Journal of Cyber Forensics and Security (CYFORS)
Vol.1, No.1 (2024)

WINNOWING ALGORITHM FOR MOBILE-BASED THESIS TITLE SIMILARITY DETECTION

Andra Ramadan Pratama [1)], Suci Mutiara [2)]

[1,2)] Information Technology, Institut Informatics and Business Darmajaya
email : 1911010125.1911010125@mail.darmajaya.ac.id [1)], sucimutiara@darmajaya.ac.id [2)]

Abstract

Determining a thesis title can be a difficult task for students due to concerns about the similarity of proposed titles and the possibility of plagiarism. Currently, the checking and management of thesis titles in the Informatics Engineering study program at the Institute Informatika dan Bisnis Darmajaya is still done conventionally. This study aims to build a system that can automatically check thesis titles to prevent plagiarism. Title checking is performed using the winnowing algorithm, which is an algorithm for comparing similarities between texts or documents with document fingerprinting techniques. By comparing two very different thesis titles using the winnowing algorithm, a similarity percentage of 6.08% is obtained. System testing is conducted using black box testing methods to test the system's user functions and interface. The result of the system can facilitate students in determining the similarity of titles they will propose early on, and the system can help the academic group supervisors to inventory the thesis titles submitted by students and build a new culture in managing thesis titles digitally.

***Keywords*:**
Winnowing, Plagiarism, Black Box Testing, Mobile

Introduction

A thesis is a scientific paper written by undergraduate program students (S1) which discusses a particular topic or field based on studies written by experts, the results of field research, or a development (experiment) [1]. Before starting research, students first determine the topic that will be addressed in the research by reading national and international journals, following research carried out by lecturers, reading suggestions in final assignments that have been made by previous students, or observing problems that occur in scope of higher education and outside higher education [2].

The increasing number of students can cause plagiarism in the thesis title [3]. The process of checking thesis titles in the Informatics Engineering Study Program at the Institute of Informatics and Business Darmajaya is still carried out conventionally. The checking process is carried out by checking one by one the thesis titles submitted by students with data on previous students' thesis titles. Apart from that, the checking process is also carried out by groping, namely by remembering whether the thesis title already exists or not. Apart from that, management of thesis title data is also still done manually so it does not rule out the possibility that there is data that is lost and difficult to find [4]. This problem triggers plagiarism in the title of the thesis that students will submit [5]. Plagiarism is an act of violation, namely stealing another author's work without citing the original reference [6].

To overcome the problem of checking and managing thesis titles, a solution is needed, one solution that can be done is to build a website-based application for checking and managing thesis titles that is user friendly and has a responsive appearance. A website is a page with a collection of text, images and sound that combine to present information [7]. The website was chosen because it is multiplatform and portable. Apart from that, websites can also be built to adjust the screen size of the user's device (responsive) so that the website has one more characteristic, namely being mobile. So that if the website is opened on a small or large device screen, the display will not be damaged [8].

The aim of this research is to build a website-based system to check the level of similarity of the thesis titles that will be submitted by students and to manage the thesis titles which will be carried out by Lecturers in the Scientific Section Group (KBK) of the Informatics Engineering Study Program at the Institute of Informatics and Business Darmajaya. Website development will be carried out using the PHP (Hypertext Preprocessor) programming language and using MySQL (My Structured Query Language) as a data storage medium (database). PHP is a website programming language that is executed on the server side so that the information received by the user is always up to date and also PHP is open source and MySQL is a database management system (DBMS) that has relationships between tables [9].

The website will be built by applying an algorithm that functions to calculate the level of similarity of the title of the thesis that will be submitted by students. The algorithm that will be used is the winnowing algorithm. The winnowing algorithm is an algorithm for calculating the level of similarity between two texts or documents using hashing and document fingerprinting techniques [10]. Before building the website, modelling will be carried out using the Unified Modelling Language (UML). UML is object-oriented modelling which is widely used by developers to build websites [11]–[14].

When building a website, a development method will be used, namely the prototype model software development method. The prototype method is a software development method that is widely used by developers so that during development they can interact with users [15]. The final stage of website development is testing. Testing will be carried out using a testing method, namely black box testing. Black box testing is a test that focuses on system functionality and system appearance (user interface) as parameters for user satisfaction [16].

The object of this research is the Informatics Engineering Study Program of the Institute of Informatics and Business Darmajaya which is located in Bandar Lampung, where this college has 3 faculties and 15 study programs. The number of students, especially in the Informatics Engineering Study Program, continues to increase every year, automatically influencing the number of students who will take a thesis in the future. With the development of this thesis title checking application, it is hoped that it can be used as a solution to the problem of thesis title plagiarism for students taking theses.

Literature Review

There have been several previous studies that utilized the winnowing algorithm to detect the similarity level of a text or string. The following references provide the foundation for this research.

Pratama et al. (2021) stated that the winnowing algorithm can be employed in calculating the similarity level of words by finding common fingerprints in the compared sentences. The system, based on winnowing algorithm calculations, can recommend the acceptance of a proposal title by comparing the percentage similarity of the proposed title with the accepted percentage threshold set in the system [17].

Alamsyah & Rasyidan (2019) outlined the need for implementing an algorithm to determine the similarity of thesis title texts submitted with previously existing titles, one of which involves using the winnowing algorithm to generate the percentage of similarity between the compared titles [18].

Sibarani et al. (2019) proposed that to prevent plagiarism in determining thesis titles and facilitate the Curriculum Committee in checking students' thesis titles, a system is required to detect similarity with titles submitted by students and existing titles. One approach involves using the winnowing algorithm [5].

Research Methods

The method that will be used in this research is the prototype model method as a software development method using the PHP programming language and using the MySQL database and the winnowing algorithm as an algorithm that functions to calculate similarities between thesis titles. Data collection will be carried out using several techniques, namely interviews and literature studies.

3.1 Data Collection

At this stage, the supporting data required for this research is collected. The data collection technique used is as follows.

1. Interviews
   Interviews will be conducted with the Head of the Study Program and Lecturers in the Scientific Section Group (KBK) to determine the level of plagiarism that is tolerated, system requirements, and thesis title data that students have previously created.
2. Literature Study
   Literature studies are carried out to obtain supporting data, such as how algorithms work. The study will be carried out by reading various references such as journals, books/e-books, and also student theses.

3.2 Design System

At this stage, modelling or design of the system to be built will be carried out. The design that will be created in this research is in the form of a use case diagram which can be seen in figure 1, a system design using DFD level 0 or a context diagram which can be seen in figure 2 and also a database design using a class diagram which can be seen in figure 1.

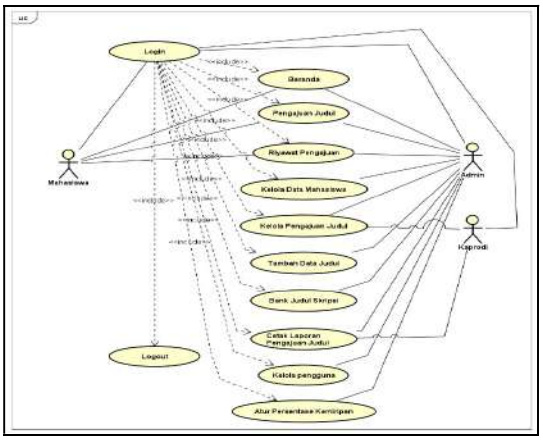

Figure 1. Use Case Diagram

In Figure 1 it can be seen that there are 3 actors or 3 users on the website that will be built, namely students as the main users, Head of the Study Program, and Lecturers in the Scientific Section Group (KBK). Activities that students can carry out are logging in, accessing the homepage, submitting/checking titles, logging out, and checking history. Activities that can be carried out by KBK lecturers are logging in, accessing the homepage, managing student data, managing thesis title data, submitting student titles, reporting on student title submissions and logging out. Activities that can be carried out by the Head of the Study Program are logging in, managing thesis title submissions, viewing reports on thesis title submissions, and logging out.

Figure 2. Context Diagram

The context diagram that can be seen in Figure 2 is a depiction pattern that functions to show the interaction of the system with the environment where the system will be placed [19].

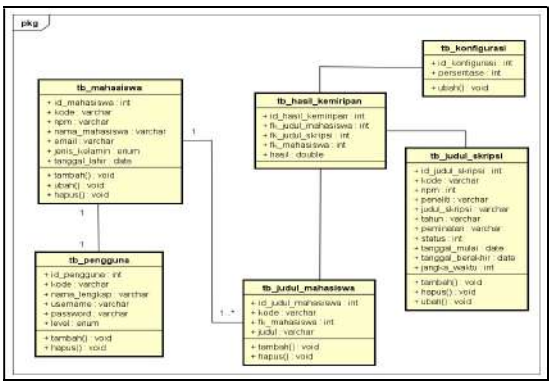

Figure 3. Class Diagram

Figure 3 is the database design used on the website. The database built consists of 6 tables, namely tb_mahasiswa, tb_pengguna, tb_hasil_kemiripan, tb_judul_mahasiswa, tb_konfigurasi, and tb_judul_skripsi

3.3 Winnowing Algorithm

Detection of similarities in thesis titles can be detected using one of the algorithms, namely the winnowing algorithm. Apart from using hashing techniques, the winnowing algorithm also uses document fingerprinting techniques. Document fingerprinting is a grouping of windows formed from a hashing technique obtained from converting each string into a decimal number using ASCII (American Standard Code for Information Interchange) and using the Jaccard coefficient to display the percentage of the fingerprinting value results [10], [20]. The implementation of the winnowing algorithm is as follows.

1. Get 2 thesis titles as samples to compare.
   Title text 1:

"IMPLEMENTASI ALGORITMA *WINNOWING* UNTUK DETEKSI TINGKAT KEMIRIPAN JUDUL SKRIPSI BERBASIS *MOBILE*"

   Title text 2:

"RANCANG BANGUN *WEBSITE* PROGRAM STUDI TEKNIK INFORMATIKA INSTITUT INFORMATIKA DAN BISNIS DARMAJAYA MENGGUNAKAN METODE *EXTREME PROGRAMMING* (XP)"

2. Text Preprocessing
   Remove all characters that are not A-Z, a-z, and 0-9 and change them to lower case (to lower case).
   Text preprocessing title 1:

implementasialgoritmawinnowinguntukdeteksitingkatkemiripanjudulskripsiberbasismobile

   Text preprocessing title 2:

rancangbangunwebsiteprogramstuditeknikinformatikainstitutinformatikadanbisnisdarmajayamenggunakanmetodeextremeprogrammingxp

3. Formation of N-Gram Series
   Create an N-Gram for each thesis title. The N-Gram quantities are prime numbers (2, 5, 7, 11,

13, 17, and so on). Determine the size of the N-Gram is 5, so n = 5.
Formation of N-Gram title 1:

imple mplem pleme lemen ement menta entas ntasi tasia asial sialg ialgo algor lgori gorit oritm ritma itmaw tmawi mawin awinn winno innow nnowi nowin owing wingu ingun ngunt guntu untuk ntukd tukde ukdet kdete detek eteks teksi eksit ksiti sitin iting tingk ingka ngkat gkatk katke atkem tkemi kemir emiri mirip iripa ripan ipanj panju anjud njudu judul uduls dulsk ulskr lskri skrip krips ripsi ipsib psibe siber iberb berba erbas rbasi basis asism sismo ismob smobi mobil obile

Formation of N-Gram title 2:

ranca ancan ncang cangb angba ngban gbang bangu angun ngunw gunwe unweb nwebs websi ebsit bsite sitep itepr tepro eprog progr rogra ogram grams ramst amstu mstud studi tudit udite ditek itekn tekni eknik kniki nikin ikinf kinfo infor nform forma ormat rmati matik atika tikai ikain kains ainst insti nstit stitu titut ituti tutin utinf tinfo infor nform forma ormat rmati matik atika tikad ikada kadan adanb danbi anbis nbisn bisni isnis snisd nisda isdar sdarm darma armaj rmaja majay ajaya jayam ayame yamen ameng mengg enggu nggun gguna gunak unaka nakan akanm kanme anmet nmeto metod etode todee odeex deext eextr extre xtrem treme remep emepr mepro eprog progr rogra ogram grami ramin aming mingx ingxp

4. Rolling Hash
   Rolling hashing is a rolling hashing technique using ASCII. The calculation in the text of title 1 of the first part is "imple" with the base value (b) being b = 2 and the length of the N-Gram (n) being n = 5 using the following formula.

$$Hash = c_1 x b^{(n-1)} + c_2 x b^{(n-1)} + \cdots + c_n x b^{(n-1)} \; x \; b$$

So, the values obtained are as follows

Hash = ((106 * 16) + (109 * 8) + (112 * l4) + (108 * 2) + (101 * 1)) * 2
Hash = (1680 + 872 + 448 + 216 + 101) * 2
Hash = 3317 * 2

The hash value for the "**imple**" string is 6634. Rolling hash is carried out until all n characters have been formed in the n-gram process.

Hash value in title 1:

6634 6766 6758 6568 6456 6642 6538 6822
6798 6388 6774 6410 6328 6658 6636 6898
6886 6714 6918 6632 6508 7030 6682 6854
6888 6942 7014 6632 6776 6746 7114 6940
7042 6892 6498 6362 6554 6854 6516 6778
6928 6702 6898 6566 6644 6462 6534 6438
6878 6560 6482 6724 6666 6832 6580 6674
6380 6786 6748 6942 6610 7048 6818 6948
6766 6894 6688 6858 6776 6388 6250 6458
6662 6258 6462 6938 6712 6914 6684 6594

Hash value in title 2:

6670 6264 6526 6208 6274 6560 6286 6214
6376 6782 6726 7056 6854 6878 6372 6482
6916 6700 6902 6586 6936 6898 6718 6562
6764 6466 6924 7082 7036 6850 6426 6672
6834 6458 6662 6696 6556 6614 6608 6714
6582 6868 6842 6602 6422 6846 6488 6486
6356 6714 6940 7074 7020 6826 7152 7084
6902 6608 6714 6582 6868 6842 6602 6422
6836 6442 6384 6116 6234 6298 6608 6386
6730 6940 6714 6616 6730 6294 6400 6786
6518 6254 6518 6454 6920 6302 6602 6462
6680 6514 6650 6902 6536 6250 6494 6372
6758 6676 6578 6894 6604 6336 6500 6738
7230 6982 6764 6460 6678 6586 6936 6898
6718 6542 6712 6334 6700 6648

5. Window Formation and Rolling Hash Values
   Perform grouping (windowing) on each hash value. Window formation is the same as n-gram by determining the window size (w). It is known that the value of w is w = 4.
   Window in title 1:

W-1: {6634 6766 6758 6568}	W-41: {6928 6702 6898 6566}
W-2: {6766 6758 6568 6456}	W-42: {6702 6898 6566 6644}
W-3: {6758 6568 6456 6642}	W-43: {6898 6566 6644 6462}
W-4: {6568 6456 6642 6538}	W-44: {6566 6644 6462 6534}
W-5: {6456 6642 6538 6822}	W-45: {6644 6462 6534 6438}
W-6: {6642 6538 6822 6798}	W-46: {6462 6534 6438 6878}
W-7: {6538 6822 6798 6388}	W-47: {6534 6438 6878 6560}
W-8: {6822 6798 6388 6774}	W-48: {6438 6878 6560 6482}
W-9: {6798 6388 6774 6410}	W-49: {6878 6560 6482 6724}
W-10: {6388 6774 6410 6328}	W-50: {6560 6482 6724 6666}
W-11: {6774 6410 6328 6658}	W-51: {6482 6724 6666 6832}
W-12: {6410 6328 6658 6636}	W-52: {6724 6666 6832 6580}
W-13: {6328 6658 6636 6898}	W-53: {6666 6832 6580 6674}
W-14: {6658 6636 6898 6886}	W-54: {6832 6580 6674 6380}
W-15: {6636 6898 6886 6714}	W-55: {6580 6674 6380 6786}
W-16: {6898 6886 6714 6918}	W-56: {6674 6380 6786 6748}
W-17: {6886 6714 6918 6632}	W-57: {6380 6786 6748 6942}
W-18: {6714 6918 6632 6508}	W-58: {6786 6748 6942 6610}
W-19: {6918 6632 6508 7030}	W-59: {6748 6942 6610 7048}
W-20: {6632 6508 7030 6682}	W-60: {6942 6610 7048 6818}
W-21: {6508 7030 6682 6854}	W-61: {6610 7048 6818 6948}
W-22: {7030 6682 6854 6888}	W-62: {7048 6818 6948 6766}
W-23: {6682 6854 6888 6942}	W-63: {6818 6948 6766 6894}
W-24: {6854 6888 6942 7014}	W-64: {6948 6766 6894 6688}
W-25: {6888 6942 7014 6632}	W-65: {6766 6894 6688 6858}
W-26: {6942 7014 6632 6776}	W-66: {6894 6688 6858 6776}
W-27: {7014 6632 6776 6746}	W-67: {6688 6858 6776 6388}
W-28: {6632 6776 6746 7114}	W-68: {6858 6776 6388 6250}
W-29: {6776 6746 7114 6940}	W-69: {6776 6388 6250 6458}
W-30: {6746 7114 6940 7042}	W-70: {6388 6250 6458 6662}
W-31: {7114 6940 7042 6892}	W-71: {6250 6458 6662 6258}
W-32: {6940 7042 6892 6498}	W-72: {6458 6662 6258 6462}
W-33: {7042 6892 6498 6362}	W-73: {6662 6258 6462 6938}
W-34: {6892 6498 6362 6554}	W-74: {6258 6462 6938 6712}
W-35: {6498 6362 6554 6854}	W-75: {6462 6938 6712 6914}

W-36: {6362 6554 6854 6516}	W-76: {6938 6712 6914 6684}
W-37: {6554 6854 6516 6778}	W-77: {6712 6914 6684 6594}
W-38: {6854 6516 6778 6928}	
W-39: {6516 6778 6928 6702}	
W-40: {6778 6928 6702 6898}	

Window in title 2:

W-1: {6670 6264 6526 6208}	W-59: {6714 6582 6868 6842}
W-2: {6264 6526 6208 6274}	W-60: {6582 6868 6842 6602}
W-3: {6526 6208 6274 6560}	W-61: {6868 6842 6602 6422}
W-4: {6208 6274 6560 6286}	W-62: {6842 6602 6422 6836}
W-5: {6274 6560 6286 6214}	W-63: {6602 6422 6836 6442}
W-6: {6560 6286 6214 6376}	W-64: {6422 6836 6442 6384}
W-7: {6286 6214 6376 6782}	W-65: {6836 6442 6384 6116}
W-8: {6214 6376 6782 6726}	W-66: {6442 6384 6116 6234}
W-9: {6376 6782 6726 7056}	W-67: {6384 6116 6234 6298}
W-10: {6782 6726 7056 6854}	W-68: {6116 6234 6298 6608}
W-11: {6726 7056 6854 6878}	W-69: {6234 6298 6608 6386}
W-12: {7056 6854 6878 6372}	W-70: {6298 6608 6386 6730}
W-13: {6854 6878 6372 6482}	W-71: {6608 6386 6730 6940}
W-14: {6878 6372 6482 6916}	W-72: {6386 6730 6940 6714}
W-15: {6372 6482 6916 6700}	W-73: {6730 6940 6714 6616}
W-16: {6482 6916 6700 6902}	W-74: {6940 6714 6616 6730}
W-17: {6916 6700 6902 6586}	W-75: {6714 6616 6730 6294}
W-18: {6700 6902 6586 6936}	W-76: {6616 6730 6294 6400}
W-19: {6902 6586 6936 6898}	W-77: {6730 6294 6400 6786}
W-20: {6586 6936 6898 6718}	W-78: {6294 6400 6786 6518}
W-21: {6936 6898 6718 6562}	W-79: {6400 6786 6518 6254}
W-22: {6898 6718 6562 6764}	W-80: {6786 6518 6254 6518}
W-23: {6718 6562 6764 6466}	W-81: {6518 6254 6518 6454}
W-24: {6562 6764 6466 6924}	W-82: {6254 6518 6454 6920}
W-25: {6764 6466 6924 7082}	W-83: {6518 6454 6920 6302}
W-26: {6466 6924 7082 7036}	W-84: {6454 6920 6302 6602}
W-27: {6924 7082 7036 6850}	W-85: {6920 6302 6602 6462}
W-28: {7082 7036 6850 6426}	W-86: {6302 6602 6462 6680}
W-29: {7036 6850 6426 6672}	W-87: {6602 6462 6680 6514}
W-30: {6850 6426 6672 6834}	W-88: {6462 6680 6514 6650}
W-31: {6426 6672 6834 6458}	W-89: {6680 6514 6650 6902}
W-32: {6672 6834 6458 6662}	W-90: {6514 6650 6902 6536}
W-33: {6834 6458 6662 6696}	W-91: {6650 6902 6536 6250}
W-34: {6458 6662 6696 6556}	W-92: {6902 6536 6250 6494}
W-35: {6662 6696 6556 6614}	W-93: {6536 6250 6494 6372}
W-36: {6696 6556 6614 6608}	W-94: {6250 6494 6372 6758}
W-37: {6556 6614 6608 6714}	W-95: {6494 6372 6758 6676}
W-38: {6614 6608 6714 6582}	W-96: {6372 6758 6676 6578}
W-39: {6608 6714 6582 6868}	W-97: {6758 6676 6578 6894}
W-40: {6714 6582 6868 6842}	W-98: {6676 6578 6894 6604}
W-41: {6582 6868 6842 6602}	W-99: {6578 6894 6604 6336}
W-42: {6868 6842 6602 6422}	W-100: {6894 6604 6336 6500}
W-43: {6842 6602 6422 6846}	W-101: {6604 6336 6500 6738}
W-44: {6602 6422 6846 6488}	W-102: {6336 6500 6738 7230}
W-45: {6422 6846 6488 6486}	W-103: {6500 6738 7230 6982}
W-46: {6846 6488 6486 6356}	W-104: {6738 7230 6982 6764}
W-47: {6488 6486 6356 6714}	W-105: {7230 6982 6764 6460}
W-48: {6486 6356 6714 6940}	W-106: {6982 6764 6460 6678}
W-49: {6356 6714 6940 7074}	W-107: {6764 6460 6678 6586}
W-50: {6714 6940 7074 7020}	W-108: {6460 6678 6586 6936}
W-51: {6940 7074 7020 6826}	W-109: {6678 6586 6936 6898}
W-52: {7074 7020 6826 7152}	W-110: {6586 6936 6898 6718}
W-53: {7020 6826 7152 7084}	W-111: {6936 6898 6718 6542}
W-54: {6826 7152 7084 6902}	W-112: {6898 6718 6542 6712}
W-55: {7152 7084 6902 6608}	W-113: {6718 6542 6712 6334}
W-56: {7084 6902 6608 6714}	W-114: {6542 6712 6334 6700}
W-57: {6902 6608 6714 6582}	W-115: {6712 6334 6700 6648}
W-58: {6608 6714 6582 6868}	

6. Fingerprint Selection from Each Window
Fingerprint selection is done by taking the smallest value from each window. The fingerprint formed is as follows.
Fingerprint in title 1:

6568 6456 6456 6456 6456 6538 6388 6388
6388 6328 6328 6328 6328 6636 6636 6714
6632 6508 6508 6508 6508 6682 6682 6854
6632 6632 6632 6632 6746 6746 6892 6498
6362 6362 6362 6362 6516 6516 6516 6702
6566 6566 6462 6462 6438 6438 6438 6438
6482 6482 6482 6580 6580 6380 6380 6380
6380 6610 6610 6610 6610 6766 6766 6688
6688 6688 6388 6250 6250 6250 6250 6258
6258 6258 6462 6684 6594

Fingerprint in title 2:

6208 6208 6208 6208 6214 6214 6214 6214
6376 6726 6726 6372 6372 6372 6372 6482
6586 6586 6586 6586 6562 6562 6466 6466
6466 6466 6850 6426 6426 6426 6426 6458
6458 6458 6556 6556 6556 6582 6582 6582
6582 6422 6422 6422 6422 6356 6356 6356
6356 6714 6826 6826 6826 6826 6608 6608
6582 6582 6582 6582 6422 6422 6422 6384
6116 6116 6116 6116 6234 6298 6386 6386
6616 6616 6294 6294 6294 6294 6254 6254
6254 6254 6302 6302 6302 6302 6462 6462
6514 6514 6250 6250 6250 6250 6372 6372
6578 6578 6336 6336 6336 6336 6500 6738
6460 6460 6460 6460 6586 6586 6542 6542
6334 6334 6334

7. Similarity Using the Jaccard Coeficient
The formula used to calculate similarity using the Jaccard coefficient is as follows.

$$similarity = \frac{intersection}{(union - intersection)} * 100$$

Based on the stages that have been carried out, it is known:

a. Number of fingerprints for title 1 = 77
b. b. Number of fingerprints for title 2 = 115
c. c. Combined (union) fingerprint titles 1 and 2 = 192
d. d. Fingerprint intersection of titles 1 and 2 = 11
e. e. Union minus intersection (union – intersection) = 181

Thus, the following calculation results are obtained.

$$similarity = \frac{11}{(181)} * 100$$
$$similarity = 6.08\%$$

Based on the results obtained, it can be concluded that the title of thesis 1 with the title of thesis 2 is **not similar / not the same**, with the percentage of similarity obtained being 6.08%.

Results and Discussion

The result of the research carried out is a website-based application that is used to check thesis titles, submit thesis titles, and manage thesis titles in the Informatics Engineering Study Program at the Institute of Informatics and Business Darmajaya.

4.1 Website Appearance

The display of the website for checking student thesis titles that has been built is as follows.

1. Login View

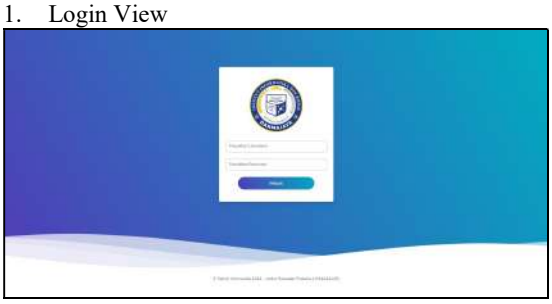

Figure 4. Login View

Figure 4 is the first page displayed when a user accesses the website. Users must enter their credentials, namely a username and password in order to use the website.

2. Home View

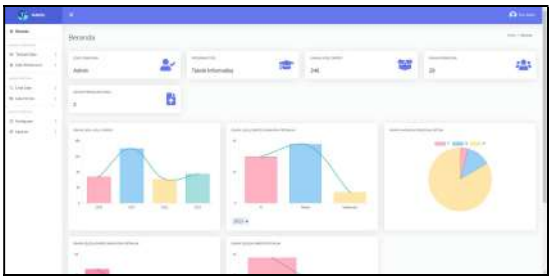

Figure 5. Home View

Figure 5 is the home page that is displayed when the user successfully enters the website using their respective credentials. This page contains data information on the system such as the number of students, and thesis title data graphs.

3. Menu Display for "Manage User Data"

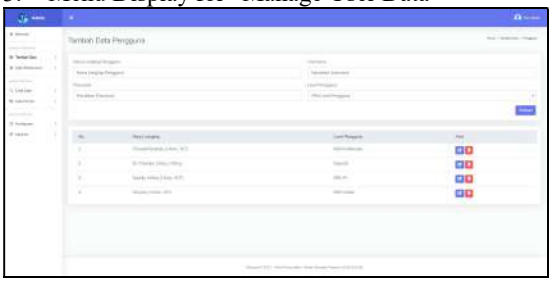

Figure 6. Manage User Data

Figure 6 is a page used by the admin to manage special user data, namely KBK Lecturers and Heads of Study Programs so they can enter and manage the website that has been built with their respective roles in accordance with the use case diagram that has been designed.

4. Menu Display for "Student Title Submission"

Figure 7. Student Title Submission

Figure 7 is the page that will be used to manage the thesis titles submitted by students. This page will contain the title of the thesis, student, and the percentage of similarity to the title of the proposed thesis.

5. Menu Display for "Manage Student Data"

Figure 8. Manage Student Data

Figure 8 is the page that will be used to manage student data. On this page CRUD operations (create, read, update and delete) can be performed.

6. Menu Display for "Thesis Title Bank"

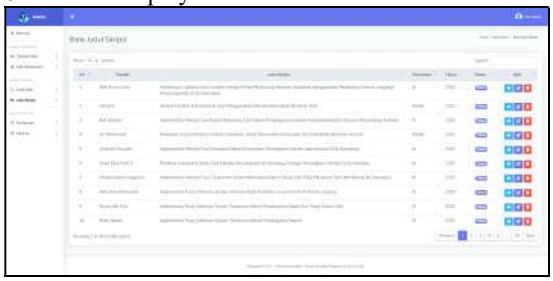

Figure 9. Thesis Title Bank

Figure 9 is a display of the thesis title bank page. These thesis titles are thesis titles in the Information Engineering Study Program at the Institute of Informatics and Business Darmajaya from 2020 to 2023.

7. Menu Display for "Student Title Report"

Figure 10. Student Title Report

Figure 10 is a display of the thesis title submission report made by students. The report will be in PDF form which has headers containing university information and a body containing NPM data, student name, thesis title and level of plagiarism.

8. Menu Display for "Thesis Title Checker"

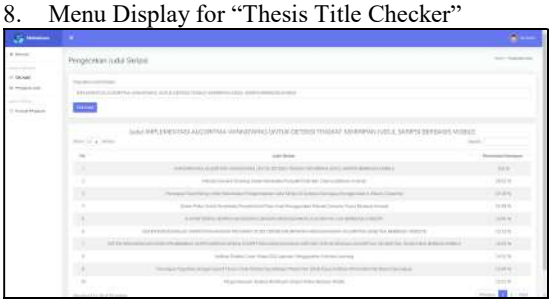

Figure 11. Thesis Title Checker

Figure 11 is a display of the page used to check the title of the thesis. This page contains a thesis title form and a button for checking. The output from this page is the percentage of plagiarism with titles in the system database.

9. Menu Display for "Thesis Title Submission"

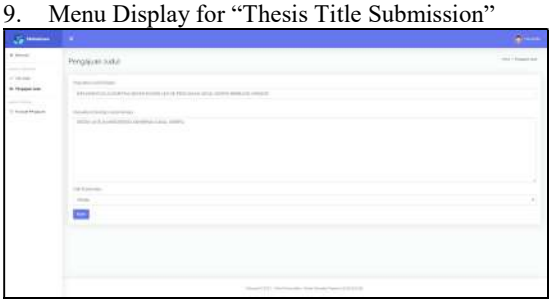

Figure 12. Thesis Title Submission

Figure 12 is a display of the thesis title submission page. On this page there is a form that must be filled in, namely the thesis title and a brief description of the proposed title.

10. Menu Display for "Thesis Title Submission History"

Figure 13. Thesis Title Submission History

Figure 13 is a display of the history page for thesis title submissions that have been made by students. This page displays data in the form of the thesis title, submission status, and percentage of plagiarism.

4.2 System Testing

System testing is carried out on websites that have been built using black box testing. Black box testing is a testing technique that focuses on the functionality and appearance of the user interface. In functionality testing, a series of test cases are carried out by the user. The functions tested include login, manage users, manage students, manage thesis titles, print title submission reports, check titles, submit titles, and check the history of thesis title submissions. The results of functionality testing showed that the website built was in line with expectations and all functionality ran well without any errors.

Testing the user interface is carried out by using an extension called Mobile Device Simulator. With this extension, you can change the desktop display to a smartphone, tablet or laptop display with various screen sizes. Testing was carried out using 3 different screen sizes, namely smartphone, tablet and laptop (desktop). The results of testing the user interface showed that the website built was mobile and responsive on various types of device screen sizes without damaging the appearance of the website.

Conclusions and Recommendations

Based on the research that has been carried out, several conclusions are obtained as follows.

1. This system can help students in submitting their thesis titles. Students can find out in advance about the level of similarity of the title of the thesis that will be submitted.
2. This system can help Scientific Section Group Lecturers (KBK) in managing and inventorying thesis titles submitted by students.
3. This system was built using responsive website technology (uncluttered layout on various screen sizes) so that it can be run on various mobile platforms (smartphones, tablets and PC/laptops).
4. Building a new culture internally in the Informatics Engineering Study Program by

utilizing digitalization of the student thesis process.

References

[1] H. Wakhyudin and A. D. S. Putri, "Analisis Kecemasan Mahasiswa Dalam Menyelesaikan Skripsi," *WASIS J. Ilm. Pendidik.*, vol. 1, no. 1, pp. 14–18, 2020, doi: 10.24176/wasis.v1i1.4707.

[2] W. Astuti, "Analisis String Matching Pada Judul Skripsi Dengan Algoritma Knuth-Morris Pratt (Kmp)," *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 167–172, Aug. 2018.

[3] A. Amrulloh, D. Januarita, A. Kusuma, Y. Saintika, and A. D. Septiadi, "Model rapid application development (rad) pada pengembangan aplikasi penjadwalan mata kuliah," *J. TEKINKOM*, vol. 6, pp. 26–32, 2023, doi: 10.37600/tekinkom.v6i1.579.

[4] Muliadi, M. R. Said, and E. Sofyan, "Peqguruang: Conference Series," *J. Peqguruang Conf. Ser.*, vol. 3, pp. 1–4, 2021.

[5] L. Sibarani, M. Magdalena, and A. Dharma, "Analisa Perbandingan Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing Dan Algoritma Rabin Karp," *REMIK (Riset dan E-Jurnal Manaj. Inform. Komputer)*, vol. 4, no. 1, p. 69, 2019, doi: 10.33395/remik.v4i1.10174.

[6] M. A. Shadiqi, "Memahami dan Mencegah Perilaku Plagiarisme dalam Menulis Karya Ilmiah," *Bul. Psikol.*, vol. 27, no. 1, p. 30, 2019, doi: 10.22146/buletinpsikologi.43058.

[7] T. A. Kinaswara, N. R. Hidayati, and F. Nugrahanti, "Rancang Bangun Aplikasi Inventaris Berbasis Website Pada Kelurahan Bantengan | Kinaswara | Prosiding Seminar Nasional Teknologi Informasi dan Komunikasi (SENATIK)," *Pros. Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2, no. 1, pp. 71–75, 2019, [Online]. Available: http://prosiding.unipma.ac.id/index.php/SENATIK/article/view/1073

[8] M. D. Ariawan, A. Triayudi, and I. D. Sholihati, "Perancangan User Interface Design dan User Experience Mobile Responsive Pada Website Perusahaan," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 161, 2020, doi: 10.30865/mib.v4i1.1896.

[9] A. A. Rizky and I. Ramdhani, "Perancangan Sistem Informasi Perekrutan Karyawan Berbasis Web Menggunakan PHP dan MySQL DI PT. Ria Indah Mandiri," *J. Manaj. Inform.*, vol. 9, no. 1, pp. 49–57, 2019, doi: 10.34010/jamika.v9i1.1651.

[10] S. Schleimer, D. S. Wilkerson, and A. Aiken, "Winnowing: Local Algorithms for Document Fingerprinting," Apr. 2003, doi: 10.1145/872757.872770.

[11] R. Destriana, S. Maulana Husain, N. Handayani, and A. Tegar Prahara Siswanto, *Diagram UML Dalam Membuat Aplikasi Android Firebase "Studi Kasus Aplikasi Bank Sampah."* Tangerang: Deepublish, 2021.

[12] M. Teguh Prihandoyo, "Unified Modeling Language (UML) Model Untuk Pengembangan Sistem Informasi Akademik Berbasis Web," *J. Inform. J. Pengemb. IT*, vol. 3, no. 1, Jan. 2018.

[13] I. Tanu Kusnadi, A. Supiandi, R. Nurafifah Syabaniah, and R. Oktapiani, *Pemodelan Sistem Berbasis Objek with UML*, 1st ed. Yogyakarta: GRAHA ILMU, 2019.

[14] R. S. Pressman, *Software Engineering*. 2010.

[15] Rusdiansyah, "MEMBANGUN PROTOTYPE SISTEM INFORMASI ARSIP ELEKTRONIK SURAT PERJANJIAN KERJASAMA PADA BUSINESS SUPPORT DEPARTEMENT," vol. 14, no. 2, p. 157, 2018, [Online]. Available: www.bsi.ac.id

[16] S. Supriyono, "Software Testing with the approach of Blackbox Testing on the Academic Information System," *IJISTECH (International J. Inf. Syst. Technol.*, vol. 3, no. 2, pp. 227–233, 2020, [Online]. Available: https://ijistech.org/ijistech/index.php/ijistech/article/view/54

[17] N. Putra and B. Pratama, "Pengembangan Sistem Informasi Pengelolaan Judul Skripsi dan Tugas Akhir dengan Fitur Deteksi Kemiripan Menggunakan Algoritma Winnowing," 2021.

[18] N. Alamsyah and M. Rasyidan, "Deteksi Plagiarisme Tingkat Kemiripan Judul Skripsi Pada Fakultas Teknologi Informasi Menggunakan Algoritma Winnowing," Oktober-Desember, Dec. 2019.

[19] M. Muliadi, M. Andriani, and H. Irawan, "Perancangan Sistem Informasi Pemesanan Kamar Hotel Berbasis Website (Web) Menggunakan Data Flow Diagram (Dfd)," *JISI J. Integr. Sist. Ind.*, vol. 7, no. 2, p. 111, 2020, doi: 10.24853/jisi.7.2.111-122.

[20] N. Fitrianingsih, M. Asfi, D. Prasetyo, R. Perdana Kusuma, and M. Afif Sulhan, "Deteksi Tingkat Kemiripan Judul Menggunakan Algoritma OliverPada Sistem Informasi Pengajuan Skripsi," *J. MEDIA Inform. BUDIDARMA*, vol. 6, pp. 2148–2156, Oct. 2022, doi: 10.30865/mib.v6i4.4409.