic Suite has performance value of 61.90%. Autopsy 4.1.1 has performance value of 9.52% **[3]**.

The next three tier levels are through procedures of physical extractions. The third level is through methods of Hex Dumping, requiring connectivity (Wi-Fi, wired, et cetera) between mobile device and digital forensic workstation. Hex Dumping forces a boot loader onto the device dumping the information that is harvested on the protected parts of the memory (RAM). Forensic analyst use a flasher box connecting the device data port to the digital forensic workstation, then the device is placed in a diagnostic mode, where the analyst can send commands and the flasher box "captures' sections of the memory transporting the data back to the workstation **[2].**

The handling of digital evidence can become an evidence of a determination thatcrimes have been committed or may give links between crime and its victims or crime and the culprit. Soft System Methodology (SSM) is a method of evaluation tocompare a conceptual model with a process in the real world, so deficiencies of the conceptual model can be revealed thus it canperform corrective action against the conceptual model, thus there is no difference between the conceptual model and the real activity. Evaluation on the IDFIF stage is only done on a reactive and proactive process stages in the process so that the IDFIF model can be more flexible and can be applied on the investigation process of a smartphone **[9].**

The research data mining resulted that the cosine similarity method has the highest value compared to Jaccard Similarity. While from the results of grouping with the SNN (Shared Nearest Neighbor) method, the parameters greatly influence the formation of groups, the greater the value, the fewer groups will be formed **[13]**.

Text mining in general is a theory about processing a large number of documents from time to time using several analyzes, the purpose of processing the text is to know and extract useful information from data sources by identifying and exploring interesting patterns in the case of text mining, sources the data used is a collection or collection of unstructured documents and requires a grouping to find out similar information **[7].** Whereas social media is an integral part of online life as a social website and web-based communication. E-communication through chat servers, Instant Messaging Systems, Internet Relay Chat (IRC) is one type of communication that is growing rapidly but suspicious messages are sent via Instant Messenger (IM) and Social Networking Sites (SNS) which are not bound, leading to disruption to network communication and cyberspace security. Text mining techniques are an effective way to predict and detect criminal activity.

3 RESEARCH METHOD

The research method is the stage of conducting a simulation of a case study in exploring the messages of instant messenger conversations in this case WhatsApp, to obtain digital evidence of crime. With various simulations and stages carried out, it also aims to identify or classify digital evidence of crime from the Whatsapp messenger messaging chat application that had previously been done by both sides of the bill or more. Where in the research phase there are several process stages as in Figure 1;

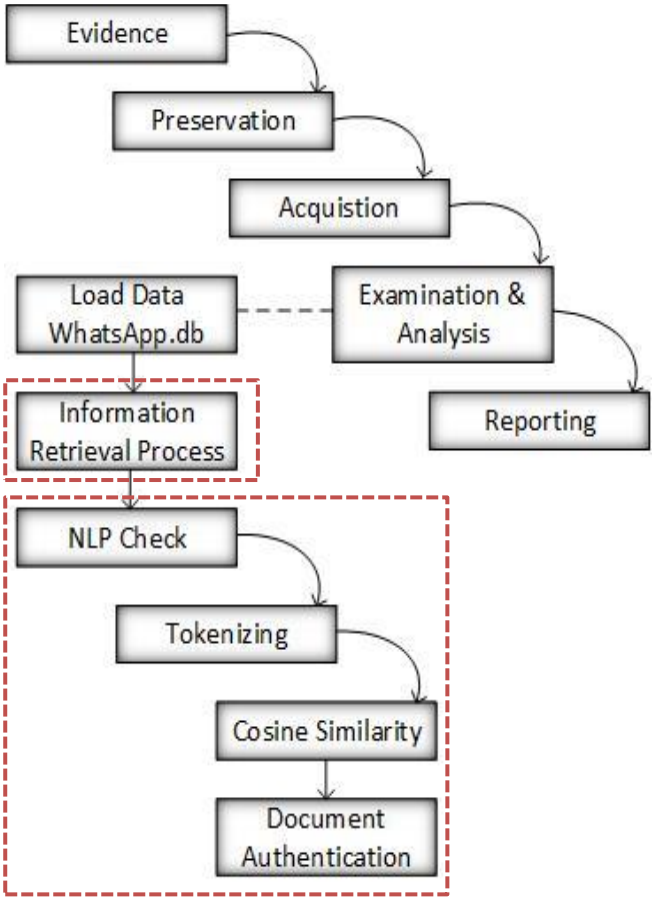

Figure 1. Stages of Analysis of Digital Evidence

In the part of Figure 1. it describes the stages of analyzing cases from evidence findings to weighting calculations from digital evidence which is obtained in the form of paragraph messages chat scenario performed by both actors through a smartphone. The stages are as follows:

1. Evidence; The main evidence obtained from the perpetrator is a smartphone.

2. Preservation; The process by which the investigator maintains the smartphone evidence to conduct data searches, data collection, and documentation of evidence in the form of artifacts from the two perpetrators.
3. Acquistion; The process of collecting and obtaining digital evidence is a conversation message which is then performed imaging of the conversation message for further investigation.
4. Examination and Analysis; is a process where the investigator can explore, analyze, and uncover from imaging results from the stage of previous acquisition to obtain data relating to the perpetrator or crime that is found in the WhatsApp application. From this stage it is then broken down again to find out the frequency of a word in a paragraph or document by calculating the weight of each word.
5. Load data; Conversation message data between the two actors that have been obtained from the stage of examination and analysis is then collected and processed with excel per document to load into the program so that the conversation can be processed by the system.
6. Tokenizing; Stages for cutting strings based on each of the words and their punctuation marks on the conversations of the actors' messages in the form of symbols and changing capital letters to lowercase letters.
7. NLP Check; API (Application Programming Interface), that will provide NLP as Service and Knowledge as service. Our main Language Processing Core are focussed on Indonesian Language.
8. Cosine similarity; Stages where to count sentences in one document with another and categorize or give a percentage value whether there is an indication of difference or not.
9. Document Process; to document the words of the perpetrator and the percentage obtained from the weighting of each paragraph word in the form of table data.

Research scenarios are the initial stages in mobile forensic methods to describe the research that will be conducted, the thing that needs to be done is to search, collect and document evidence. For testing in the study, the samples of evidence analyzed were two smartphones which were scanned as evidence in a crime case. From one of the smartphones in a rooted condition and one still in a state of unroot with the condition of the active password security and active screen security. At this stage documentation of things related to the smartphone is done.

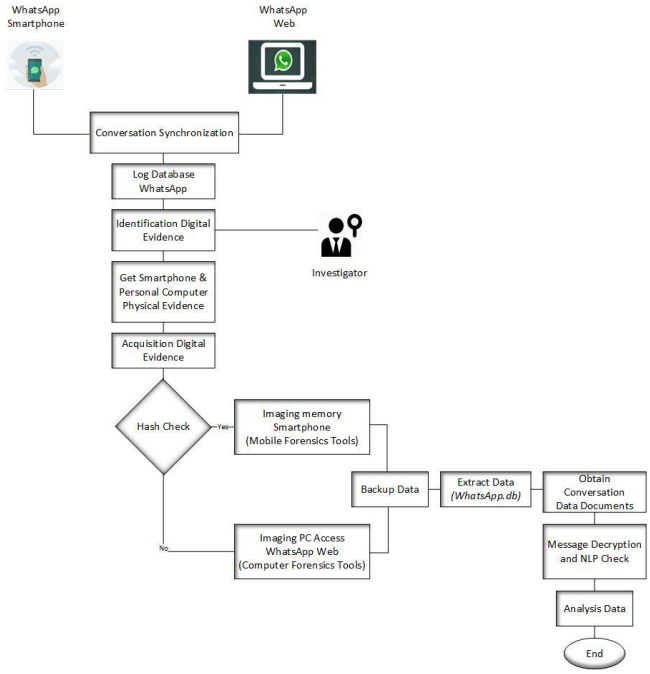

Figure 2. Scenario Analysis of WhatsApp on the Android & Windows Platform

In Figure 2. it can be described that the perpetrator has a conversation through the WhatsApp application, where the WhatsApp application stores various kinds of log messages from the conversation interaction. Both in the form of pictures, videos, paragraph messages, etc. The investigator identifies the case in the Case Event Place based on what he found that can be used as evidence for a legal trial. Where the activity is an activity carried out with the aim of securing existing digital evidence, so that it is not contaminated by other things. In the process of identifying the evidence, the investigator also obtained a finding of physical evidence, namely the perpetrator's smartphone. After the perpetrators get the evidence, then he conducts the process if evidence of mobile forensics. Namely by connecting the offender's smartphone to the computer using a data cable so that it is detected by tools forensics for the evidence imaging process. And if the smartphone is not detected by tools forensics, the investigator can do imaging by taking the SD-card (evidence) first and then using a card reader to get access to evidence so that the SD-card is read by the Investigator application.

Data backup is done to copy the original source of information from the imaging results to evidence that has not been carried out several stages. Where it has a function to maintain the authenticity of digital evidence that is obtained if the investigator loses data due to natural disasters, viruses, errors etc. by using TWRP (Team Win Recovery Project) tools. With TWRP you can retrieve data on the Android system partition and data on internal memory without having to rooting, activating USB Debugging, and also not being affected by the state of the smartphone's screen security is active or not. Then the investigator can carry out the stages of extracting data on the evidence contained in the

smartphone, which is where the stage has the function to open any files or files contained in the results of previous imaging such as videos, images, and message conversations. With so many files that emerged from the extracted file, the investigator began to search for files where the file was a paragraph of conversation messages between the two actors, according to the focus of the research. After all data related to the WhatsApp application has been obtained, then the next step is to decrypt the encrypted WhatsApp application database using the WhatsApp Viewer application. Then the investigator decrypts an encrypted paragraph message and performs a weighting calculation based on the decrip of the paragraph messages sent by the perpetrator to then analyze the paragraph weighting data by comparing the results of the weekly calculation of each document. And the final result of the investigation process, the investigator can record the results of imaging evidence, file encryption and decryption, weighting, and the results of the analysis as an investigation report.

4 RESULT AND DISCUSSION

The results and discussion of the following research discuss how the research steps from data analysis and results obtained from the study. The discussion in this chapter covers the system identification study phase which is used for the smartphone target research object and the data calculation analysis. The identification phase is carried out to obtain evidence obtained on a smartphone in the form of a message conversation. The analysis phase is used to find the value of the results of calculating the weight of each word using the text mining with information retrival approach method.

A. Data Analysis

Investigation using WhatsApp Viewer v1.12 can be done as a way to process the WhatsApp database according to those found on smartphones and computers. From the feature or menu file in the whatsApp viewer, it has several types of files to be able to encrypt messages. Among them are cryp5, cryp7, cryp8, and cryp12 which are compatible with the current version of the WhatsApp application. Whereas to encrypt .db data can be done on one smartphone and computer, where on both evidence devices save some history of the WhatsApp message database. The database has a gap that can be used by forensic investigators in authenticating conversations involving synchronizing data for WhatsApp applications, both web base and smartphone applications.

Mobile forensic research on this smartphone can also be applied by using forensic Oxygen software tools as file explorer and SQLite database as Tools to access conversational data on computers with browsers as their forensic internet application. Remember the .db and key files on WhatsApp as shown in Figure 3.;

Figure 3. ".db" file and Key in WhatsApp Database

In Figure 3. shows some encrypted whatsapp databases and also keys that can be generated using whatsapp viewer, so the investigator can open the contents of the message directly and analyze the data.

a) Crime Simulation & Authentication

The simulation process of proof of action or authentication of WhatsApp conversations is a stage in identifying conversational messages that have been decrypted by the system to be used as evidence by investigators between computer devices and smartphones, whether there is an indication of interference by others or not. Data simulation or analysis starts from a number of clustering text processes or words of conversations of actors who are then made a comparison, where the investigator performs several steps to get the results of word weighting and the percentage of similarities that will later be used as comparative values.

The scenario of case simulations carried out by the two actors can be described for a proof object. Where the two devices synchronize data and through messages sent and summarized in a WhatsApp database, and the weighting is used as proof as well as a method to get the value of each message sent whether there is an indication of other people's interference or not in the scenario Figure 4.;

Figure 4. Crime and Authentication Scenarios

The evidence of the perpetrators in this study investigators can do several processes using the Indonesian NLP text matching method, measurements, and cosine similarity tests used in handling a case. Then the investigator can analyze the data with a comparison based on the investigation he did.

b) Flowchart of Forensic Text Mining Investigation

Investigators in conducting message investigations in this study can be seen in the investigator's standard operational workflow. Where in the process there are several stages as shown in Figure 5.;

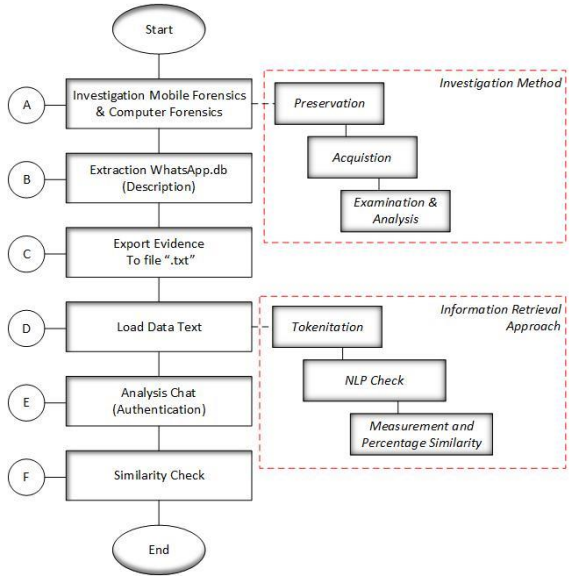

Figure 5. Flowchart WhatsApp Messager Investigation

The WhatsApp Smartphone (Android) database can be opened using tools forensics, the arrangement of key file storage folders and WhatsApp .db data as in Table 1. ;

Table 1. *File Directory Database Whatsapp Smartphone*

File Type	File Name	Storage Location
Db WhatsApp	*WhatsApp\Database*	*Mgstore.db.crypt 12*
Key WhatsApp	*\data\data\com.WhatsApp\files*	*Key*
Db WhatsApp	*\data\data\com.WhatsApp\databases*	*Web_sessions.db*

The WhatsApp Web (Windows) database can be opened using tools forensics, then the arrangement of the Whatsapp.SQLiteDB data storage folder as shown in Table 2.;

Table 2. File Directory Database Whatsapp Web

File Type	Storage Location	File Name
Db SQLite	*C:\Users\Adm\AppData\Local \Google\Chrome\User\Data\ Default*	*History*
Db SQLite	*C:\Users\Adm\AppData\Local \Google\Chrome\User\Data\ Default*	*Login*
Cache	*C:\Users\Adm\AppData\Local \Google\Chrome\User\Data\ Default\cache*	*Cache*

Export evidence WhatApp Smartphone is intended to make WhatsApp text conversation files into .txt files in this case only limited to Indonesian, so the conversation looks like Figure 6.;

Figure 6. WhatsApp Smartphone Conversation Files in Indonesian Language

Export evidence WhatsApp Web is intended to create text conversation files after synchronization becomes a .txt file like Figure 7.;

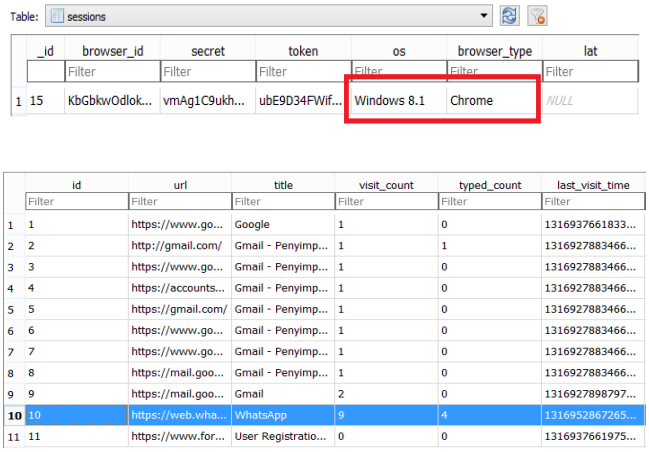

Figure 7. WhatsApp Web Conversation Files

From the evidence.txt export file as shown in Figure 7, it converts using CSVConver tools to a .csv file so that it can be accessed using exel before clustering. Using API NLP http; // nlp.yuliadi.pro/ looks like Figure 8; **[8].**

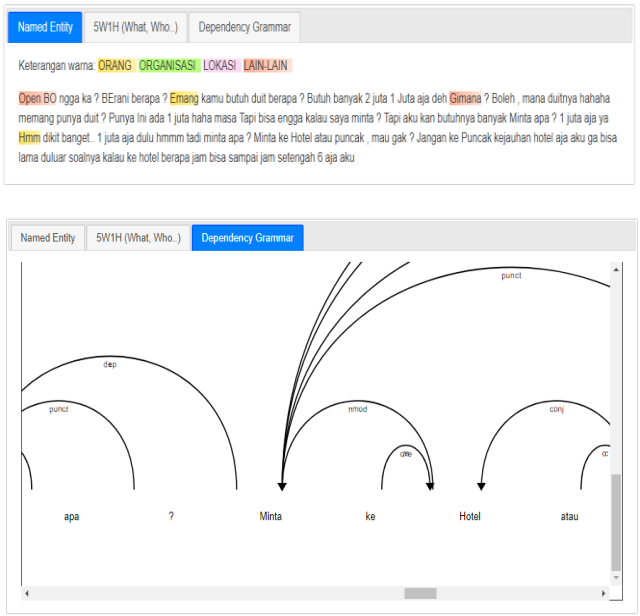

Figure 8. Indonesian Language API NLP

Load data is intended to process or process the conversational data of the perpetrator to analyze the weighting using the text mining stages such as tokenisation. After obtaining a percentage of data, it can be analyzed by categorizing the percentage into moderate, sufficient, or proven indications of interference. others as in Table 2.;

Table 2. Cosine similarity Authentication results

Conversation Document	Similarity Value	Similarity Value (%)	Text Mining Results
Chat WhatsApp Smartphone	0,1285	12,85 %	Similarity
Chat WhatsApp Web	0,5419	54, 19 %	Is being

5. Conclusion

Based on the results of the study "Analysis of the Forensics WhatsApp Messenger Method Using Information Retival Approach" it can be concluded that a mobile forensics analysis related to cases involving digital evidence smartphone with the WhatsApp messenger application and crime has been identified involving two mutually integrated devices namely WhatsApp Smartphone and WhatsApp web. All crimes involving digital evidence of instant conversations need to handle special incident handlers by using the standard of forensic investigations of crimes involving WhatsApp instant messaging applications. The application of the text mining Information Retival Approach method can be used as an investigative method, especially the generation of text in conversation applications and helps the digital evidence exploration process related to cybercrime in the realm of instant messenger with the WhatsApp application and can be applied to other similar instant messaging applications.

References

[1] A Marfianto, I Riadi. WhatsApp Messenger Forensic Analysis Based on Android Using Text Mining Method. International Journal of Cyber-Security and Digital Forensics (IJCSDF) 7(3): 319-327. (2018).

[2] Ayers, R., Brothers, S. & Jansen, W. Guidelines on mobile forensic devices. (NIST SpecialPublication 800-101 Revision 1)," NIST Spec. Publ., Vol. 1, no. 1, p. 85. (2014).

[3] I. Riadi, Sunardi;, and A. Firdonsyah, Forensic Tools Performance Analysis on Android-based Blackberry Messenger using NIST Measurements, International Journal of Electrical and Computer Engineering (IJECE), vol. 8, no. 5. (2018).

[4] M. S. L. Ade Chania Sion Sagala and R. F.Rahmat, Detection of Similarities to Text Documents Using Combination of Enhanced Confix Stripping Stemmer and Winnowin Algorithms Detection of Similarities in Text Text with Enhanced Algorithm Combination Stripping and Winnowing Algorithm. (2015).

[5] N Anwar, I. Riadi. Forensic Investigative Analysis of WhatsApp Messenger Smartphone Against WhatsApp Web-Based, J. Ilm. Tech. Electro Comput. and Inform., vol. 3, no. 1, pp. 1-10. (2017).

[6] N. S. Thakur, WhatsApp Forensic Analysis on Android Smartphones. (2013).

[7] O. Somantri and S. Wiyono, K-Means Method for Optimization of Classification of Student's Final Project Using Support Vector Machine (SVM), Sci. J. Informatics, vol. 3, no. 1, pp. 34-45. (2016).

[8] Rahmi, F. and Wibisono, Y. The SMS Spam Filtering application on Android uses Naive Bayes , Unpublished manuscript. (2016).

[9] Ruuhwan, I Riadi, Y Prayudi Evaluation of Integrated Digital Forensics Investigation Framework for the Investigationof Smartphones Using Soft System Methodology, International Journal of Electrical and Computer Engineering (IJECE), Vol.7, No.5, October2017, pp. 2806~2817. (2017).

[10] S. Ikhsani and C. Hidayanto, Whatsapp and LINE Messenger Forensic Analysis with Strong and Valid Evidence in Indonesia. Tek. ITS, vol. 5, no. 2, pp. 728-736. (2016).

[11] WhatsApp Inc, "WhatsApp," [Online]. Available https : www.whatsapp.com/about/. (2018).

[12] Y. N. Kunang, Implementation of forensic procedures for whatsapp applications on android phones, vol. 11. (2017).

[13] Zahrotun, L. Comparison Jaccard similarity, Cosine Similarity and Combined Both of the Data Clustering With Shared Nearest Neighbor Method. Computer Engineering and Applications, 5(11), 2252–4274. (2016).