Journal of Engineering Science and Technology
Vol. 17, No. 2 (2022) 1408 - 1421

CYBERBULLYING DETECTION ON INSTANT MESSAGING SERVICES USING ROCCHIO AND DIGITAL FORENSICS RESEARCH WORKSHOP FRAMEWORK

IMAM RIADI[1], SUNARDI[2], PANGGAH WIDIANDANA[3,*]

[1]Department of Information System, Faculty of Applied Science and Technology,
[2]Department of Electrical Engineering, Faculty of Industrial Technology,
[3]Department of Informatics, Faculty of Industrial Technology,
Universitas Ahmad Dahlan,
Jl. Ahmad Yani, Tamanan, Banguntapan, Bantul, Yogyakarta 55191, Indonesia
*Corresponding Author: panggah1808048029@webmail.uad.ac.id

Abstract

In recent years cybercrime has increased rapidly in line with the growth of smartphone users. The development of messenger applications has opened up opportunities for cybercrime. Any cybercrime will leave digital evidence that can be tracked and used as digital evidence that applies in court. One of the acts of cybercrime is cyberbullying, in Indonesia, it is ranked 3rd in the world and if it is not followed up, cases of cyberbullying will be rampant. This study carried out the acquisition of digital evidence on the WhatsApp instant messenger application using the Digital Forensics Research Workshop (DFRWS) framework which is a scientific method that has a basis and is proven in court. The Rocchio method is used to investigate cyberbullying cases. Data collection in this study is with dummy data simulated in a cyberbullying scenario on WhatsApp Group Messenger. Research has successfully proven the occurrence of cyberbullying in groups. Obtained the highest R value (0.25) which exceeds the threshold value to say cyberbullying occurs (0.15). Therefore, based on the results of the study, it can be concluded that cyberbullying has occurred in the case scenario studied in this study with 3 bullying perpetrators with different levels, namely Dave, Alex, and Clara, while Jhon is the victim of bullying.

Keywords: Cyberbullying, DFRWS, Digital evidence, Digital forensics, Rocchio.

1. Introduction

Digital evidence in recent years is found so much, this is because cybercrime has increased rapidly and the growth of technology services, applications, and smartphone users [1, 2]. Statista's online web survey reports that smartphone users in the world always increase from 2.5 billion users in 2016 to 3.5 billion users in 2020 and 3.8 billion in 2021 [3]. These data prove that the growth is very fast. This fact has implications for both positive and negative impacts [4]. Technology in a smartphone is directly proportional to the features offered to meet user needs [5]. One application that is well known and fulfils the needs of various users is instant messenger. Applications on smartphones will open opportunities for cybercriminals to commit crimes via messenger. Digital crime is traces of criminal activity that are analysed to be used as evidence. WhatsApp is one of the most popular instant messaging applications [6]. The online survey institute dailysocial.id in 2017 reported that 97.24% of respondents had used the WhatsApp application, and 61.81% of them said that they used the application frequently [7]. The negative impact of the WhatsApp application is cyberbullying. In Indonesia, cyberbullying cases are ranked 3rd in the world because they are easy to do without having to face to face and are considered safe for bullies with victims via WhatsApp. Cyberbullying makes psychosocial stress very serious [8].

Each crime can be traced to digital evidence [9, 10]. WhatsApp is an application that is often used as a reference for investigators in the field of digital forensics in Indonesia [11]. Several researchers researched to discuss digital crime, including a study entitled Investigation of Cyberbullying on Kik Messenger using the National Institute of Standards Technology Method to produce data acquisition from the Kik Messenger application, but this study has not been able to identify cases of cyberbullying systematically, so further research is needed regarding evidence that has been obtained from the forensic results. Another similar research entitled Audio Forensics on Smartphone with Digital Forensics Research Workshop (DFRWS) Method; the results that have been obtained are being able to acquire digital evidence in the form of audio.

This research was conducted for the acquisition of digital evidence on WhatsApp using the Digital Forensics Research Workshop Framework (DFRWS) and investigations of cyberbullying cases using the Rocchio method. DFRWS is a scientific method that has a basis and is proven to be more complete than NIJ and NIST [12]. The Rocchio method is a relevant feedback method that is very effective in the process of searching for library catalogues [13]. The Rocchio method will be carried out through a text mining process, namely, the TF-IDF method which will assist the calculation process of the Rocchio method for the case scenario in this study. The research that will be carried out will make the cyberbullying scenario in a group of four people and one in this group becomes the victim of bullying in that group.

2. Research Method

2.1. DFRWS framework

DFRWS is the use of scientific methods that have a basis and proven for the maintenance, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence originating from digital sources

to facilitate or continue the reconstruction of events that contain criminal offenses or help to anticipate unauthorized actions that prove disruptive for planned operations.

The stages in conducting investigations in digital forensics using the DFRWS framework as follows [14]:

i. Identification for determining the needs that will be needed for investigations and digital evidence searches.
ii. Maintenance is to maintain the evidence and ensure the authenticity or integrity of the evidence so that the evidence is genuinely valid/valid.
iii. Collection for identification to gather sources of evidence that have the potential to become strong evidence.
iv. Examination to determine what will be analysed or better known as data filtering, so that the investigator can focus more on carrying out the next step.
v. Analysis to find and process data, including where the data is obtained from, who made it, and how the data was generated.
vi. Presentation is reporting and presenting the results of the analysis so that the public can understand it.

2.2. Text mining

Text mining is the process of extracting patterns from a large number of unstructured data sources [15]. The purpose of text mining is to find new information that was not previously disclosed by text data processing based on specific regulations. The process of mining the text can be seen in Fig. 1.

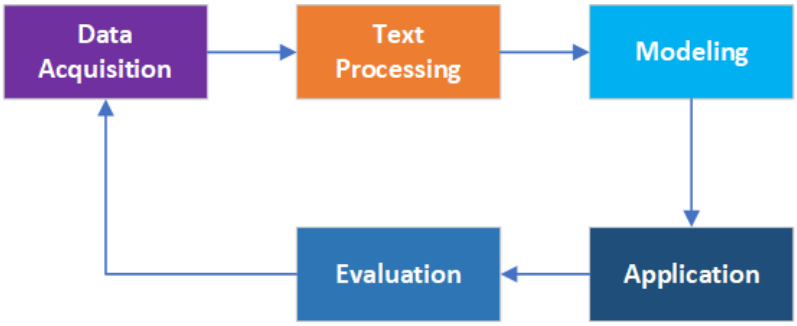

Fig. 1. The process of extracting the text.

Figure 1 describes the sequence of the process of extracting the text, which starts with data acquisition and then processes the text, then modelling, and the results of the modeling will produce applications and evaluation.

Data acquisition is the stage to collect data or can be interpreted of data retrieval from documents to be processed. The preprocessing is the initial stage for text processing [16]. Preprocessing is the initial phase of text mining. This stage covers all the routines and procedures of preparing data for use in the knowledge discovery operation of the text mining system [17].

The preprocessing is a stage to process text data consisting of several steps, namely parsing, stop words removal, stemming, and analysing, as shown in Fig. 2.

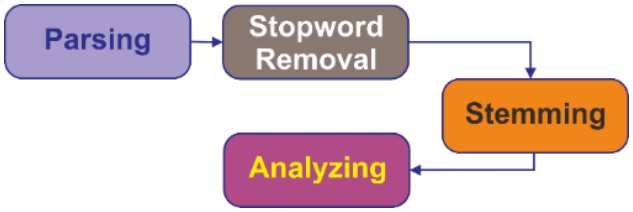

Fig. 2. Preprocessing stage.

Figure 2 shows the flow of stages for preprocessing with the following explanation [18]:

- Parsing or tokenizing is changing the form of sentences into several words in sentences and eliminating punctuation in sentences [19].
- Eliminating stop words is the process of removing essential words in a sentence from the previous stage. These words are already listed in the dictionary (stop list) [20].
- Stemming is the process of removing conjunctions or the operation of returning words to a root word [21].
- The analysis phase is the search for words that lead to cyberbullying in identifying cases of cyberbullying.

The results of text processing will be modeling, which is the application of methods such as machine learning methods, classification, and clustering. Evaluation is carried out to test the calculation results in modeling, whether valid or not. At this stage, recall and precision are used.

The application is the result of applying minimal text, examples that can be seen, such as search, categorization, spam detection, etc. Application and evaluation can be fixed repeatedly if there is something that needs to be improved in the process of text mining starting from data acquisition.

The results of text mining will speed up the process of detection of cyberbullying in a conversation that has occurred within the group. Words will be detected less and less because it has been selected using text mining.

2.3. Term frequency - Inverse document frequency (TF-IDF)

Term Frequency (TF) is a way to find the weight of a document [22]. Where will be seeking the number of occurrences of the term in the paper? The higher the appearance of time, it will affect the amount of weight and the suitability value. TF is frequency of the term t in each document, which is then used to calculate the TF-IDF weighting, as shown in Eq. (1).

$$W(d,t) = TF(d,t) \tag{1}$$

Inverse Document Frequency (IDF) is a method for calculating the distribution of terms in documents [23]. IDF can be seen in Eq. (2).

$$idf_t = log_{10}\left(\frac{N}{df_t}\right) \tag{2}$$

where N is the total documents in a conversation that occurred in the WhatsApp application, and df_t is the number of documents containing the target word. The

smaller the number of documents containing the target word, the greater the weight of the IDF.

TF-IDF is a multiplication of TF and IDF [24] therefore term weight value (β) obtained as Eq. (3).

$$\beta = TF \times IDF \tag{3}$$

2.4. Rocchio method

Rocchio method relevance feedback is a strategy query reformulation that is most popular because it is often used to help novice users retrieve information systems [25]. In the relevance feedback cycle, the user is presented with the document search results, after which the user can check and mark the documents that are truly relevant [26, 27]. Relevance feedback is useful for bringing the query closer to the average of relevant documents and away from the average of irrelevant documents. This can be done through the addition of query terms and adjusting the weight of the query term so that it matches the usefulness of the term in its function of distinguishing relevant and irrelevant documents [28]. The Rocchio technique applies these limits in the form of centroids to give these limits [29]. Centroid is a class c is the average of all vectors in class c. The centroid formula can be seen in Eq. (4).

$$\mu(c) = \frac{1}{|Dc|} \Sigma_{d \in Dc} \vec{v} \tag{4}$$

where D_c is the set of documents in the corpus in class c., and $\vec{v}$ is the normalized document vectors

Determining the similarity of two space vector models there are two ways, namely by measuring distance or by measuring similarity. In determining the distance (distance) between two vector space models used Euclidean distance. The space vector formula can be seen in Eq. (5).

$$d(d_1 - d_2) = \sqrt{\Sigma_{i=1}^{n} (w_{i.d1} - w_{i.d2})^2} \tag{5}$$

where $d_1 - d_2$ is the distance vector space from document 1 and document 2, $w_{i.d1} - w_{i.d2}$ is the word weight of document 1 is reduced by the weight of document 2.

Calculating the similarity (similarity) between two document vectors is in Eq. (6)

$$sim(d_1 . d_2) = \frac{\vec{v}(d_1).\vec{v}(d_2)}{|\vec{v}(d_1)|.|\vec{v}(d_2)|} \tag{6}$$

If a query is processed into a space vector, it can be compared with each centroid class in the corpus. With two approaches looking for the similarity of two space vectors. Query vectors considered to be similar to a class centroid can be done using distance or using similarity. If using distance, what you are looking for is the class that has the smallest distance to the query. And if you use the similarity you are looking for is the class that has the greatest similarity to the query, like the one below:

- Vector space

$$\min |\vec{\mu}(c) - \vec{v}(d)| \tag{7}$$

- Similarity

$$\max \cos((\vec{\mu}(c), \vec{v}(d)) \tag{8}$$

2.5. Case scenario

Figure 3 is a cyberbullying simulation using instant messenger in a group, and one member in the group is a victim of bullying from other members. The victim reported to the investigator and proceeded to the trial investigation using the DFRWS method.

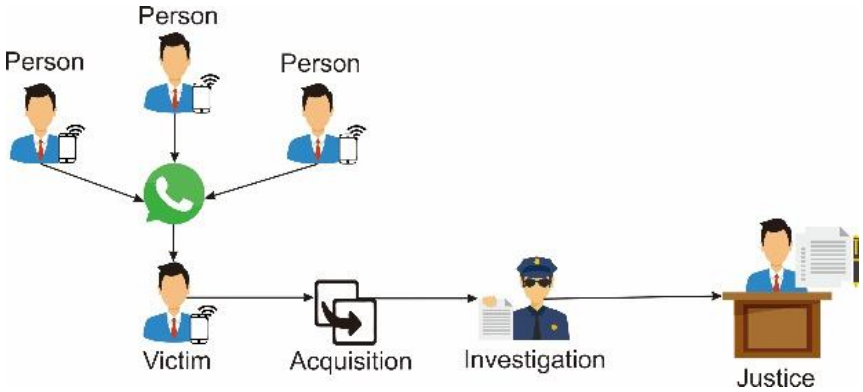

Fig. 3. Cyberbullying scheme.

3. Results and Analysis

3.1. Identification

Identification has been done by providing information on digital evidence that has been obtained as shown in Fig. 4. There are several forms to identify evidence such as the time when imaging will be occurred, the language used, case labels, case evidence numbers, case notes, and so forth. Identification needs to make it easy for digital forensic processes. MobilEdit application can identify which is the initial stage of the DFRWS method, the application has made it easier to make reports that are in the final stages of the DFRWS method.

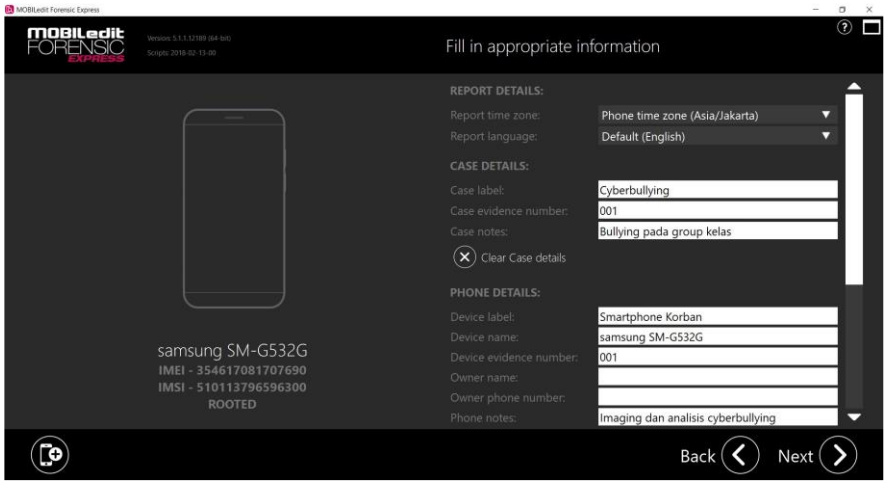

Fig. 4. Identification.

3.2. Preservation

Preservation is carried out to maintain the authenticity and security of evidence so that the imaging process is needed to retrieve proof on the victim's smartphone that has been obtained. The imaging process can be seen in Fig. 5 which is carried out in MOBILedit application for security and maintain the authenticity of the evidence. This stage is the stage that will determine whether the evidence is valid or not in court as digital evidence.

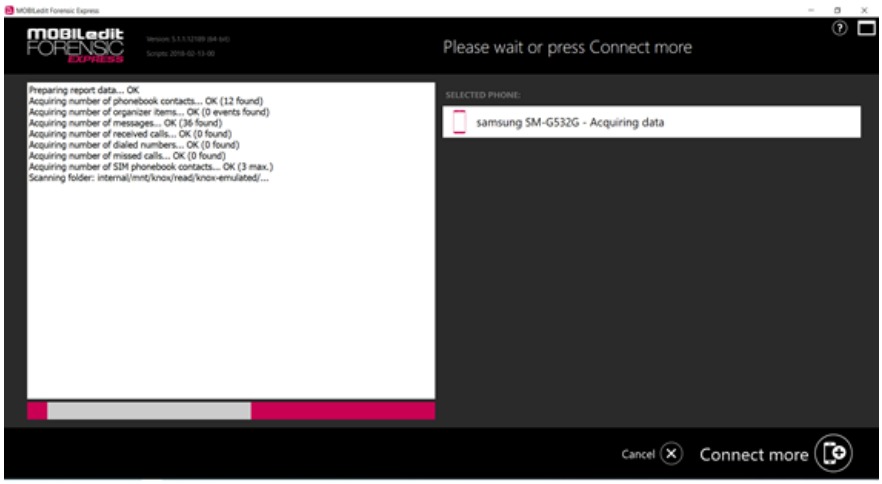

Fig. 5. Preservation.

3.3. Collection

The collection process is imported from the result of creating a physical image or the imaging process of the victim's smartphone at the preservation stage. Fig. 6 is a collection process that will be continued at the next stage, namely the examination stage. The collection stage is used to dissect the contents of the evidence that has been done imagin the previous stage for the forensic investigation process.

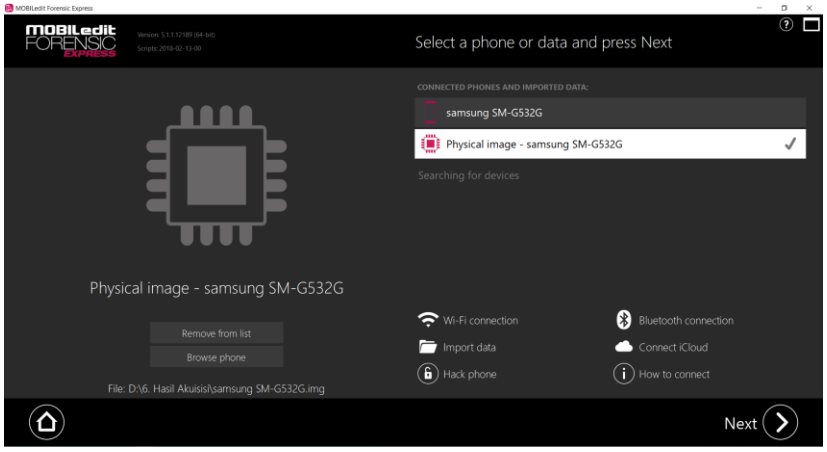

Fig. 6. Collection.

3.4. Examination

Examination the process is done by looking for applications that will be investigated. In this study, the use to be analysed or investigated is WhatsApp Messenger, Fig. 7 is the process of examination or filter of evidence following the needs of investigators in analysing digital evidence. The examination stage helps investigators be safe from misuse of evidence, because in the investigation process an investigator is only.

Figure 7 is the application result that will be investigated further at the Analysis stage, a way to select an application by entering the name of the application to be investigated.

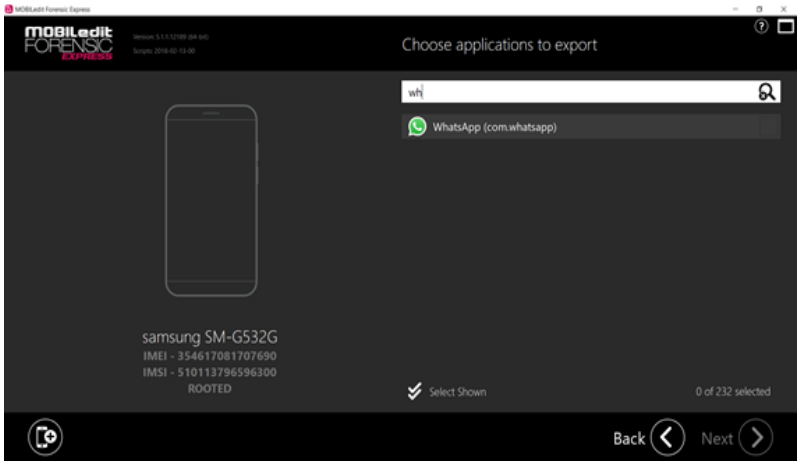

Fig. 7. Examination.

3.5. Analysis

The analysis process is carried out by analysing the results obtained using the MOBILedit application and then filtering those needed for the cyberbullying identification process. The analysis process shown in Fig. 8.

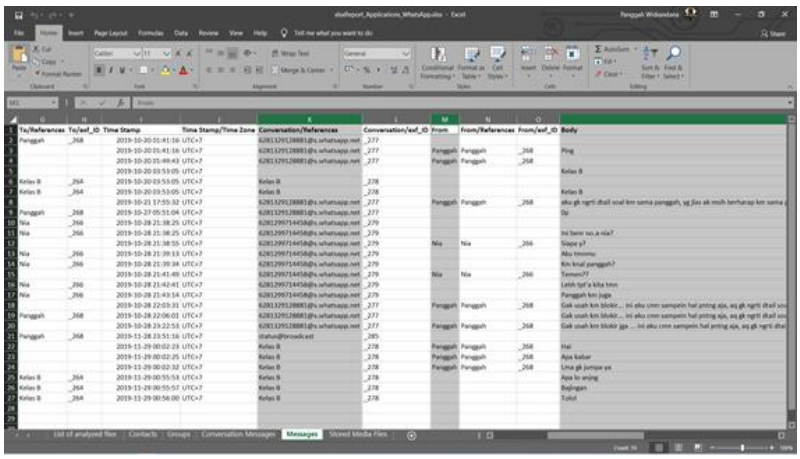

Fig. 8. Analysis.

Figure 8 is the process of cyberbullying analysis of the results of the mobile edit application. The results that have been found are then filtered to be taken several fields such as Conversation/ Reference, From, and Body needed for text mining to identify cyberbullying. Areas that have been selected will be known as the group name, sender's name, and the conversation that occurred in the group. The cyberbullying analysis process on WhatsApp Messenger is the result of selecting the fields needed for the cyberbullying identification process so that it can be preprocessing and applying the cosine similarity method to identify cyberbullying done in the group. The stage is carried out to identify cyberbullying that is data taken from the simulation data of a group of four people who have a conversation in the group, conversation dialogue can be seen in Table 1.

Table 1. Conversation.

User	Conversation
Dave	What the fuck friend! why didn't you say it first?
Alex	Is it true that Jhon did that ??
Clara	Hell yeah!
John	I'm sorry, this is indeed my fault
Dave	Shit! I get the bad score read from you
Dave	Get off my back now! go away!
Jhon	I'm sorry, I will clarify to the teacher
Clara	Let him fix everything Dave, be patient
Alex	You have to take responsibility, Jhon

The conversations that have been carried out are the result of data acquisition and then preprocessing will be carried out to speed up and make it easier to identify cases of cyberbullying. The first stage of tokenizing, filtering and stemming the results can be seen in the Table 2.

Table 2. Prepocessing.

User	Conversation
Dave	fuck friend didn't say firs
Alex	damn true jhon
Clara	hell yeah
Jhon	I'm sorry indeed my fault
Dave	shit get bad score of
Dave	get back go away
Jhon	I'm sorry clarify teacher
Clara	let fix everything dave patient
Alex	to take respon jhon

After preprocessing, the next step will be the TF-IDF calculation using Eqs. (1), (2), and (3) to obtain the weight of words or the results of multiplication between TF and IDF to meet the components in Eq. (4). Table 2 is a TF calculation piece and IDF value, the value of the product can be seen in Table 3. The yellow blocks in Tables 3 and 4 are the markers for bullying dirty word query lines.

Table 3. TF-IDF.

Term	TF				DF	IDF
	Dave	Alex	Clara	Jhon		
fuck	1.00	0.00	0.00	0.00	1.00	0.60
friend	1.00	0.00	0.00	0.00	1.00	0.60
didn't	1.00	0.00	0.00	0.00	1.00	0.60
say	1.00	0.00	0.00	0.00	1.00	0.60
first	1.00	0.00	0.00	0.00	1.00	0.60
shit	1.00	0.00	0.00	0.00	1.00	0.60
get	2.00	0.00	0.00	0.00	1.00	0.60
bad	1.00	0.00	0.00	0.00	1.00	0.60

Table 4. TF-IDF Multiplication.

Dave	Alex	Clara	Jhon
0.60	0.00	0.00	0.00
0.60	0.00	0.00	0.00
0.60	0.00	0.00	0.00
0.60	0.00	0.00	0.00
0.60	0.00	0.00	0.00
0.60	0.00	0.00	0.00
1.20	0.00	0.00	0.00
0.60	0.00	0.00	0.00

The results of the TF-IDF multiplication calculation for each document will be carried out multiplication of the query which can be seen in Table 5.

Table 5. TF-IDF Multiplication query and document.

Dave	Alex	Clara	Jhon
0,36	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0,36	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.72	0.36	0.36	0.00

The green line indicates the total number of times the query weight is weighted by the weight of each document. After searching with the query then looking for vector space which can be seen in Table 6.

Table 6. Vector space.

Q	Dave	Alex	Clara	Jhon
0.36	0.36	0.00	0.00	0.00
0.00	0.36	0.00	0.00	0.00
0.00	0.36	0.00	0.00	0.00
0.00	0.36	0.00	0.00	0.00
0.00	0.36	0.00	0.00	0.00
0.36	0.36	0.00	0.00	0.00
0.00	1.45	0.00	0.00	0.00
0.00	0.36	0.00	0.00	0.00
1.45	5.80	3.26	2.54	4.71
1.20	2.41	1.81	1.59	2.17

Green coloured lines indicate the total amount of vector space multiplication and blue coloured lines are the root of the amount of vector space. Tables 4 and 5 will be entered using Eq. (6).

$$R(Dave) = \frac{0.72}{(1.20 \times 2.40)} = 0.25$$

$$R(Alex) = \frac{0.36}{(1.20 \times 1.81)} = 0.16$$

$$R(Clara) = \frac{0.36}{(1.20 \times 1.59)} = 0.18$$

$$R(Jhon) = \frac{0.00}{(1.20 \times 2.17)} = 0.00$$

3.6. Presentation

Calculations using Rocchio relevance get the result that the highest value is obtained by a user named Dave with Rocchio reaching number (0.25). Alex obtains the lowest value with (0.16). A person's value can be said to do bullying if someone feels depressed behavior more than 61% [30]. The threshold in bullying is said to be big if the perpetrators commit 61% bullying and make the victim depressed and so on. Research on the basis of 61% will be multiplied by the maximum limit of someone doing bullying, in this case the one who has the highest level of bullying is "Dave" so that a calculation of 61% of the value of "Dave" and produces a threshold (0.15) is the limit that determines who the perpetrators are who have been bullying the group. The threshold value concludes that there are 3 people who carry out bullying with different levels, namely Dave, Alex, and Clara, while Jhon is a victim of bullying. The data obtained in this study is a continuation of previous research, namely after obtaining digital evidence from the acquisition results, it will be continued to detect cyberbullying cases. The reporting format that will be submitted to the court is documentation of each step in the digital forensic research workshop framework.

4. Conclusion

The process of collecting evidence using the digital forensics research workshop framework can produce evidence that has validity and can apply the rocchio method in the analysis phase to help process evidence in order to identify cyberbullying cases that are in the group of an instant messaging. The highest rocchio value that has been obtained for Dave is 11.10 higher than threshold of (0.15). Therefore, it can be said that the identification obtained is based on the level of cyberbullying that has been occured. Future research is expected to compare identification methods to prove the effectiveness of the method in identifying cyberbullying cases.

Acknowledgment

This research was funded by Universitas Ahmad Dahlan.

References

1. Akbar, Z.; Nugraha, B.; and Alaydrus, M. (2016). WhatsApp forensics pada android smartphone : A survey. *Sinergi*, 20(3), 207-212.

2. Albert; Kristanda, M.B.; and Hansun, S. (2017). Studi Kelayakan dan Perancangan Aplikasi Pencarian Buku pada Katalog Perpustakaan Menggunakan Rocchio Relevance Feedback -(Studi Kasus: Perpustakaan UMN) (Feasibility study and design of book search applications in library catalogues using Rocchio relevance feedback). *Jurnal Ultimatics*, 8(1), 37-43.
3. Anwar, N.; and Riadi, I. (2017). Analisis Investigasi Forensik WhatsApp Messanger Smartphone Terhadap WhatsApp Berbasis Web (WhatsApp messenger smartphone forensic investigation analysis against web-based WhatsApp). *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 3(1), 1-10.
4. Soyusiawaty, D.; Jones, A.H.S.; and Widiandana, P. (2018). Similarity detection of student assignments using Rocchio method. 2018 12*th International Conference on Telecommunication Systems, Services, and Applications* (*TSSA*). Yogyakarta, Indonesia, 1-4.
5. Dragut, E.C.; Fang, F.; Sistla, A.P.; Yu, C.T.; and Meng, W. (2009). Stop word and related problems in web interface. *Proceedings of the VLDB Endowment*, 2(1), 349-260.
6. Suprianto; Sunardi; Fadlil, A.; and Sunardi, S. (2019). Aplikasi sistem temu kembali angket mahasiswa menggunakan metode generalized vector space model (Application of information retrieval for opinion student questionnaire using generalized vector space model methods). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 6(1), 33-40.
7. Harjanta, A.T.J. (2015). Preprocessing Text untuk meminimalisir kata yang tidak berarti dalam proses text mining (Preprocessing text to minimize meaningless words in the text mining process). *Journal of Informatics UPGRIS*, 1(1), 1-9.
8. Kwan, I.; Dickson, K.; Richardson, M.; MacDowall, W.; Burchett, H.; Stansfield, C.; Brunton, G.; Sutcliffe, K.; and Thomas, J. (2020). Cyberbullying and children and young people's mental health: A systematic map of systematic reviews. *Cyberpsychology, Behavior, and Social Networking*, 23(2), 72-82.
9. Larasati, T.D.; and Hidayanto, B.C. (2017). Live forensics analysis for comparison of instant messenger applications on Windows 10 operating system. *Seminar Nasional Sistem Informasi Indonesia* (*SESINDO*). Surabaya, Indonesia, 245-256.
10. Lumbanraja, F.R. (2013). Text data search system using Rocchio classification method (Case study: Thesis text document). *Kumpulan Makalah Seminar dan Rapat Tahunan BKS PTN Wilayah Barat*. Bandar Lampung, Indonesia, 217-224.
11. Faiz, M.N; Prabowo, W.A.; Sidiq, M.F. (2018). Studi komparasi investigasi digital forensik pada tindak kriminal (Comparative study of forensic digital investigations on crime). *Journal of Informatics, Information System, Software Engineering and Applications* (*INISTA*), 1(1), 63-70.
12. Nurhayati, S. (2010). *Implementation of text mining for classification of traditional arts using the NBC* (*Naïve Bayes Classifier*) *method*. Fakultas Teknik dan Ilmu Komputer, Universitas Komputer Indonesia.
13. Pandie, M.M; and Weismann, I.T.J. (2016). Pengaruh cyberbullying di media sosial terhadap perilaku reaktif sebagai pelaku maupun sebagai korban cyberbullying pada siswa Kristen SMP nasional Makassar (The effect of

cyberbullying on social media on reactive behavior as perpetrators and victims of cyberbullying in Christian students at Makassar national junior high school). *Jurnal Jaffray*, 14(1), 43-62.

14. Prasetya, D.D.; Wibawa, A.P.; and Hirashima, T. (2018). The performance of text similarity algorithms. *International Journal of Advances in Intelligent Informatics*, 4(1), 63-69.
15. Purbo, O.W. (2017). Text mining : Analisis MedSos, kekuatan brand dan intelejen di internet (*Text mining MedSos analysis, brand strength and intelligence on the internet*). Yogyakarta: Andi Offset.
16. Riadi, I.; Sunardi; S.; and Kadim, A.Z. (2019). Monitoring log aplikasi mobile native menggunakan framework grr rapid response (Monitoring mobile native application logs using the grr rapid response framework). *Jurnal Buana Informatika*, 10(1), 1-10.
17. Riadi, I.; Sunardi; S.; and Widiandana, P. (2019). Mobile forensics for cyberbullying detection using term frequency - Inverse document frequency (TF-IDF). *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika* (*JITEKI*), 5(2), 68-76.
18. RSA. (2016). 2016: Current state of cybercrime. Retrieved March 21, 2022, from https://www.moonstone.co.za/upmedia/uploads/library/Moonstone%20Library/MS%20Industry%20News/2016-current-state-of-cybercrime.pdf.
19. Ruthven, I.; and Lalmas, M. (2003). A survey on the use of relevance feedback for information access systems. *The Knowledge Engineering Review*, 18(2), 95-145.
20. Sahu, S.K.; Sarangi, S.; and Jena, S.K. (2014). A detail analysis on intrusion detection datasets. *Advance Computing Conference* (*IACC*), 2014 *IEEE International*. Haryana, India, 1347-1350.
21. Selberg, E.W. (1997). Information retrieval advances using relevance feedback. Retrieved March 21, 2022, from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.207.6933&rep=rep1&type=pdf
22. Statista (2020). Number of smartphone users worldwide from 2016 to 2027. Retrieved March 21, 2022, from https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/.
23. Suryana; A.L.; Akbar, R.E.; and Widiyasono, N. (2016). Investigasi email spoofing dengan metode digital forensics research workshop (DFRWS) (Email spoofing investigation using digital forensics research workshop (DFRWS) method). *Jurnal Edukasi dan Penelitian Informatika* (*Jepin*) 2(2), 111-117.
24. Syahib; M.I.; Riadi, I.; and Umar, R. (2020). Akuisisi bukti digital aplikasi viber menggunakan metode national institute of standards technology (NIST) (Digital evidence acquisition of viber application using national institute of standards technology method). *Jurnal Sains Komputer & Informatika* (*J-Sakti*), 4(1), 170-178.
25. Tala, F.Z. (2003). *A study of stemming effects on information retrieval in Bahasa Indonesia*. The Netherlands: Universiteitvan Amsterdam.
26. Uden, M.V. (2011). Rocchio: Relevance feedback in learning classification algorithms. Retrieved March 21, 2022, from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.53.3960&rep=rep1&type=pdf

27. Umar, R.; Riadi, I.; and Muthohirin, B.F. (2019). Live forensics of tools on android devices for email forensics. *Telkomnika* (*Telecommunication Computing Electronics and Control*), 17(4), 1803-18099.
28. Weiss, S.M.; Indurkhya, N.; Zhang, T.; and Damerau, F.J. (2005). *Text mining: Predictive methods for analyzing unstructered information*. New York: Springer.
29. Riadi, I.; Sunardi; and Widiandana, P. (2020). Investigasi cyberbullying pada WhatsApp Menggunakan digital forensics research workshop (Investigating cyberbullying on WhatsApp using digital forensics research workshop). *Jurnal RESTI* (*Rekayasa Sistem Dan Teknologi Informasi*), 4(4), 730-735.
30. Zebua, F. (2017). DailySocial Report: Instant Messaging Survey 2017. Retrieved March 21, 2022, from: https://dailysocial.id/post/laporan-dailysocial- survey-instant-messaging-2017.