IT Journal Research and Development (ITJRD)
Vol.7, No.1, August 2022, E-ISSN : 2528-4053 | P-ISSN : 2528-4061
DOI : 10.25299/itjrd.2022.7897

Implementation of the Winnowing Algorithm in Detecting Plagiarism in Title and Abstract of Student's Final Project

Dewi Nasien[1] and Eka Putra Yansen[2]
Department of Information Technology, Faculty of Computer Science, Institut Bisnis dan Teknologi Pelita Indonesia [1,2]
dewinasien@lecturer.pelitaindonesia.ac.id[1], eka.yansen@student.pelitaindonesia.ac.id[2]

Article Info

Article history:

Received Oct 21, 2021
Revised Mar 18, 2022
Accepted Aug 11, 2022

Keyword:

Plagiarism detection
Winnowing algorithm
Title and abstract
Final project

ABSTRACT

Title and research abstract in a university / college are things that are quite common. Every student who will do the final project will definitely encounter these two things. The number of titles and abstracts that have been scattered on the internet and collected at one university / college must have similarities. To detect the similarity of the title text and abstract, it is necessary to apply an algorithm to the system that can determine the similarity of the text, one of which is the winnowing algorithm. The winnowing algorithm is an algorithm that can be used to detect similarities in a document text, where the text will be converted into hash values which will then form a window based on the previous hash values and the similarity percentage will be searched. The evidence from this study shows that winnowing method is considered as a benchmark for the acceptance of the student's final project and reduce plagiarism activities from their title and abstract writing. With the creation of this system, universities / colleges can use this system as a benchmark for the acceptance of a title and research abstract at the university / college.

Corresponding Author:

Dewi Nasien
Department of Information Technology, Faculty of Computer Science
Institut Bisnis dan Teknologi Pelita Indonesia
Jl. Jend. Ahmad Yani, Pekanbaru, Indonesia
Email: dewinasien@lecturer.pelitaindonesia.ac.id

1. INTRODUCTION

Technology has become increasingly sophisticated in today's modern era, allowing most activity to be done quickly. Conversely, there are some people who still craving for maximum results without working hard for it. Simply by getting into the internet, all information is accessed quickly and efficiently [1]. Nonetheless, when surfing the internet to ascertain information, news content frequently looks almost similar or 100% similar. Information has been discovered easily through the internet, which is widely misused by numerous parties causing plagiarism [2].

The rapid development of technology also dramatically influences the level of plagiarism in education contexts [3]. For instance, in scientific research, when the researchers find writings that will be chosen as their related references, they tend to choose copypaste techniques to complete the tasks instantly [4]. The similarities are usually uncovered in final semester student research titles and abstracts. If undetected, plagiarism can facilitate students progressing through courses without achieving the desired learning goals, which affects both their ability to complete subsequent courses, program objectives or even final project [5].

Talking about the thesis, of course, it cannot be detached from plagiarism owing to it is a scientific writing from the students' original research as a requirement for obtaining a bachelor's degree [6]. This issue is encountered when searching for reference sources through internet media (Google Scholar, Scopus, and so forth) for students about to enter the final semester. Somehow, the lecturer's capacity to exercise control and selection is still restricted by needing to verify and find out with each lecturer's or supervisor's memory abilities, which may be limited, so that titles occasionally pass the observation, resulting in duplication [7]. It might lead the reader or their advisor to feel uninterested since they have read the title or abstract before. In resolving whether a thesis title is accepted or not, it is to prove or evaluate the title with a list of existing ones, but this will take time and considering there are many applicants for thesis titles [8].

Winnowing algorithm selection is assumed on [9], which states that the winnowing algorithm is one of the algorithms in the document fingerprinting method. The input of this algorithm is a text document that is administered in order to get an output hash value [10]. This method accurately identifies copy text, including small parts similar in a set of documents through the resulting fingerprint. Through fingerprint matching, similarity values will be collected between documents. Meanwhile, as states by [11] the Winnowing Algorithm is performed to detect the similarity of words/sentences (common subsequence) in two or more texts being assessed. [12] elaborate winnowing algorithm is classified as an extrinsic plagiarism detection method. This algorithm identifies parts of sentences or words that have similarities from the document tested for plagiarism with the source document [13, 17].

Derived from the explanation, the researchers explored how to build a website design to check the similarity of titles and abstracts and apply the winnowing method, which is expected to examine it on students' final assignments. A winnowing algorithm can be utilized to detect similarities in text and documents. As defined by [16], the title in final project is an initial description of the document content. It is the preliminary direction of the research content, where there is a resemblance to the final project with the content which suspected to be similar to the existing ones.

Finally, research by [6] entitled "Implementasi Pengecekan Plagiarisme Proposal Tugas Akhir Mahasiswa Teknik Informatika UPN Veteran Yogyakarta", they conducted the winnowing algorithm to check the similarity of the proposal document. The results explicated that the smaller the N-gram selection, the higher the percentage of similarity values (76% and 89%). This happens because, with fewer N-grams, the string is cut smaller, the probability of finding the same character set is more outstanding. The larger the N-gram, the more characters it contains compared to the smaller N-gram, affecting the character set to decrease so that the similarity values (up to 57% and 68%).

From the explanation above, two questions are needed to be adjust for this study; 1) How to design a website to examine the similarity of titles and abstracts? 2) How the winnowing method thoroughly check the similarity of titles and abstracts in the final project? This study proposes creating a website design to verify the similarity of titles and abstracts, then using the winnowing method to accomplish the similarity in the final project. It is expected to contribute to presenting information related to the similarity of the title and abstract in the final project as a benchmark for accepting it by the campus. Furthermore, with this application program, it is hoped that the plagiarism activities for titles and abstracts of students' final assignments can be minimized. It also provides insights and great opportunities for other researchers to fill the gaps in this study as future research.

2. RELATED LITERATURE

In this study, many were inspired by previous research related to resolving similarity of text or documents by using one of the methods such as the winnowing method, which can assist authors in completing this research. According to [2], the Final Project is a document that represents research conducted by undergraduate students. In producing a quality Final Project, competent research is required; one of the factors is originality. The ability of students to create original research is an essential factor. Plagiarized text is the most common form of plagiarism. Detection of plagiarism is divided into two based on the task, i.e., intrinsic and extrinsic. The winnowing algorithm is an

extrinsic plagiarism detection method; it extends the Rabin-Karp Fingerprint method by adding a window feature to optimize the detection results. This study applies a Winnowing algorithm to detect text-similarity in the final project document of Dian Nuswantoro University students.

In [11] is also conducted study entitled "Implementasi deteksi plagiarisme menggunakan metode N-gram dan Jaccard similarity terhadap algoritma Winnowing". The method is N-gram, Jaccard similarity and winnowing. The results exhibited that the right N-gram is necessitated. The degree of similarity between the two documents/samples produced similarity values of different values if the N-gram and W-gram values entered are of different values. If the value entered is small, the degree of similarity outcome has a high value. If the N-gram and W-gram values entered are greater, the similarity value or the degree of similarity between documents had low results.

As stated by [18], from research, about 89% of students agree and understand the meaning of plagiarism in education and suggest that the material about it is explained at the beginning of the lecture. Nevertheless, about 65% admitted that they felt perplexed about the meaning of plagiarism, 59% stated that they were not given enough tutorials to avoid plagiarism in completing their assignments. *Winnowing algorithm* is an algorithm used for plagiarism detection. The input of this algorithm is a text document that is processed to produce output in a collection of hash values. The hash value is a numeric value formed from the ASCII calculation of each character. These sets of hash values are, from now on, referred to as fingerprints. Base value, gram, and window length are very influential in calculating the similarity of document similarity. If the greater the gram, the smaller the document similarity.

According to [14], the process of managing existing thesis titles and distributing information to students, and determining whether or not the titles submitted by students are accepted. It is still done manually, namely by checking one by one, and it is less effective. The winnowing algorithm's detection system makes it easier for the final project coordinator or the Head of the Study Program to delimit the percentage of similarity with an existing title. The system will ask for input in a title that will be checked for similarities and display the results to the user. Of the 117 existing thesis titles, 11 titles are similar to the submitted ones, with a more remarkable similarity equal 20 percent. With this system, it is expected that these processes will become more accessible, faster, and more effective.

According to [15], one of the challenges in the academic field is preventing the proliferation of plagiarism activities. One way that can be done is to detect plagiarism early on student work, especially thesis. The application of detection of plagiarism indications using the n-grams and winnowing method is the goal of this study and finding an effective nominal n. The words in the thesis document are represented in hash form. Then a selection is made using a winnowing algorithm to determine the document's fingerprint to be stored in the database. The test was performed using a sample of student thesis documents. The final result shows that the system can detect plagiarism according to word similarity consistently based on several run test scenarios. The n is the most effective in word similarity detection, namely n=7, with a percentage of 3.07% based on the difference between testing using the system and testing manually.

The subsequent research came from [14], which explored "Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing." Determining whether or not to accept the titles submitted by students is performed manually, namely by checking one by one, time-consuming and is less effective. The detection system by the winnowing algorithm is designed to facilitate the final project coordinator or the Head of Study Program in establishing the percentage of similarity with an existing title. The system probed for input in a title will be tested for similarities and display the results to the user. Of the 117 existing thesis titles, 11 titles are the same as the submitted ones, with a greater degree of similarity equal to 20 per cent.

3. METHODOLOGY

A step is needed to get cluster data with the optimal K number, as illustrated in Figure 1.

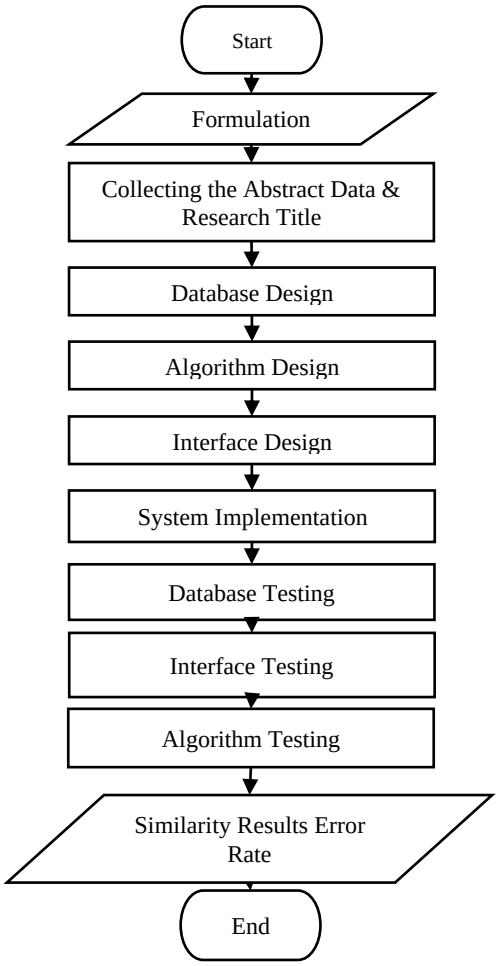


Figure 1. Plagiarism Detection System Flowchart

3.1. Formulation

How to implement the winnowing algorithm in the abstract similarity detection system and research titles on students' final assignments to be able to provide a presentation of similarities using PHP and Mysql languages and become a benchmark for the acceptance of an abstract or research title. At this stage, checking the sample titles and abstracts that have been downloaded from Google Scholar. Some several titles and abstracts are similar to each other, only in different ways of writing. The results of the similarity level calculation can be acknowledged as a benchmark for the acceptance of a title or research abstract at other universities.

3.2. Data Collection Abstract and Research Title

In this study, the data used was dummy data from the Google Scholar journal. The data is applied as a reference for conducting experiments on the system that will be created later as presented in Table 1.

Table 1. The Dummy Data

No	Researchers' Name	Research Title
1	(Tjiawi *et al*., 2018)	Perancangan Aplikasi Pendeteksi Tingkat Kesamaan Antar Dokumen dengan Algoritma Winnowing

2	(Wulan *et al*., 2018)	Deteksi Plagiasi Dokumen Skripsi Mahasiswa Menggunakan Metode N-Grams dan Winnowing
3	(Nurdin & Amin, 2017)	Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing
4	(Alamsyah, 2017)	Perbandingan Algoritma Winnowing dengan Algoritma Rabin Karp Untuk Mendeteksi Plagiarisme Pada Kemiripan Teks Judul Skripsi Fakultas
5	(Ulfa *et al*., 2016)	Pendeteksian Tingkat Similaritas Dokumen Berbasis Web Menggunakan Algoritma Winnowing

3.3 Database Design

At this stage, the database design that will be employed for the similarity detection system in titles and abstracts will be done. The database design contains a login table, abstract fingerprint table, title fingerprint table, research abstract table, research title table, category table and basic word table. Each table will have an input form.

3.4. Algorithm Design

For the algorithm design, Winnowing is a fingerprint document algorithm to perform the process of checking word similarity, which produces output in the form of a collection of hash values formed from the ASCII calculation of each character. This collection of hashes is called a fingerprint used to check for plagiarism. The steps of the winnowing algorithm are: removing irrelevant characters, forming an n-gram series, calculating the hash of each gram (rolling hash), forming a window, selecting the smallest fingerprint hash from each window, and calculating the equation with the Jaccard Coefficient.

3.5 Interface Design

In interface design, it consists of login menu design, register menu, main menu, and research menu list. This design is the initial stage for users before accessing or using the application.

3.5.1 Register Menu Design

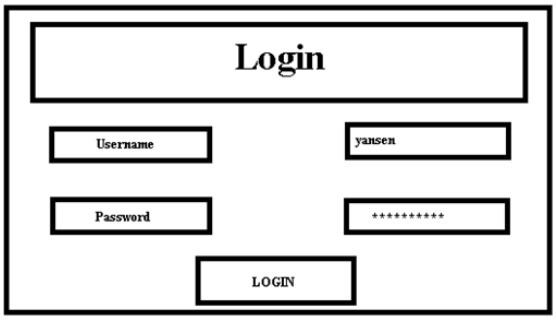


Figure 2. Login Menu

Figure 2 demonstrated that a login page design will be found when the user accesses the system first. Here, the user is required to enter an email and password to log into the system.

3.5.2 Register Menu Design

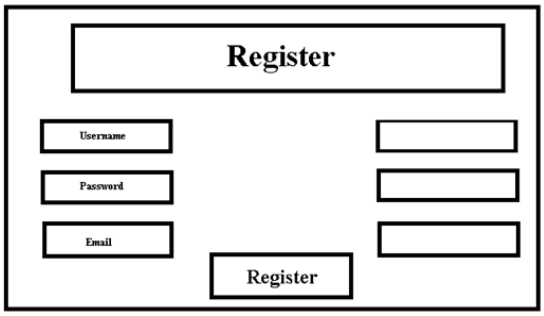


Figure 3. Register menu

The register menu design is presented in Figure 3, where users can register to enter the system. Users are asked to fill in their username, email, password, and confirm password.

3.5.3 Register Menu Design

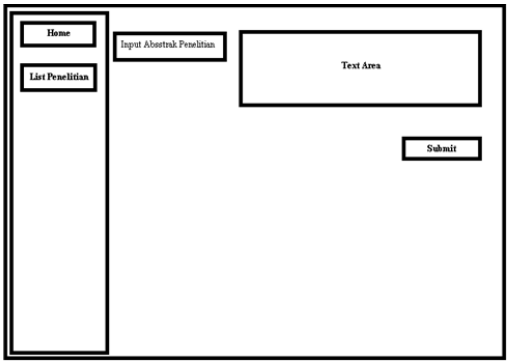


Figure 4. Main Menu

There is a design for the main menu when the user has logged in, visible in Figure 4. Other sub-menus include the category menu, research title menu, and abstract research menu.

3.5.4 Register Menu Design

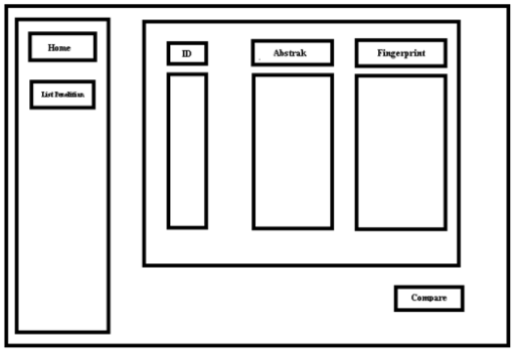


Figure 5. Research List Menu

Figure 5 describes the text results from the title and abstract that have been converted into fingerprints and are ready to be juxtaposed.

3.5 System Implementation

Calculating the similarity with the winnowing algorithm begins with removing punctuation marks (noise and whitespace) which is then performed by a series of grams according to the predetermined K value. Each series will be in the form of a hash value, which its functions will then process to generate a window along with the K value. After the window is formed, start the

fingerprint process to choose the smallest hash value from each window to be used as a fingerprint. The final step is to process the fingerprint using the Jaccard coefficient equation.

3.5.1 Database Testing

For instance, researchers took two article titles entitled "Sistem Deteksi Kemiripan Pada Judul dan Abstrak Tugas Akhir Mahasiswa Menggunakan Algoritma *Winnowing*" and "Sistem Deteksi Plagiarisme Pada Judul dan Abstrak Tugas Akhir Mahasiswa Menggunakan Algoritma *Rabin Karp".* Performing similarity calculations takes at least two texts to be compared with each other. After getting the fingerprint hash value from the window, the next step is to calculate the similarity, error rate, and sensitivity values.

3.5.2 Interface Testing

This section of the research also included user interface design. The system is examined to see whether the method was appropriately built or whether the system functioned properly. The interface testing would be explained in the next session.

3.5.3 Algorithm Testing

1. Undertaking stemming such as removing spaces between sentences, changing capital letters to normal, removing punctuation marks, and finding the sentence core.

   Test Text : sistemdeteksimiripjudulabstraktugasmahasiswaalgoritmawinnowing

   Compared Text : sistemdeteksiplagiarismejudulabstraktugasmahasiswaalgoritmarabinkarp

1. K-gram circuit with K= 3.

   Test Text:

   "sis","ist","ste","tem","emd","mde","det","ete","tek","eks","ksi","sim","imi","mir","iri","rip","ipj","pju","jud","udu","dul","ula","lab","abs","bst","str","tra","rak","akt","ktu","tug","uga","gas","asm","sma","mah","aha","has","asi","sis","isw","swa","waa","aal","alg","lgo","gor","ori","rit","itm","tma","maw","awi","win","inn","nno","now","owi","win","ing"

   Compared Text:

   "sis","ist","ste","tem","emd","mde","det","ete","tek","eks","ksi","sip","ipl","pla","lag","agi","gia","iar","ari","ris","ism","sme","mej","eju","jud","udu","dul","ula","lab","abs","bst","str","tra","rak","akt","ktu","tug","uga","gas","asm","sma","mah","aha","has","asi","sis","isw","swa","waa","aal","alg","lgo","gor","ori","rit","itm","tma","mar","ara","rab","abi","bin","ink","nka","kar","arp"

2. Determine the hash value by entering a series of grams into the hash function.

   Test Text:

   10375,9656,10460,10414,9262,9830,9125,9326,10412,9259,9807,10369,9591,9888,9636,10291,9619,10143,9739,10494,9261,10546,9719,8854,9089,10473,10519,10214,8936,9828,10552,10501,9331,9001,10393,9806,8890,9412,8997,10375,9659,10483,10609,8838,8932,9786,9456,10122,10295,9658,10474,9821,9033,10694,9605,10011,10028,10167,10694,9598

Compared Text :

10375,9656,10460,10414,9262,9830,9125,9326,10412,9259,9807,10372,9621,10141,9724,8889,9385,9492,8988,10294,9649,10397,9844,9252,9739,10494,9261,10546,9719,8854,9089,10473,10519,10214,8936,9828,10552,10501,9331,9001,10393,9806,8890,9412,8997,10375,9659,10483,10609,8838,8932,9786,9456,10122,10295,9658,10474,9816,8980,10205,8844,8993,9602,9970,9654,8995

3. Create windows along with the k=3

Test Text:

[10375,9656,10460],[9656,10460,10414],[10460,10414,9262],[10414,9262,9830],[9262,9830,9125],[9830,9125,9326],[9125,9326,10412],[9326,10412,9259],[10412,9259,9807],[9259,9807,10369],[9807,10369,9591],[10369,9591,9888],[9591,9888,9636],[9888,9636,10291],[9636,10291,9619],[10291,9619,10143],[9619,10143,9739],[10143,9739,10494],[9739,10494,9261],[10494,9261,10546],[9261,10546,9719],[10546,9719,8854],[9719,8854,9089],[8854,9089,10473],[9089,10473,10519],[10473,10519,10214],[10519,10214,8936],[10214,8936,9828],[8936,9828,10552],[9828,10552,10501],[10552,10501,9331],[10501,9331,9001],[9331,9001,10393],[9001,10393,9806],[10393,9806,8890],[9806,8890,9412],[8890,9412,8997],[9412,8997,10375],[8997,10375,9659],[10375,9659,10483],[9659,10483,10609],[10483,10609,8838],[10609,8838,8932],[8838,8932,9786],[8932,9786,9456],[9786,9456,10122],[9456,10122,10295],[10122,10295,9658],[10295,9658,10474],[9658,10474,9821],[10474,9821,9033],[9821,9033,10694],[9033,10694,9605],[10694,9605,10011],[9605,10011,10028],[10011,10028,10167],[10028,10167,10694],[10167,10694,9598]

Compared text:

[10375,9656,10460],[9656,10460,10414],[10460,10414,9262],[10414,9262,9830],[9262,9830,9125],[9830,9125,9326],[9125,9326,10412],[9326,10412,9259],[10412,9259,9807],[9259,9807,10372],[9807,10372,9621],[10372,9621,10141],[9621,10141,9724],[10141,9724,8889],[9724,8889,9385],[8889,9385,9492],[9385,9492,8988],[9492,8988,10294],[8988,10294,9649],[10294,9649,10397],[9649,10397,9844],[10397,9844,9252],[9844,9252,9739],[9252,9739,10494],[9739,10494,9261],[10494,9261,10546],[9261,10546,9719],[10546,9719,8854],[9719,8854,9089],[8854,9089,10473],[9089,10473,10519],[10473,10519,10214],[10519,10214,8936],[10214,8936,9828],[8936,9828,10552],[9828,10552,10501],[10552,10501,9331],[10501,9331,9001],[9331,9001,10393],[9001,10393,9806],[10393,9806,8890],[9806,8890,9412],[8890,9412,8997],[9412,8997,10375],[8997,10375,9659],[10375,9659,10483],[9659,10483,10609],[10483,10609,8838],[10609,8838,8932],[8838,8932,9786],[8932,9786,9456],[9786,9456,10122],[9456,10122,10295],[10122,10295,9658],[10295,9658,104

74],[9658,10474,9816],[10474,9816,8980],[9816,8980,10205],[8980,10205,8844],[10205, 8844,8993],[8844,8993,9602],[8993,9602,9970],[9602,9970,9654],[9970,9654,8995]

4. Selecting the fingerprint hash value from the window

   Test Text:

   9656,9262,9125,9259,9591,9636,9619,9739,9261,8854,9089,10214,8936,9828,9331,9001, 8890,8997,9659,8838,8932,9456,9658,9033,9605,10011,10028,9598

   Compared text:

   9656,9262,9125,9259,9621,8889,8988,9649,9252,9261,8854,9089,10214,8936,9828,9331, 9001,8890,8997,9659,8838,8932,9456,9658,8980,8844,8993,9602,8995

5. Calculating similarity, error rate, and sensitivity.

   Number of slices (title1, title2): 19

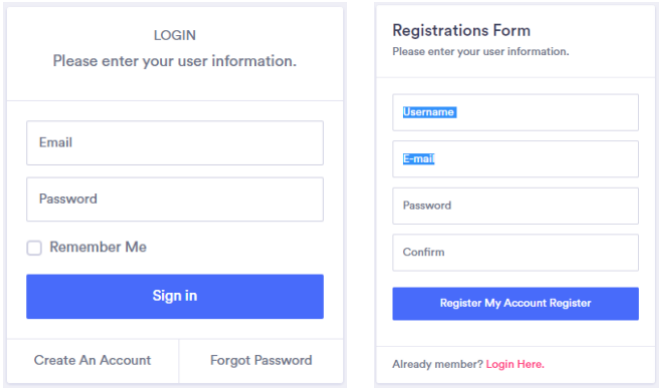
   Number of combined (head11, title2): 38

   Similartias (slice/merge): $\frac{19}{38} x100\% = 50\%$

4. RESULT AND DISCUSSION

4.1 Program Module Design

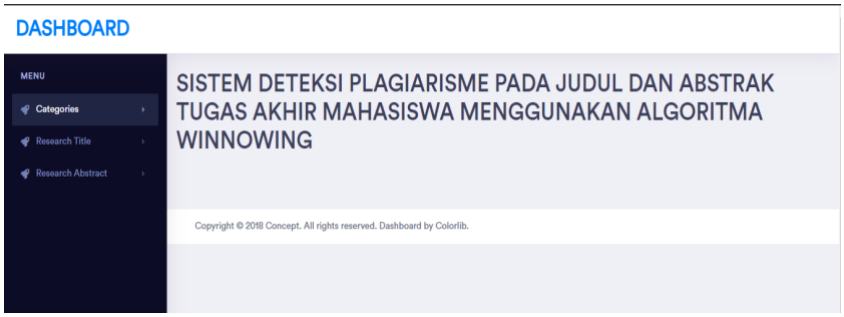
The form and appearance of the program module design can be seen below.



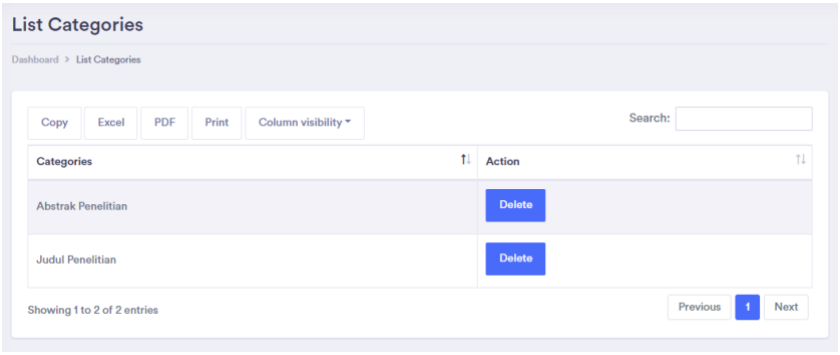
(a) (b)

Figure 6. Program Module Design (a) User Login Form (b) Register Form

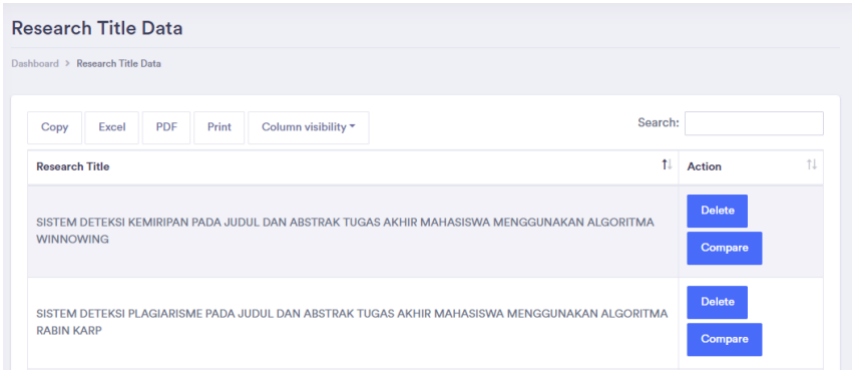
The user must register an e-mail and password in the login form to enter the system, as given in Figure 6 part (a). While in part (b), the registration form allows users to register their accounts to access the system.



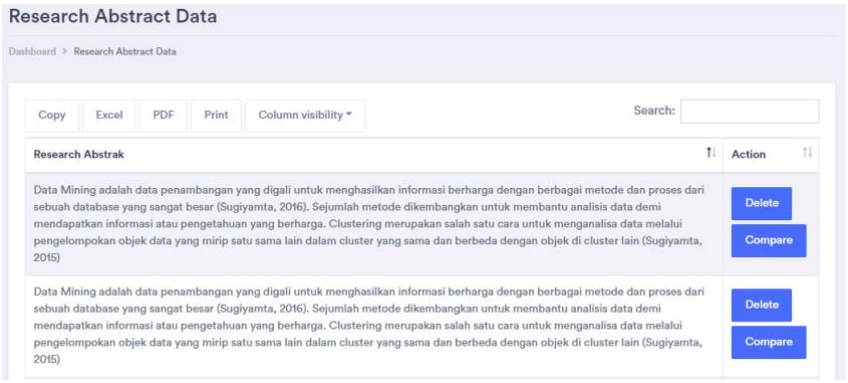
(a)


List Categories
Dashboard > List Categories
Copy
Excel
PDF
Print
Column visibility
Search:
Categories
Action
Abstrak Penelitian
Delete
Judul Penelitian
Delete
Showing 1 to 2 of 2 entries
Previous
1
Next

(b)


Research Title Data
Dashboard > Research Title Data
Copy
Excel
PDF
Print
Column visibility
Search:
Research Title
Action
SISTEM DETEKSI KEMIRIPAN PADA JUDUL DAN ABSTRAK TUGAS AKHIR MAHASISWA MENGGUNAKAN ALGORITMA WINNOWING
Delete
Compare
SISTEM DETEKSI PLAGIARISME PADA JUDUL DAN ABSTRAK TUGAS AKHIR MAHASISWA MENGGUNAKAN ALGORITMA RABIN KARP
Delete
Compare

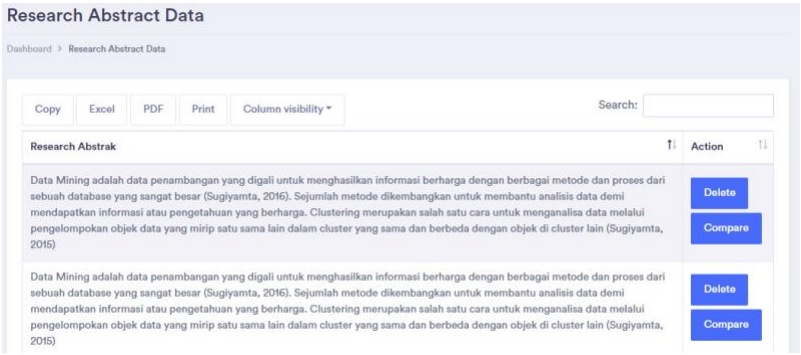
(c)


Research Abstract Data
Dashboard > Research Abstract Data
Copy
Excel
PDF
Print
Column visibility
Search:
Research Abstrak
Action
Data Mining adalah data penambangan yang digali untuk menghasilkan informasi berharga dengan berbagai metode dan proses dari sebuah database yang sangat besar (Sugiyamta, 2016). Sejumlah metode dikembangkan untuk membantu analisis data demi mendapatkan informasi atau pengetahuan yang berharga. Clustering merupakan salah satu cara untuk menganalisa data melalui pengelompokan objek data yang mirip satu sama lain dalam cluster yang sama dan berbeda dengan objek di cluster lain (Sugiyamta, 2015)
Delete
Compare
Data Mining adalah data penambangan yang digali untuk menghasilkan informasi berharga dengan berbagai metode dan proses dari sebuah database yang sangat besar (Sugiyamta, 2016). Sejumlah metode dikembangkan untuk membantu analisis data demi mendapatkan informasi atau pengetahuan yang berharga. Clustering merupakan salah satu cara untuk menganalisa data melalui pengelompokan objek data yang mirip satu sama lain dalam cluster yang sama dan berbeda dengan objek di cluster lain (Sugiyamta, 2015)
Delete
Compare

(d)


Compare Research Title
Dashboard > Compare Research Title
Copy
Excel
PDF
Print
Column visibility
Search:
Original Source
Compared to Original Source
Accuracy in Terms of Similarity(%)
Error Rate (%)
Sensitivity Rate (%)
JUDUL SISTEM DETEKSI PLAGIARISME PADA JUDUL DAN ABSTRAK TUGAS AKHIR MAHASISWA MENGGUNAKAN ALGORITMA RABIN KARP
SISTEM DETEKSI PLAGIARISME PADA JUDUL DAN ABSTRAK TUGAS AKHIR MAHASISWA MENGGUNAKAN ALGORITMA WINNOWING
70.59
29.41
0.71
JUDUL SISTEM DETEKSI PLAGIARISME PADA JUDUL DAN ABSTRAK TUGAS AKHIR MAHASISWA MENGGUNAKAN ALGORITMA RABIN KARP
SISTEM DETEKSI KEMIRIPAN PADA JUDUL DAN ABSTRAK TUGAS AKHIR MAHASISWA MENGGUNAKAN ALGORITMA WINNOWING
50
50
0.5

(e)



(f)

Figure 7. Program Module Design (a) Main menu (b) Category list menu (c) Research title menu list (d) Research abstract menu list (e) Compare title menu (f) Compare abstract menu

It appears from Figure 7, part (a) is the main menu form which will emerge when the user successfully logs into the system for the first time. From the figure (b), the category list menu is a display that will present a list of categories that have been put into the database. In this menu, there are other sub-menus that the user can access later. Part (c) is a research title list where the user opens the research title list menu. This menu will expose all research titles that have been input into the database. Additionally, when they want to proceed with the abstract list form, it is in part (d) where the user may see the abstract data. Users may compare one title to another, as seen in figure (e), where the title will be processed first with the Winnowing algorithm and Jaccard coefficient before being exhibited. Users may oppose abstracts with each other as well. Before the abstract is showed, it will be processed first by the Winnowing algorithm and Jaccard coefficients, as shown in part (f).

4.2 System Implementation and Maintenance Plan

After designing the program, the following step is to develop the implementation and maintenance of the system. At this implementation stage, software in Windows 10 as an operating system, Laravel Programming Language (PHP), a database with XAMPP, Visual Studio Code as a text editor and Google Chrome as a browser to run applications. The hardware operated is a Toshiba laptop with 6GB RAM specifications, 250GB SSD, Intel Core i3 3100 processor. In addition to the implementation stage, a system maintenance stage is required, carried out if an error or damage occurs to the system has been made.

The result of this study is coherent with several previous research related to the implementation of winnowing algorithm in detecting plagiarism. The research by [14] who determined the detection system by the winnowing algorithm. It is designed to facilitate the final project coordinator or the Head of Study Program in establishing the percentage of similarity with an existing title. The system probed for input in a title will be tested for similarities and display the results to the user. Of the 117 existing thesis titles, 11 titles are the same as the submitted ones, with a greater degree of similarity equal to 20 per cent. The previous research from [15] has discovered the effective nominal n. The words in the thesis document are represented in hash form. Subsequently, a selection is executed by a winnowing algorithm to determine the document's fingerprint to be stored in the database. The final result of the research indicated that the system could detect plagiarism conforming to word similarity consistently. As attested by various run test scenarios, nominal n is the most effective in word similarity detection, namely n = 7 with a percentage of 3.07% pertaining to the difference between testing with the system and testing manually. [11] is also conducted study in detecting plagiarism by the N-gram method and Jaccard similarity to the Winnowing algorithm. The results exhibited that the right N-gram is necessitated. When the N-gram and W-gram values were entered, the similarity value or degree of similarity between documents was low.

4. CONCLUSION

These findings enhance our understanding of applying the winnowing method can assist in locating similar results from the title and abstract of the student's final project. The evidence from this study indicates that winnowing method is considered as a benchmark for the acceptance of the student's final project and reduce plagiarism activities against their title and abstract writing.

Regarding previous research, the system created may produce a percentage of the title and abstract research similarity in this study. After observations have been made, there has been no research that simultaneously offers a similar value to the title and research. Considerably more work will need to be done to determine developers in establishing systems with a more attractive appearance and indulgent users in terms of platforms such as mobile.

REFERENCES

[1] R. Purnamasari, M. Fairuzabadi, and A. Riyadi. "Sistem Pengecekan Plagiasi Judul Tugas Akhir Menggunakan Algoritma Winnowing di Fakultas Sains dan Teknologi Universitas PGRI Yogyakarta." In *Seri Prosiding Seminar Nasional Dinamika Informatika*, vol. 5, no. 1, 2021.

[2] R. K. Wibowo, and K. Hastuti. "Penerapan Algoritma Winnowing Untuk Mendeteksi Kemiripan Teks Pada Tugas Akhir Mahasiswa." *Techno. Com*, vol. 15, no. 4, pp. 303-311, 2016.

[3] M. N. Khidfi, and J. Y. Sari. "Rancang bangun aplikasi pendeteksian kesamaan pada dokumen teks menggunakan algoritma enhanced confix stripping dan algoritma winnowing." *no. September* 2018.

[4] A. H. Purba and Z. Situmorang. "Analisis Perbandingan Algoritma Rabin-Karp Dan Levenshtein Distance Dalam Menghitung Kemiripan Teks." *Jurnal Teknik Informatika UNIKA Santo Thomas*, vol. 2, no. 2, pp. 24-32. 2017.

[5] J. Pierce and C. Zilles. "Investigating student plagiarism patterns and correlations to grades." In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, pp. 471-476. 2017.

[6] A. H. Pratomo, and A. P. Suryotomo. "Implementasi pengecekan plagiarisme proposal tugas akhir mahasiswa teknik informatika UPN Veteran Yogyakarta." In *Seminar Nasional Informatika (SEMNASIF)*, vol. 1, no. 1, pp. 221-229, 2020.

[7] N.I. Kurniati, A. Rahmatulloh and R. N. Qomar. "Web scraping and winnowing algorithms for plagiarism detection of final project titles." *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, *10*(2), pp.73-83. 2019.

[8] N. Alamsyah. "Deteksi Plagiarisme Tingkat Kemiripan Judul Skripsi dengan Algoritma Winnowing." *Technologia: Jurnal Ilmiah.* Vol. 8, no. 4, pp. 205-13, Oct, 2017.

[9] N. F. Ulfa, M. Mustikasari. "Pembuatan Aplikasi Pengukuran Tingkat Kemiripandokumen Berbasis Web Menggunakan Algoritma Winnowing." *Jurnal Ilmiah Informatika Komputer*. Vol 21, no. 2, Apr, 2017.

[10] A. P. Tjiawi, D. E. Herwindiati, and L. Hiryanto. "Perancangan aplikasi pendeteksi tingkat kesamaan antar dokumen dengan algoritma winnowing." *Computatio: Journal of Computer Science and Information Systems*, vol. 2, no. 1, pp. 36-44, 2017.

[11] S. Sunardi, A. Yudhana, and I. A. Mukaromah. "Implementasi Deteksi Plagiarisme Menggunakan Metode N-Gram Dan Jaccard Similarity Terhadap Algoritma Winnowing." 2018.

[12] A. Setiawan. "Implementasi algoritma winnowing untuk deteksi kemiripan judul skripsi studi kasus stmik budidarma." *Informasi dan Teknologi Ilmiah (INTI)*. vol 4, no. 2, 2017.

[13] M. Maskur, D. Q. Putra, and N. Hayatin. "Deteksi Kemiripan Dokumen Pengajuan Proposal Menggunakan Algoritma Biword Winnowing Pada Sistem Informasi Penelitian Dan Pengabdian." *Jurnal Repositor*, vol. 2, no. 5, pp. 571-582, 2020.

[14] N. Nurdin and A. Munthoha. "Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing." *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*.;vol. 2, no. 1, pp. 90-97, Sep, 2017.

[15] F. R. N. Wulan, A. Kunaefi, and A. Permadi. "Deteksi Plagiasi Dokumen Skripsi Mahasiswa Menggunakan Metode N-Grams Dan Winnowing." *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, vol 9. no. 2, pp. 1021-1032, 2018.
[16] L. Sugiarto, C. Mulyadi, S. Rihastuti. "Analisa Algoritma String Matching Dan Winnowing Untuk Deteksi Kemiripan Judul Tugas Akhir Perguruan Tinggi." *Jurnal Teknologi Informasi*, vol. 6, no. 2. pp. 97-106, 2020.
[17] Y. Nurdiansyah and F. N. Muharrom. "Implementation of winnowing algorithm based K-gram to identify plagiarism on file text-based document." In *MATEC Web of Conferences*, vol. 164, p. 01048. EDP Sciences, 2018.
[18] F. Abror. "Implementasi Algoritma Winnowing pada Deteksi Penjiplakan." PhD diss., Universitas Muhammdiyah Jember, 2016.