JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

SCIENTIFIC ARTICLE PLAGIARISM DETECTION USING THE QUADWORD APPROACH

By
Hari Purwanto[1], Betesda Sinaga[2], Iswandir[3], Mohamad Adila Rossa[4], Abdul Jamil[5]
[1,] Departement of Information System, Universitas Dirgantara Marsekal Suryadarma
[2,3] Department of Informatics Management, Universitas Dirgantara Marsekal Suryadarma
[4] Department of Management, Universitas Muhammadiyah Jakarta
[5] Department of Management, Universitas Muhammadiyah Brebes
Email: raldy08@gmail.com

ABSTRACT

Plagiarism is defined as the act of taking or presenting another person's work as one's own, regardless of intent. This infraction can be substantiated when supported by robust evidence. Academic works are often targeted by plagiarism, as the constrained writing proficiency among undergraduate students may foster such occurrences. A principal preventive endeavor to mitigate plagiarism within academic works involves the deployment of plagiarism detection applications. A primary preventive measure to mitigate plagiarism in academic works is the utilization of plagiarism detection software. Third-party applications typically require the purchase of a license to optimally access their full features. Cognizant of this limitation, the author developed an alternative application, although its capabilities are restricted to the detection of internal manuscripts. The application engineered herein implements the Rabin-Karp algorithm, leveraging a quadword approach. The concept of the quadword itself incorporates a phrase-based technique, functioning by generating a collection or set comprising four consecutive words within the source text

***Keywords** : Detection, Plagiarism, Rabin Karp Algorithm, Quadword*

DETEKSI PLAGIAT ARTIKEL ILMIAH MENGGUNAKAN PENDEKATAN QUADWORD

ABSTRAK

Plagiarisme didefinisikan sebagai tindakan mengambil atau menjiplak karya orang lain, baik secara sadar maupun tidak sadar, dan mengklaimnya sebagai karya pribadi. Tindakan ini dapat diketahui jika didukung oleh bukti yang kuat. Karya ilmiah sering menjadi target plagiarisme, di mana keterbatasan pengalaman menulis mahasiswa sarjana dapat mendorong terjadinya tindakan tersebut. Salah satu upaya preventif untuk menanggulangi plagiarisme pada karya ilmiah adalah melalui penggunaan aplikasi pendeteksi plagiarisme. Aplikasi pihak ketiga umumnya memerlukan pembelian lisensi untuk dapat mengakses seluruh fiturnya secara optimal. Menyadari hal tersebut, penulis mengembangkan sebuah

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

aplikasi alternatif, meskipun kemampuannya terbatas pada pendeteksian naskah internal. Aplikasi yang dirancang ini mengimplementasikan algoritma Rabin-Karp dengan memanfaatkan pendekatan quadword. Aplikasi ini dibuat dengan menggunakan algoritma Rabin Karp menggunakan pendekatan quadword. Konsep quadword itu sendiri meyerap teknik phrase-based yang cara kerjanya adalah membuat kumpulan atau himpunan yang terdiri dari empat kata yang berurutan pada teks sumber
***Kata kunci** : Aplikasi deteksi, Plagiarisme, Algoritma Rabin Karp, Quadword*

INTRODUCTION

Plagiarism refers to the taking of intellectual property, whether the work or ideas of others, and presenting it as one's own work without proper credit. This is an unacceptable act of indiscipline. Generally, many people simply copy all or part of the writing of another source without citing the original source. They often do not paraphrase, or use their own words, when incorporating material found on the internet into their assignments. Theses are often the target of plagiarism. A thesis is a typical Indonesian term that refers to a formal scientific paper, which presents the results of undergraduate research (S1) and serves as an in-depth analysis of a specific issue or phenomenon in a scientific discipline, compiled in accordance with applicable methodological rules.

Academic integrity is a fundamental pillar of the higher education and scientific research ecosystem. Plagiarism, broadly defined as the act of taking, copying, or presenting another individual's work, ideas, or concepts whether intentionally or unintentionally and claiming ownership without proper attribution of the source, is a serious ethical violation. The impact of this indiscipline not only harms the creator of the original work but also undermines the credibility of academic institutions and the validity of the knowledge produced. Formal academic papers, such as undergraduate theses, dissertations, and theses, are often the primary targets of plagiarism. Empirical observations indicate that limited writing experience and paraphrasing skills among undergraduate students can lead to verbatim duplication of content, either partially or completely, without proper attribution. This situation indicates the urgency of developing and implementing effective preventive measures and detection tools.

In practice, students utilize online plagiarism checking features to validate their theses. This online checking is typically done by comparing the thesis text against a database of

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

journals or published articles. If plagiarism verification across internal student theses is required within a single graduation period, the operator is required to use the Bulk Comparison feature available in the Plagiarism Checker X application. This feature's working mechanism adopts a 'one-to-many' approach, where one file is cross-validated against a set of other files. If a high plagiarism percentage is detected, the operator or administrator will contact the student concerned to provide an opportunity to paraphrase. This iterative workflow process can consume a significant amount of time. It would be more efficient if students independently check their theses for plagiarism and include evidence of plagiarism-free results. This check includes both internal comparisons (with theses of fellow students) and external comparisons (with journals or published articles). Due to this need, the author initiated the development of an alternative application that can help detect similarities in student theses.

One of the main preventative measures in combating plagiarism is through the adoption of plagiarism detection applications or software. While many commercial solutions are available online, these third-party applications generally require the purchase of a paid license to fully access all features. Crucial features like bulk comparison, which is necessary to cross-validate all student theses within a single graduating class, are often locked behind licensing restrictions. Manual or partial checking processes using these limited features have proven to be significantly time-consuming and inefficient.

The Rabin-Karp algorithm is an effective string matching method used to measure the degree of textual similarity (Mubarak, 2022). The fundamental procedure of this algorithm involves generating a set of n-grams from the source text, converting these n-grams to hash values, and then calculating the percentage of similarity using the Jaccard, Sorensen-Dice, or Andberg similarity coefficients. Recognizing these limitations in accessibility and time efficiency, this research focuses on developing an alternative application specifically designed for the internal context of institutions. This proposed application implements the Rabin-Karp algorithm, a string matching method known for its efficiency in measuring the degree of textual similarity. The innovation in this research lies in the use of a quadword approach (four consecutive words) for n-gram formation as the basis for hashing calculations, which is expected to improve the accuracy of detecting similarity in longer phrases compared to

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

conventional unigram or bigram methods.

LITERATURE REVIEW

1. Plagiarism

   Plagiarism, often referred to as plagiarism, is an unethical act that involves taking or copying the intellectual work—whether it be an essay, idea, concept, or creation—of another individual and then presenting it as if it were one's own original creation. This act has serious legal consequences and can be categorized as a criminal offense, especially because it essentially violates or steals the copyright (intellectual property rights) of the original creator.

   Specifically in the academic context, Regulation of the Minister of National Education of the Republic of Indonesia Number 17 of 2010 concerning the Prevention and Handling of Plagiarism in Higher Education provides a formal definition. Based on this regulation, plagiarism is defined as an act, whether done consciously (intentionally) or unconsciously (unintentionally), which aims to obtain a value or credit in a scientific work. The method used is by citing or using part or all of the scientific work of another party, but claiming it as one's own scientific work, without including accurate, precise, and adequate sources of information.

2. Plagiarism detection

   Plagiarism detection is the process of finding cases of plagiarism or copying (taking someone else's work, opinions or ideas and making them appear to be your own work without proper attribution) in a work or document.

3. Basic Concepts of Plagiarism Detection

   Essentially, plagiarism detection tools work by comparing uploaded text with billions of available sources. This process involves several key stages:

   a. Text Preprocessing : The text is cleaned and formatted to improve comparison accuracy. This includes removing punctuation, converting the text to lowercase, removing common words (stop words) such as "dan" or "yang," and stemming (reducing words to their base form, for example, "berjalan," "lari," "lelarian" becomes

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

"lari").

b. Comparison : The processed document is then compared with the reference source using various algorithms.

c. Similarity Report : The comparison results are displayed in a report highlighting sections of the text where similarities are detected, often accompanied by a plagiarism percentage.

4. Rabin-Karp Algorithm

The Rabin-Karp algorithm is an efficient string searching algorithm invented by Michael Rabin and Richard Karp. Its basic concept is to use a rolling hash function to quickly compare blocks of text. Instead of comparing characters individually, the Rabin-Karp algorithm works based on three key principles:

a. Using a Hash Function

Instead of comparing an entire substring of text against the search pattern (which would be time-consuming), Rabin-Karp computes a unique numeric value, called a hash value, for the pattern and for each substring window in the text.

b. Rolling Hash Function

This is the key to Rabin-Karp's efficiency. When the search window shifts one position forward in the text, the algorithm does not recalculate the hash value for the new window from scratch. Instead, it reuses the previous hash value.

The rolling hash formula allows for the removal of the contributions of characters leaving the window and the addition of the contributions of new characters entering the window, using only a few simple mathematical operations (subtraction, multiplication, and addition). All these operations are usually performed using the prime modulus (\(p\)) to avoid dealing with very large numbers.

c. Application of Number Theory

This algorithm utilizes number theory, such as the use of modulus primes, to minimize hash collisions, which occur when two different substrings produce the same hash value.

d. Rabin Karp Algorithm Simulation

Rabin Karp is a string matching algorithm that uses a hash function to compare the

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

search string with a substring in the text (ngram). This algorithm was developed by two researchers, Michael O. Rabin and Richard M. Karp, in 1987. The following are the steps or flowchart in implementing the Rabin Karp algorithm:

Text Preprocessing

This stage begins by removing symbols and spaces in the two sentences being tested, then converting all capital letters to lowercase.

Source	Before	after
Teks 1	designing a plagiarism detection application	designing a plagiarism detection application designing a plagiarism detection application

Formation of n-gram sequences

After the text pre-processing stage is complete, the next step is to form a series of n-grams. In this test, the parameter n = 5 is used.

Source	Before	after
Teks 1	designing a plagiarism detection application	["peran","eranc","ranca","ancan","ncang","canga","angan","ngana","ganap","anapl","napli","aplik","plika","likas","ikasi","kasid","aside","sidet","idete","detek","eteks","teksi","eksip","ksipl","sipla","iplag","plagi","lagia","agiar","giari","iaris","arism","risme"]

5. Message Digest 5 (MD5)

MD5 is a one-way hash function created by Ronald Rivest in 1991. The MD5 algorithm is an algorithm that uses input values of arbitrary length and produces a random output value of 32 characters [15]. For example, there is the text "file upload is starting", and if the text is processed using MD5, it will become “3802cc03253a58a6bd68bd50155427aa”.

6. The Quadword approach

The Quadword approach in the context of plagiarism detection is a specific implementation technique, not a stand-alone algorithm. The basic concept is to use "four words" as the basic unit of comparison or "fingerprint" in text matching algorithms, such as the Rabin-Karp Algorithm or Winnowing. This approach is rooted in phrase-based

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

techniques, or N-grams (in this case, N=4), which work by dividing a document into fixed-size chunks of text, each consisting of four words. Here are the theoretical details:

a. Quadword Unit Formation

The document to be checked first undergoes a preprocessing stage (such as removing punctuation, lowercase letters, and optional stop words). After that, the text is divided into "quadrants" (quadwords).

b. Use with Matching Algorithms

The quadword approach is most often used in conjunction with hashing algorithms such as the Rabin-Karp Algorithm.

c. Simulation

The quadword concept itself is derived from phrase-based techniques, which work by creating groups or sets of four consecutive words in the source text. Here is an example of the concept of a quadword:

Source text	rancang bangun aplikasi deteksi plagiat skripsi menggunakan algoritma rabin karp dengan pendekatan quadword

After quadword tokenization is performed :

[0] = ‘designing a plagiarism detection application’
[1] = ‘building a plagiarism detection application’
[2] = ‘thesis plagiarism detection application’
[3] = 'detecting thesis plagiarism using'
[4] = ‘thesis plagiarism using an algorithm’
[5] = ‘thesis using the Rabin algorithm’
[6] = ‘using the Rabin-Karp algorithm’
[7] = ‘Rabin-Karp algorithm with’
[8] = ‘Rabin-Karp with a quadword approach’
[9] = 'Karp with a quadword approach'

RESEARCH METHODS

1. Data Collection Method

Before creating a system, a clear formulation and planning are necessary to determine the system's objectives. To support this system, adequate computer system support is

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

required, both in terms of software and hardware.

2. Implementation of the Rabin Karp Algorithm with a Quadword Approach to the System

The Rabin-Karp algorithm has been adapted for plagiarism detection in undergraduate thesis applications using a quadword-based approach. This technique is effective for searching for substring (four-word) similarities in a data corpus. The algorithm's implementation is explained as follows:

a. Text Preprocessing

Starting with the text preprocessing stage, this stage removes breaklines, removes excess whitespace, and converts all text to lowercase in texts 1 and 2.

Tabel 1. Results ext Pre-Processing teks 1 dan 2

Source	Before	After
Teks 1	Designing a thesis plagiarism detection application using the Rabin-Karp algorithm with a quadword approach.	Designing a thesis plagiarism detection application using the Rabin-Karp algorithm with a quadword approach.
Teks 2	Design a face detection application using the Python programming language and MySQL.	Design a face detection application using Python and MySQL programming languages..

b. Quadword Sequence Formation

After the text pre-processing stage, we continue with the formation of a quadword sequence from the pre-processed texts 1 and 2.

Table 2. Results of quadword formation for texts 1 and 2

Source	Befor	After
Teks 1	Designing a thesis plagiarism detection application using the Rabin-Karp algorithm with a quadword approach	‘designing a plagiarism detection application, building a plagiarism detection application, thesis plagiarism detection application, 'detecting thesis plagiarism using', ‘thesis plagiarism using algorithms’, ‘thesis using Rabin algorithms’, ‘using Rabin-Karp algorithms’, ‘Rabin-Karp algorithms with’, ‘Rabin-Karp with an approach’, 'Karp with a

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

		quadword approach'
Teks 2	Design a face detection application using the Python programming language and MySQL.	‘designing a plagiarism detection application, building a plagiarism detection application, thesis plagiarism detection application, 'detecting thesis plagiarism using', ‘thesis plagiarism using algorithms’, ‘thesis using Rabin algorithms’, ‘using Rabin-Karp algorithms’, ‘Rabin-Karp algorithms with’, ‘Rabin-Karp with an approach’, 'Karp with a quadword approach'

c. Generating an MD5 Hash from Quadwords

After generating a quadword sequence from texts 1 and 2, the next step is to convert each quadword into MD5 format.

Table 3. MD5 Hash of a Quadword Sequence

Source	Before	After
Teks 1	'designing a detection application', ‘building a plagiarism detection application’, ‘thesis plagiarism detection application’, 'detecting thesis plagiarism using', ‘thesis plagiarism using algorithms’, ‘thesis using Rabin algorithms’, ‘using Rabin-Karp algorithms’, ‘Rabin-Karp algorithms with’, ‘Rabin-Karp with an approach’, 'Karp with a quadword approach'	‘dfe66ea2374886921cd17a5ce60da971', 'c086518f3723a59c62c897c5595f1111', 'd223abdbfb202938fae79114fc2ce173', 'c44589ba35da06e4c81bd9cc16ba379a', '861b82f5aefee356f98823442ad694de', '6b1514170a3ef417756733f12fd2ed77', '359414e48828c50043c78813064415d3', '127660ded44c76ca2bd1a4b577ea1021', 'dfe86174ea4c0db58e9140c9c985129b', 'ea0bd4ab24ae416116cc520f4300c724'
Teks 2	'designing a detection application', ‘building a plagiarism detection application’, ‘thesis plagiarism detection application’, 'detecting thesis plagiarism using', ‘thesis plagiarism using algorithms’, ‘thesis using Rabin algorithms’, ‘using Rabin-Karp algorithms’, ‘Rabin-Karp	‘dfe66ea2374886921cd17a5ce60da971', '2663dae2e89e5f2b6f8323b8b33704c2', 'f6b42fb73be48c980b1049ec12a4f04b', 'fe709e0fde495c8f1f35cd97c0dbb669', '5ab1eb98e5756f24135decc670c2d438', '870cf6c963e0d5183f199911397e93c0', '48432026181e9c591e650dfd12054748', '7d2748b63921f1398c4274176f8e1fa6

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

	algorithms with', 'Rabin-Karp with an approach', 'Karp with a quadword approach'	

d. Calculating the Percentage of Plagiarism

Calculating the percentage of plagiarism is necessary to determine how much of a student's thesis has been detected as plagiarized. The following formula is used to calculate the percentage of plagiarism between Texts 1 and 2.

Percentage of Plagiarism (%) =

Description:

A: Wordset of Text 1

B: Wordset of Text 2

|A n B| = Number of words in common between A and B

|A| = Number of words in Text 1

Example: After converting a quadword to an MD5 hash, Table 3.3 shows the calculation of the percentage of plagiarism from matching Text 1 to Text 2. Table 3.3 shows the MD5 hash results for Texts 1 and 2, which are identical, namely 'dfe66ea2374886921cd17a5ce60da971'. This identical MD5 hash, when converted to a quadword, produces the phrase 'rancangbangun aplikasi sinyal', which consists of four words. Then in table 10 in text 1, namely 'Design of a plagiarism detection application for a thesis using the Rabin Karp algorithm with a quadword approach', the number of words is 13. Then it is known that the calculation of the percentage of plagiarism in text 1 and text 2 is 4/13x100% = 30%

3. Hardware and Software Requirements

Hardware Requirements

The hardware requirements for developing the application are a laptop with the following specifications:

1. Processor: Core i3 6006u
2. RAM: 8 GB
3. Internal Memory: 480 GB SSD

4. Software Requirements

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

The software requirements for developing the application are:

1. Windows 10 operating system
2. Visual Studio Code
3. XAMPP
4. Web Browser (Chrome)

RESULTS AND DISCUSSION

1. Results of Thesis Plagiarism Checking by the System

   To determine the success of the application created, namely the thesis plagiarism detection application (Plagiarism Checker F), a trial or testing of the application was conducted. The testing was conducted to determine the accuracy of the output compared to the Plagiarism Checker X application. The Plagiarism Checker X application is an application used to detect plagiarism in theses. The testing was conducted using three scenarios: one-to-one comparisons of different files, one-to-one comparisons of the same file, and one-to-many comparisons of different files.

2. One-to-One Testing of Different Files

   This test used 10 student thesis files from 2017. Files were selected randomly. Three variables were used in this test: detected plagiarized words, the number of words in the main file, and the plagiarism percentage.

Table 4. Test results on 2 different theses

No	file Comparison	Plagiarism Checker F			Plagiarism Checker X		
		words plagiarized	Word Count	%	words plagiarized	Word count	%
1	file1 dan file2	124	14505	0.85%	158	14487	1.09%
2	file2 dan file1	127	9341	1.36%	161	9338	1.70%
3	file3 dan file1	148	8613	1.72%	178	8610	2.06%
4	file4 dan file1	179	14554	1.23%	187	14567	1.28%
5	file5 dan file1	104	7086	1.47%	137	7086	1.93%
6	file6 dan file1	191	16591	1.15%	267	16635	1.60%
7	file7 dan file1	238	17548	1.36%	274	17822	1.53%
8	file8 dan file1	256	14571	1.78%	275	14575	1.88%
9	file9 dan file1	170	8834	1.92%	224	8796	2.54%
10	file10 dan file1	172	8090	2.13%	232	8087	2.86%

Table 4 shows the difference between the output of the Plagiarism Checker X (PCX)

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

application and the results from the original application (PCF). The difference is presented in Table 5.

Table 5. Calculation of the difference in testing 2 different sources

No	file comparizon	Deviation (output PCX - PCF)		
		words plagiarized	word count	%
1	file1 dan file2	34	-18	0.24%
2	file2 dan file1	34	-3	0.34%
3	file3 dan file1	30	-3	0.34%
4	file4 dan file1	8	13	0.05%
5	file5 dan file1	33	0	0.46%
6	file6 dan file1	76	44	0.45%
7	file7 dan file1	36	274	0.17%
8	file8 dan file1	19	4	0.10%
9	file9 dan file1	54	-38	0.62%
10	file10 dan file1	60	-3	0.73%
	Average	36	6	0.31

The results obtained from the one-to-one testing of different files showed an average difference in plagiarism of 36 words, a total of 6 words, and a percentage of 0.31%.

3. One-to-One Testing of the Same File

A second test was conducted on two of the same student thesis files. The test used six randomly selected 2017 student thesis files. The results of the second test can be seen in Table 6.

Table 6. Test Results on the Same 2 Sources

No	Comparison file	Plagiarism Checker F			Plagiarism Checker X		
		words plagiarized	Word count kata	%	words plagiarized	Word count	%
1	file1 dan file1	13765	14505	94.90%	14487	14487	100%
2	file2 dan file2	8827	9341	94.50%	9338	9338	100%
3	file3 dan file3	8174	8631	94.90%	8610	8610	100%
4	file4 dan file4	13565	14554	93.20%	14567	14567	100%
5	file5 dan file5	6798	7086	95.94%	7086	7086	100%
6	file6 dan file6	15885	16591	95.74%	16635	16635	100%

In the second test, the difference results were obtained in table 7 with the same difference calculation formula as the first test, namely the output results from the Plagiarism Checker

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

X (PCX) application minus the results from the application created (PCF).

Table 7 Calculation of the difference in testing of 2 identical theses

No	File comparison	Deviation (output PCX - PCF)		
		words plagiarized	Count Word	%
1	file1 dan file1	722	-18	5.10%
2	file2 dan file2	511	-3	5.50%
3	file3 dan file3	436	-21	5.10%
4	file4 dan file4	1002	13	6.80%
5	file5 dan file5	288	0	4.06%
6	file6 dan file6	750	44	4.26%
	Rata - Rata	618	2	5

The results of this second test showed an average difference in plagiarized words of 618, a total word count of 2 words, and a percentage of 5%.

4. One-to-Many Different File Test

This test is the application's main feature: comparing one source file with other uploaded sources. This test used 10 source files compared to 30 source files already available in the system. The results of this third test can be seen in Table 8.

Table 8. Test results for one to many different files

No	File Comparison	Plagiarism Checker F			Plagiarism Checker X		
		words plagiarized	Word Count	%	words plagiarized	Word Count	%
1	file 1 dan 30 file lainnya	2374	14505	16.37%	-	14487	22%
2	file 2 dan 30 file lainnya	2334	9341	24.99%	-	9338	32%
3	file 3 dan 30 file lainnya	2192	8613	25.45%	-	8610	33%
4	file 4 dan 30 file lainnya	3128	14554	21.49%	-	14567	27%
5	file 5 dan 30 file lainnya	1535	7086	21.66%	-	7086	27%
6	file 6 dan 30 file lainnya	3505	16591	21.13%	-	16635	24%
7	file 7 dan 30 file lainnya	4438	17548	25.29%	-	17822	34%
8	file 8 dan 30 file lainnya	4096	14571	28.11%	-	14575	40%
9	file 9 dan 30 file lainnya	2128	8834	24.09%	-	8918	33%
10	file 10 dan 30 file lainnya	2408	8090	29.77%	-	8087	42%

In this one-to-many test, the Plagiarism Checker X application outputs plagiarized words because the application itself does not output plagiarized words. The results of this one-to-

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

many test reveal the difference in output between the two applications, as shown in Table 9.

Table 9. Difference between PCX and PCF outputs in the one-to-many test

No	File Conmparison	Deviation	
		Word Count	%
1	file 1 dan 30 file lainnya	-18	5.63%
2	file 2 dan 30 file lainnya	-3	7.01%
3	file 3 dan 30 file lainnya	-3	7.55%
4	file 4 dan 30 file lainnya	13	5.51%
5	file 5 dan 30 file lainnya	0	5.34%
6	file 6 dan 30 file lainnya	44	2.87%
7	file 7 dan 30 file lainnya	274	8.71%
8	file 8 dan 30 file lainnya	4	11.89%
9	file 9 dan 30 file lainnya	84	8.91%
10	file 10 dan 30 file lainnya	-3	12.23%
	Average	6	7.57%

The results of this second test showed an average number of words = 6 words, and a plagiarism percentage = 7.57%.

CONCLUSION

The Thesis Plagiarism Detection Application Using the Rabin-Karp Algorithm with the Quadword Approach can be concluded as follows: (a). The results of the one-to-one plagiarism check test on different files using the application produced good results due to the small difference in output values compared to the Plagiarism Checker X application. (b). The results of the one-to-one plagiarism check test on the same file produced less than satisfactory results because the output should have been 100%. (c). The thesis plagiarism detection application using the Rabin-Karp algorithm with the quadword approach can detect similarities in phrases between two different texts. (d). Detection Effectiveness: The quadword approach, which is generally combined with algorithms such as Rabin-Karp, is

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

effective in detecting textual similarities or "copy-paste" type plagiarism (verbatim plagiarism) in scientific papers. (e). Result Accuracy: The test results show that the application with this approach is able to provide an accurate percentage of plagiarism levels and has a small difference in output values compared to commercial plagiarism detection applications (e.g., Plagiarism Checker X). (f). Limitations: This approach has limitations in detecting plagiarism that uses paraphrasing techniques or word modification with synonyms, because small changes in the sequence or words can change the quadword hash value significantly.

Suggestion **:** The following are suggestions for developing the application of Plagiarism Detection for Scientific Works Using the Quadword Approach: (a). Developing a plagiarism detection method, not only through text but also able to detect through images contained in the document, (b). Ability to Process Long Texts: The system built with this approach is able to process long text documents, such as theses or final assignments, (c). Combination of Methods: It is recommended to combine the quadword approach with other methods or algorithms (for example, stemming, stopword removal, or more sophisticated Natural Language Processing (NLP) techniques) to improve detection accuracy, especially in dealing with paraphrasing and changes in sentence structure. (d). Feature Development: Further development can include adding features to detect more covert plagiarism of ideas, which are difficult to identify only based on the similarity of keywords or phrases. (e). Database Improvement: Enriching the database (corpus) of comparative documents will improve the system's ability to identify a wider range of plagiarized sources. (f). Application Accessibility: Developed applications should be made more accessible, for example by developing them for web-based or mobile platforms (Android and iOS), so that they can be used more widely by academics.

BIBLIOGRAPHY

Sari, Y., Khatimi, H., & Fajrin, R. A. (2014). Deteksi Plagiarisme menggunakan Algoritma Levenshtein Distance. Jurnal Teknologi Informasi Universitas Lambung Mangkurat, 6(1), 31-38. (Meskipun fokus pada Levenshtein Distance, artikel ini mengulas metode string matching secara umum).

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

Akbar, A., & Picard, M. (2019). Understanding plagiarism in Indonesia from the lens of Plagiarism Policy: Lessons For Universities. International Journal for Educational Integrity, 15(1), 1–17

Rustan. (2021). Uji Plagiarism pada Tugas Mahasiswa Menggunakan Algoritma Winnowing. Jurnal Applied Computer Science and Technology (JACOST), 2(2), 108–112. Artikel ini mendemonstrasikan penerapan praktis algoritma winnowing untuk menghitung nilai similaritas.

Mowbray, M. (2003). "Winnowing: local algorithms for document fingerprinting". Paper ini adalah sumber fundamental yang memperkenalkan algoritma winnowing (umumnya dirujuk dalam publikasi teknis, seperti yang disebutkan dalam konteks repositori UB).

Kaur, H., & Gupta, P. (2012). Plagiarism detection methods and tools: an overview. International Journal of Soft Computing and Engineering (IJSCE), 1(6), 218-223.

Bensalem, H., Jaafar, J. B., & Omri, M. N. (2025). Plagiarism types and detection methods: a systematic survey. Frontiers in Computer Science, 1504725. Tinjauan sistematis yang sangat baru dan komprehensif mengenai metode deteksi saat ini.

Foltynek, T., Rybicka, M., & Zelaszczyk, M. (2019). Academic Plagiarism Detection: A Systematic Literature Review. ACM Computing Surveys (CSUR), 52(2), 1-38.

M. S. Ramli, S. Cokrowibowo, and M. F. Rustan, “Uji Plagiarism pada Tugas Mahasiswa Menggunakan Algoritma Winnowing,” Journal of Applied Computer Science and Technology, vol. 2, no. 2, pp. 108–112, Dec. 2021, doi: 10.52158/jacost.v2i2.177.

D. G. Fatimah, “Ketakutan Akan Kegagalan Dan Intensi Plagiarisme Pada Mahasiswa,” Jurnal Psikologi Ulayat, vol. 5, no. 1, p. 45, Apr. 2018, doi: 10.24854/jpu12018-177.

A. Mubarak, “Implementasi Algoritma Rabin-Karp Untuk Pendeteksian Plagiarisme Pada File Dokumen Berupa Text Berbasis Web,” Journal of Information System Research, vol. 3, pp. 150–154, 2022

L. M. Febriansyah and S. E. Wahyuningrum, “Analysis Winnowing Algorithm For Text Plagiarism Detection Using Three Method Similarity,” 2019.

Y. Fadhillah, “114-218-1-SM-rabin karp,” Journal of Computer & Information Technology, vol. 2, pp. 1–20, 2021.

T. T. I. B. Billhaqqi, G. W. Wicaksono, and C. S. K. Aditya, “Analisis Perbandingan Algoritma Rabin-Karp Dan Winnowing Dalam Penilaian Jawaban Otomatis,” Seminar Nasional Teknologi dan Rekayasa, pp. 269–276, 2020.

T. Winarti, W. Setiawan, Iswoyo, and E. Pujiastuti, “Deteksi Kemiripan Dokumen Bahasa Indonesia Dengan Menggunakan Model Ruang Vektor,” Semarang, 2019.

I. Saputra and S. D. Nasution, “Perbandingan Performa Algoritma Md5 Dan Sha-256 Dalam Membangkitkan Identitas File,” Sains Komputer & Informatika (J-

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

SAKTI), vol. 6, no. 1, pp. 172–187, 2022.

Nst, V. F. H., Isnaini, D. B. J., Supriadi, S., Syafrizal, S., & Ichsan, R. N. (2025). Model Of Human Resource Collaboration Strategy In Strengthening Msme Halal Products In The Indonesian Nias Islands. *Jurnal Ilmiah METADATA*, *7*(3), 62-79.

Ichsan, R. N., Nst, V. F. H., Supriadi, S., Syafrizal, S., & Lubis, F. P. A. (2025). Sharia principles, digital transformation, and local economy: Challenges and opportunities for Sharia cooperatives in Langkat Regency. *Jurnal Ilmiah METADATA*, *7*(3), 30-41.

Ichsan, R. N., Siregar, B. A., Suma, D., Nst, V. F. H., & Lubis, F. P. A. (2025). Halal Industry In The Fulfillment Of Sharia Maqasid: A Qualitative Study On Halal Business Actors In North Sumatra. *Jurnal Ilmiah METADATA*, *7*(2), 80-97.

Wijaya, D. M., Nst, V. F. H., & Isnaini, D. B. Y. (2025). Designing A Talent Management Strategy To Address Organizational Transformation Challenges: A Case Study of PT. Sentosa Deli Mandiri. *Moneter: Jurnal Keuangan dan Perbankan*, *13*(1), 125-138.

Nst, V. F. H., Ichsan, R. N., Supriadi, S., & Lubis, F. P. A. (2025). Edukasi Konsep Pariwisata Ramah Muslim Bagi Pelaku Usaha Pariwisata Di Kabupaten Langkat, Sumatera Utara. *Jurnal Pengabdian Masyarakat Hablum Minannas*, *4*(1), 26-36.

Nst, V. F. H., Wijaya, D. M., Azaman, A., & Nasti, N. (2025). Sustainability Performance Management Integration: A Systemic Approach In Improving The Organizational Competitiveness Of PT. Sentosa Deli Mandiri. *Moneter: Jurnal Keuangan dan Perbankan*, *13*(1), 114-124.

Nst, V. F. H., Wijaya, D. M., & Azaman, A. (2025). Pengaruh Modal Intelektual Dan Komitmen Organisasional Terhadap Kinerja Pegawai Dengan Organizational Citizenship Behavior (Ocb) Sebagai Variabel Intervening Pada Pemerintahan Kota Medan. *Jurnal Ilmiah METADATA*, *7*(1), 1-15.

Nst, V. F. H., Asmuni, A., & Anggraini, T. (2024). Review Of Fiqh Muamalah On The Forms Of Online Buying And Selling Contracts In The Tiktok Shop Application. *Jurnal Review Pendidikan dan Pengajaran (JRPP)*, *7*(3), 10804-10812.

Ichsan, R. N., Nst, V. F. H., Nasution, L., & Hutabarat, L. (2024). The effect of halal labeling on the performance of small and medium enterprise (SME) in medan city. *Jurnal Mantik*, *8*(1), 421-427.

Lubis, M. R., Ichsan, R. N., Nasution, L., Nst, V. F. H., & Lubis, D. (2024). Analysis Of Factors Affecting The Amount Of People's Business Credit Loans In Lubuk Pakam District, Deli Serdang Regency, North Sumatra Province. *Jurnal Ekonomi*, *13*(02), 915-923.

Nst, V. F. H., Majid, M. S. A., & Harahap, I. (2024). The Role Of Imports In Development According To Islamic And Conventional Macroeconomic Perspectives. *Moneter: Jurnal Keuangan dan Perbankan*, *12*(1), 100-106.

Devi, R. S., Lubis, M. A., Nst, V. F. H., & Sihombing, A. (2024). Persaingan Usaha Tidak

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

Sehat Berdasarkan Undang-Undang Nomor 5 Tahun 1999 Tentang Larangan Praktek Monopoli Dan Persaingan Usaha Tidak Sehat. *Jurnal Ilmiah METADATA*, *6*(1), 108-118.

Nasution, L., Ichsan, R. N., Nst, V. F. H., & Rizkina, S. (2024). Pendampingan Akreditasi Institusi Perguruan Tinggi Di Akademi Keperawatan Hkbp Balige. *Pedamas (Pengabdian Kepada Masyarakat)*, *2*(01), 113-117.

Nst, V. F. H., Nasution, M. Y., & Sugianto, S. (2024). Relationship ushul Fiqh, Qowa'id Fiqih dan Maqashid Al-Syariah With Islamic Economy. *Jurnal Ilmiah Ekonomi Islam*, *10*(1), 1017-1023.

Nst, V. F. H., Tarigan, A. A., & Nasution, Y. S. J. (2023). Prinsip Equilibrium Perilaku Berkonsumsi Dalam Perspektif Al Qur'an Surat Al Furqon Ayat 67. *Management Studies and Entrepreneurship Journal (MSEJ)*, *4*(6), 10024-10034.

Lubis, M. R., Siregar, G. T., Nurita, C., Nst, V. F. H., & Lubis, D. (2023). Peningkatan Kesadaran Hukum Masyarakat: Memahami Perbedaan Tindak Pidana Penipuan dan Penggelapan. *Bulletin of Community Engagement*, *3*(2), 261-270.

Ichsan, R. N., Nst, V. F. H., Nasution, L., & Hutabarat, L. (2024). The effect of halal labeling on the performance of small and medium enterprise (SME) in medan city. *Jurnal Mantik*, *8*(1), 421-427.

Lubis, M. A., Siregar, G. T., Lubis, M. R., Nst, V. F. H., & Ichsan, R. N. (2023). Prosedur Jual Beli Tanah Dan Bangunan Warisan Yang Dilakukan Dihadapan Ppat (Procedure For Sale And Purchase Of Heritage Land And Buildings Carried Out Before The Ppat). *PKM Maju UDA*, *4*(3), 1-13.

Ichsan, R. N., Syahbudi, M., & Nst, V. F. H. (2023). Development of Islamic Human Resource Management in The Digital Era For MSMEs and Cooperatives in Indonesia. *IQTISHODUNA: Jurnal Ekonomi Islam*, *12*(2), 497-512.

Ichsan, R. N., Tanjung, A. M., & Nst, V. F. H. (2023). Pemanfaatan Website Online Single Submission (Oss) Dalam Kegiatan Usaha Mikro Kecil Menengah Dikota Medan Berbasis Maqashid Syariah. *Jurnal PKM Hablum Minannas*, *2*(2), 57-72.

Ichsan, R. N., Lubis, M. A., Nst, V. F. H., & Panggabean, N. R. (2023). Sosialisasi Peningkatan Usaha Mikro Kecil Dan Menengah Berbasis Manajemen Syariah Di Kecamatan Medan Area Kota Medan. *PKM Maju UDA*, *4*(2), 42-49.

Nst, V. F. H., Suma, D., Siregar, B. A., Ichsan, R. N., Panggabean, N. R., & Sibarani, J. P. (2023). Pendampingan Pemasaran Keripik Ubi Dalam Meningkatkan Penjualan Berbasis Digital Di Desa Marendal 1 Kecamatan Patumbak, Deli Serdang-Sumatera Utara. *Jurnal PKM Hablum Minannas*, *2*(1), 45-52.

Ammar, D., Danialsyah, D., Lubis, M. F. R., Purba, A. R., & Nst, V. F. H. (2023). Pelaksanaan Pemberian Marga Dalam Sistem Perkawinan Etnik Mandailing (Studi Di Lembaga Adat Budaya Mandailing Medan). *Jurnal PKM Hablum Minannas*, *2*(1), 68-79.

JURNAL ILMIAH METADATA, ISSN :2723 -7737, Vol. 8 No. 1
Edition January 2026, Published: 28-01-2026, Page 126-144
DOI: 10.47652/metadata.v8i1.936

Siregar, G., Lubis, M. A., Lubis, M. R., Nst, V. F. H., & Nasution, L. (2023). Perbuatan Melawan Hukum Akibat Membangun Di Atas Tanah Wakaf (Unlawful Actions Caused By Building On The Waqf Land). *PKM Maju UDA*, *4*(1), 31-38.

Nst, V. F. H., Nasution, Y. S. J., & Siregar, S. (2024). Implementation Of Wakaf As A Tool Of Social Finance To Achieve The Sdgs In Indonesia Case Study On Indonesian Waqf Board. *Moneter: Jurnal Keuangan Dan Perbankan*, *12*(3), 623-634.

Ichsan, R. N., Nst, V. F. H., Nasution, L., & Hutabarat, L. (2024). *Buku Pelatihan Dan Pengembangan SDM*. CV. Sentosa Deli Mandiri.

Ichsan, R. N., Nst, V. F. H., & Panggabean, N. R. (2024). *Buku Ajar Sistem Informasi Manajemen (SIM)*. CV. Sentosa Deli Mandiri.

Ichsan, R. N., Syahbudi, M., Barus, E. E., & Nst, V. F. H. (2024). The Role Of Islamic Banking Literacy And Ease Of Use On Achieving Sustainable Development Goals And Maqashid Al-Shariah In Indonesia. *International Journal Of Economics And Finance Studies*, *16*(2), 190-208.

Ichsan, R. N., Syahbudi, M., Barus, E. E., & Nst, V. F. H. (2024). The Role Of Islamic Banking Literacy And Ease Of Use On Achieving Sustainable Development Goals And Maqashid Al-Shariah In Indonesia. *International Journal Of Economics And Finance Studies*, *16*(2), 190-208.

Nst, V. F. H., Asmuni, A., & Anggraini, T. (2024). Review Of Fiqh Muamalah On The Forms Of Online Buying And Selling Contracts In The Tiktok Shop Application. *Jurnal Review Pendidikan Dan Pengajaran (JRPP)*, *7*(3), 10804-10812.

Ichsan, R. N., Nst, V. F. H., Nasution, L., & Hutabarat, L. (2024). The Effect Of Halal Labeling On The Performance Of Small And Medium Enterprise (Sme) In Medan City. *Jurnal Mantik*, *8*(1), 421-427.

Lubis, M. R., Ichsan, R. N., Nasution, L., Nst, V. F. H., & Lubis, D. (2024). Analysis Of Factors Affecting The Amount Of People's Business Credit Loans In Lubuk Pakam District, Deli Serdang Regency, North Sumatra Province. *Jurnal Ekonomi*, *13*(02), 915-923.

Nst, V. F. H., Majid, M. S. A., & Harahap, I. (2024). The Role Of Imports In Development According To Islamic And Conventional Macroeconomic Perspectives. *Moneter: Jurnal Keuangan Dan Perbankan*, *12*(1), 100-106.