

**IMPLEMENTASI SISTEM DETEKSI INDIKASI  
PLAGIARISME MENGGUNAKAN METODE N-GRAMS DAN  
DOCUMENT FINGERPRINTING BERBASIS WEB**

**SKRIPSI**



**DISUSUN OLEH :**

**FITRI RATNA NING WULAN**

**H76214010**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL**

**SURABAYA  
2018**

## **LEMBAR PERNYATAAN KEASLIAN**

Saya yang bertanda tangan dibawah ini:

Nama : Fitri Ratna Ning Wulan  
NIM : H76214010  
Program Studi : Sistem Informasi  
Angkatan : 2014

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: IMPLEMENTASI SISTEM DETEKSI INDIKASI PLAGIARISME MENGGUNAKAN METODE *N-GRAMS* DAN *DOCUMENT FINGERPRINTING* BERBASIS WEB. Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya akan menerima sanksi yang telah ditetapkan. Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 03 Agustus 2018

Yang membuat pernyataan,



Fitri Ratna Ning Wulan  
NIM: H76214010

**PERSETUJUAN PEMBIMBING**

Skripsi oleh :

NAMA : FITRI RATNA NING WULAN

NIM : H76214010

JUDUL : IMPLEMENTASI SISTEM DETEKSI INDIKASI  
PLAGIARISME MENGGUNAKAN METODE N-GRAMS DAN  
*DOCUMENT FINGERPRINTING* BERBASIS WEB

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 19 Juli 2018

Dosen Pembimbing I

Dosen Pembimbing II

Anang Kunaefi, M. Kom  
NIP. 197911132014031001

Andhy Permadi, M. Kom  
NIP. 198110142014031002

## PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh Fitri Ratna Ning Wulan ini telah dipertahankan

Didepan tim Penguji Skripsi

Surabaya, 25 Juli 2018

Mengesahkan Fakultas Sains dan Teknologi

Universitas Islam Negeri Sunan Ampel Surabaya

Susunan Dewan Pengaji

Penguji I

Penguji III

Penguji II

Andhy Permadi, M. Kom  
NIP. 198110142014031002

Moch. Yasin, M. Kom, M.B.A.  
NIP. 198808302014031001

Nita Yalina, S. Kom, M. MT  
NIP. 19870282014032003

Mengesahkan  
Dekan Fakultas Sains dan Teknologi  
UIN Sunan Ampel Surabaya

Dr. Eni Purwati, M.Ag  
NIP. 196512211990022001



**KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA  
PERPUSTAKAAN**

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300  
E-Mail: perpus@uinsby.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Fitri Ratna Ning Wulan  
NIM : 1176214010  
Fakultas/Jurusan : Sains dan Teknologi/Sistem Informasi  
E-mail address : fitiratna96@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi     Tesis     Desertasi     Lain-lain (.....)  
yang berjudul :

**IMPLEMENTASI SISTEM DETEKSI INDIKASI PLAGIARISME  
MENGGUNAKAN METODE N-GRAMS DAN DOCUMENT FINGERPRINTING  
BERBASIS WEB**

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 03 Agustus 2018

Penulis

( ..... )  
*Fitri Ratna Ning Wulan*

## ABSTRAK

# **IMPLEMENTASI SISTEM DETEKSI INDIKASI PLAGIARISME MENGGUNAKAN METODE *N*-GRAMS DAN *DOCUMENT FINGERPRINTING* BERBASIS WEB**

## Oleh: **Fitri Ratna Ning Wulan**

Plagiarisme dalam dunia pendidikan sangat diperhatikan, mahasiswa sering melakukan tindakan plagiasi salah satunya mencantumkan karya orang lain tanpa menyertakan sumber. Plagiarisme dapat diminimalisir dengan mendeteksi kesamaan kata pada dokumen, salah satu metode yang digunakan yaitu metode *n-grams*. Penelitian ini mencoba untuk mengimplementasikan sistem deteksi indikasi plagiarisme menggunakan metode *n-grams* dan *document fingerprinting* dengan mencari nilai *n-grams* yang paling efektif digunakan dalam mendeteksi indikasi plagiarisme dari perbandingan hasil pengujian dengan sistem dan secara manual. Penelitian ini menggunakan metode *n-grams* dengan  $n = 3,4,5,6,7$  untuk mencari nilai  $n$  yang paling efektif, algoritma *winnowing* dan *jaccard coefficient* untuk menentukan prosentase kemiripan. Hasil dari penelitian ini yaitu sistem mampu mendeteksi indikasi plagiarisme berdasarkan kesamaan kata dan nilai  $n$  yang paling efektif digunakan untuk mendeteksi kesamaan dokumen adalah  $n = 6$  dan  $n = 7$  berdasarkan 3 skenario dengan prosentase selisih rata-rata (akurasi) antara pengujian secara sistem dan manual paling kecil adalah 3,19% dan 2,16%.

**Kata kunci:** Plagiarisme, *N-Grams*, Document Fingerprinting, Rolling Hash, Winnowing, Jaccard Coefficient.

## ABSTRACT

# **IMPLEMENTATION DETECTION SYSTEM OF PLAGIARISM INDICATION USING N-GRAMS AND DOCUMENT FINGERPRINTING METHODS BASED ON WEB**

By:  
**Fitri Ratna Ning Wulan**

In education, plagiarism is very concerned, students often do plagiarism one of them using the work of others without mentioning the sources. Plagiarism could be minimized by detecting the same word in a document, one of the method used is the *n-grams* method. This research is trying to implement detection system of plagiarism indication using *n-grams* and *document fingerprinting* methods by finding the most effective *n-grams* that is used in detecting plagiarism indications from comparing test results of systems and manual. This research use *n-grams* method with  $n = 3,4,5,6,7$  to find the most effective *n* value, *winnowing algorithm* and *jaccard coefficient* to determine the percentage of similarity. The result of this research is the system that can detect indication of plagiarism based on similarity of words. The most effective *n* value that is used to detect similarity document are  $n = 6$  and  $n = 7$  based on 3 scenarios with the smallest percentage of average difference (accuracy) between testing by system and manual are 3.19% and 2.16%.

**Keywords:** Plagiarisme, *N-Grams*, Document Fingerprinting, Winnowing Algorithm, Jaccard Coefficient.

DAFTAR ISI

LEMBAR PERNYATAAN KEASLIAN .....	i
LEMBAR PERSETUJUN PEMBIMBING .....	ii
LEMBAR PENGESAHAN .....	iii
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI.....	iii
ABSTRAK .....	v
ABSTRACT .....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR .....	ix
DAFTAR TABEL .....	xi
DAFTAR LAMPIRAN .....	xiv
<b>BAB 1 .....</b>	<b>1</b>
1.1    Latar Belakang .....	1
1.2    Perumusan Masalah .....	3
1.3    Batasan Masalah .....	3
1.4    Tujuan Penelitian .....	3
1.5    Manfaat Penelitian .....	3
1.6    Sistematika Penulisan .....	4
<b>BAB 2 .....</b>	<b>5</b>
2.1    Tinjauan Penelitian Terdahulu .....	5
2.2    Teori dasar yang digunakan .....	6
2.2.1    Plagiarisme.....	6
2.2.2 <i>Natural Language Processing (NLP)</i> .....	7
2.2.3 <i>Rolling Hash</i> .....	11
2.2.4    Algoritma <i>Document Fingerprinting (Winnowing)</i> .....	12
2.2.5 <i>Jaccard Coefficient</i> .....	13
2.2.6    Bahasa Pemrograman <i>PHP</i> .....	13
2.2.7    Basis Data ( <i>Database</i> ) .....	14
2.3    Integrasi Keilmuan.....	15
<b>BAB 3 .....</b>	<b>17</b>
3.1    Tahapan Penelitian.....	17
3.1.1    Studi Literatur .....	17

3.1.2	Pengumpulan Data .....	17
3.1.3	Pengolahan Data .....	18
3.1.4	Desain dan Implementasi .....	18
3.1.5	Pengujian.....	18
3.2	Jadwal Penelitian .....	19
<b>BAB 4</b>	.....	<b>20</b>
4.1	Pengumpulan Data .....	20
4.2	Proses Sistem .....	23
4.2.1	Alur Proses Sistem.....	23
4.2.2	Gambaran Pemodelan <i>Data Flow Diagram</i> .....	24
4.3	Pengolahan Data .....	27
4.3.1	<i>Preprocessing</i> .....	27
4.3.2	N-Grams.....	31
4.3.3	Hashing .....	33
4.3.4	Algoritma Winnowing .....	34
4.4	Desain dan Implementasi .....	37
4.4.1	Desain Basis Data .....	37
4.4.2	Implementasi.....	38
4.5	Pengujian.....	41
4.5.1	Pengujian Sistem.....	41
4.6	Pembahasan.....	84
<b>BAB 5</b>	.....	<b>86</b>
5.1	Kesimpulan .....	86
5.2	Saran .....	86
<b>DAFTAR PUSTAKA</b>	.....	<b>87</b>
<b>LAMPIRAN</b>	.....	<b>90</b>

## **DAFTAR GAMBAR**

Gambar 2. 1 Contoh N-Grams .....	11
Gambar 2. 2 Rumus Rolling Hash 1 .....	12
Gambar 2. 3 Rumus Rolling Hash 2 .....	12
Gambar 2. 4 Rumus Jaccard Coefficient.....	13
Gambar 3. 1 Tahapan Penelitian .....	17
Gambar 4. 1 Alur Sistem.....	23
Gambar 4. 2 DFD Context diagram .....	24
Gambar 4. 3 DFD Level 0.....	25
Gambar 4. 4 DFD Level 1.....	26
Gambar 4. 5 Alur Preprocessing .....	27
Gambar 4. 6 Alur Rangkaian N-Grams .....	31
Gambar 4. 7 Alur Hashing .....	33
Gambar 4. 8 Alur Algoritma Winnowing .....	35
Gambar 4. 9 Halaman Utama.....	39
Gambar 4. 10 Halaman Deteksi Kalimat .....	39
Gambar 4. 11 Halaman Unggah Dokumen Pembanding .....	40
Gambar 4. 12 Halaman Deteksi Dokumen .....	40
Gambar 4. 13 Unggah Dokumen Pembanding Sistem Informasi .....	42
Gambar 4. 14 Hasil Pengujian Skenario 1 .....	43
Gambar 4. 15 Grafik pengujian n = 3 skenario 1 prodi sistem informasi .....	46
Gambar 4. 16 Grafik pengujian skenario 1 n=4 prodi sistem informasi .....	47
Gambar 4. 17 Grafik pengujian n=5 skenario 1 prodi sistem informasi .....	48
Gambar 4. 18 Grafik pengujian n=6 skenario 1 prodi sistem informasi .....	50
Gambar 4. 19 Grafik pengujian n=7 skenario 1 prodi sistem informasi .....	51
Gambar 4. 20 Hasil Pengujian sistem skenario 1 prodi Biologi .....	53
Gambar 4. 21 Grafik pengujian n=3 skenario 1 prodi biologi .....	55
Gambar 4. 22 Grafik Pengujian n=4 skenario 1 prodi Biologi .....	56
Gambar 4. 23 Grafik Pengujian n=5 skenario 1 prodi Biologi .....	57
Gambar 4. 24 Hasil Pengujian N=6 Skenario 1 prodi Biologi.....	58

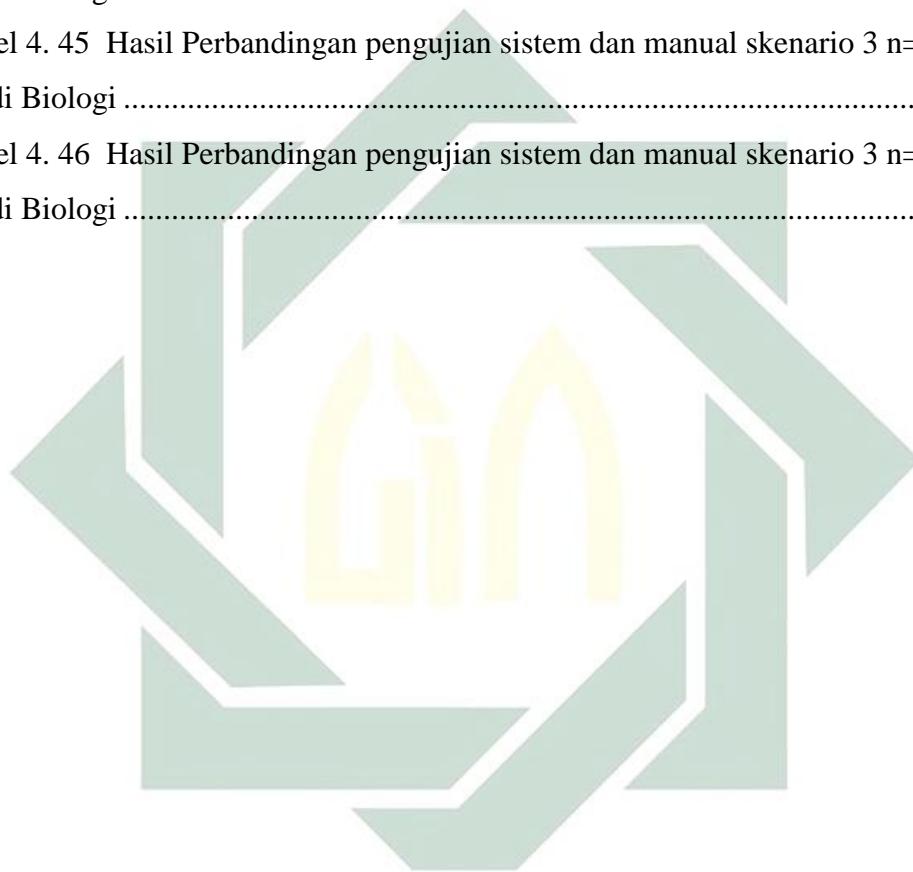
Gambar 4. 25 Grafik Pengujian n=7 skenario 1 prodi Biologi .....	60
Gambar 4. 26 Tampilan Upload Dokumen Sistem Informasi Skenario 2 .....	62
Gambar 4. 27 Hasil Pengujian sistem skenario 2 prodi Sistem Informasi .....	63
Gambar 4. 28 Hasil Pengujian sistem skenario 2 prodi Biologi .....	65
Gambar 4. 29 Tampilan Upload Dokumen Sistem Informasi Skenario 3 .....	67
Gambar 4. 30 Hasil Pengujian sistem skenario 3 prodi Sistem Informasi .....	68
Gambar 4. 31 Grafik Pengujian n=3 skenario 3 prodi sistem informasi.....	70
Gambar 4. 32 Grafik pengujian n=4 skenario 3 prodi sistem informasi.....	71
Gambar 4. 33 Grafik pengujian n=5 skenario 3 prodi Sistem Informasi .....	72
Gambar 4. 34 Grafik Pengujian n=6 skenario 3 prodi Sistem Informasi .....	74
Gambar 4. 35 Grafik Pengujian n=7 skenario 3 prodi Sistem Informasi .....	75
Gambar 4. 36 Hasil Pengujian sistem skenario 3 prodi Biologi .....	77
Gambar 4. 37 Grafik pengujian n=3 Skenario 3 Prodi Biologi .....	79
Gambar 4. 38 Hasil Pengujian n=4 skenario 3 prodi Biologi .....	80
Gambar 4. 39 Grafik Pengujian n=5 skenario prodi Biologi .....	81
Gambar 4. 40 Grafik Pengujian n=6 skenario 3 prodi Biologi .....	82
Gambar 4. 41 Grafik Pengujian n=7 skenario 3 prodi Biologi .....	83
Gambar 4. 42 Hasil Pengujian Rata-Rata Selisih (Akurasi) .....	85

## DAFTAR TABEL

Tabel 2. 1 Awalan & Akhiran .....	9
Tabel 2. 2 Peluruhan Awalan & Imbuhan.....	10
Tabel 3. 1 Jadwal Penelitian.....	19
Tabel 4. 1 Data Proposal skripsi Prodi Sistem Informasi .....	20
Tabel 4. 2 Data Proposal skripsi Prodi Biologi.....	21
Tabel 4. 3 Atribut Basis Data.....	37
Tabel 4. 4 Spesifikasi Perangkat Keras.....	38
Tabel 4. 5 skenario 1 dokumen pembanding sistem informasi .....	41
Tabel 4. 6 skenario 1 dokumen testing sistem informasi .....	42
Tabel 4. 7 Hasil Pengujian Sistem dokumen sistem informasi .....	44
Tabel 4. 8 Hasil Pengujian Manual skenario 1 dokumen sistem informasi .....	45
Tabel 4. 9 Hasil Pengujian secara sistem dan manual N=3 Skenario 1 .....	46
Tabel 4. 10 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=4 prodi sistem informasi.....	47
Tabel 4. 11 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=5 prodi sistem informasi.....	49
Tabel 4. 12 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=6 prodi sistem informasi.....	50
Tabel 4. 13 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=7 prodi sistem informasi.....	51
Tabel 4. 14 Dokumen Testing skenario 1 Prodi Biologi.....	52
Tabel 4. 15 Dokumen Pembanding Skenario 1 Prodi Biologi .....	52
Tabel 4. 16 Hasil Pengujian sistem skenario 1 prodi Biologi .....	54
Tabel 4. 17 Hasil Pengujian Manual Skenario 1 Prodi Biologi .....	54
Tabel 4. 18 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=3 prodi Biologi .....	55
Tabel 4. 19 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=4 prodi Biologi .....	56

Tabel 4. 20 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=5 prodi Biologi .....	58
Tabel 4. 21 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=6 prodi Biologi .....	59
Tabel 4. 22 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=7 prodi Biologi .....	60
Tabel 4. 23 Dokumen Testing skenario 2 prodi sistem informasi .....	61
Tabel 4. 24 Dokumen pembanding skenario 2 prodi sistem informasi.....	61
Tabel 4. 25 Hasil Pengujian skenario 2 prodi sistem informasi.....	64
Tabel 4. 26 Dokumen Testing skenario 2 prodi Biologi.....	64
Tabel 4. 27 Dokumen Pembanding skenario 2 prodi Biologi.....	64
Tabel 4. 28 Hasil Pengujian sistem skenario 2 prodi Biologi.....	66
Tabel 4. 29 Dokumen Testing skenario 3 prodi Sistem Informasi .....	67
Tabel 4. 30 Dokumen Pembanding skenario 3 prodi sistem informasi .....	67
Tabel 4. 31 Hasil Pengujian sistem skenario 3 prodi Sistem Informasi.....	69
Tabel 4. 32 Hasil pengujian manual skenario 3 prodi sistem informasi .....	69
Tabel 4. 33 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=3 prodi sistem informasi.....	71
Tabel 4. 34 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=4 prodi sistem informasi.....	72
Tabel 4. 35 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=5 prodi sistem informasi.....	73
Tabel 4. 36 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=6 prodi sistem informasi.....	74
Tabel 4. 37 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=7 prodi sistem informasi.....	75
Tabel 4. 38 Dokumen Testing Skenario 3 Prodi Biologi .....	76
Tabel 4. 39 Dokumen Pembanding Skenario 3 Prodi Biologi .....	76
Tabel 4. 40 Hasil Pengujian sistem skenario 3 prodi Biologi .....	78
Tabel 4. 41 Hasil pengujian manual skenario 3 prodi sistem informasi .....	78

Tabel 4. 42 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=3	
prodi Biologi .....	79
Tabel 4. 43 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=4	
prodi Biologi .....	80
Tabel 4. 44 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=5	
prodi Biologi .....	81
Tabel 4. 45 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=6	
prodi Biologi .....	82
Tabel 4. 46 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=7	
prodi Biologi .....	84



## **DAFTAR LAMPIRAN**

Lampiran 1 Source Code pemrosesan Data .....	90
Lampiran 2 Source Code Upload Dokumen .....	92

# BAB 1

## PENDAHULUAN

## 1.1 Latar Belakang

Teknologi informasi saat ini memudahkan masyarakat dalam mendapatkan informasi, salah satunya dengan memanfaatkan internet. Berdasarkan hasil *survey* Asosiasi Penyelenggara Jasa Internet Indonesia Pengguna (APJII) tentang penetrasi dan perilaku pengguna internet Indonesia tahun 2017 disebutkan bahwa pengguna internet di Indonesia meningkat dari tahun 2016 yang berjumlah 132,7 juta orang menjadi 143,26 juta orang pada tahun 2017. Berdasarkan hasil *survey* tersebut dapat dikatakan adanya internet berpengaruh dalam kehidupan masyarakat. Banyaknya sumber yang bisa menjadi referensi di internet, dapat memicu terjadi tindakan plagiasi (Stamatatos, 2009).

Plagiat menurut Kamus Besar Bahasa Indonesia adalah karangan atau pendapat orang lain yang diambil dan diakui sebagai karangan atau pendapat sendiri (Bahasa, 2008). Plagiarisme dalam dunia pendidikan sangat diperhatikan seperti dalam penulisan skripsi mahasiswa. Skripsi merupakan tugas yang wajib dikerjakan mahasiswa strata tingkat 1 di semester akhir sebagai syarat kelulusan.

Berdasarkan hasil survei yang dilakukan *Pew Research Center* yang dilakukan musim semi 2011 bersama asosiasi *Chronicle of High Education* dengan responden sebanyak 1,055 mahasiswa melalui *survey* secara *online*. Disebutkan bahwa dengan adanya teknologi memberikan dampak negatif kepada mahasiswa diantaranya plagiarism dan kecurangan dalam mengerjakan tugas. Penelitian tersebut menyebutkan 89% mahasiswa berpendapat bahwa komputer dan teknologi merupakan sumber dari adanya plagiarism (Parker, Lenhart, & Moore, 2011).

Plagiarisme berdasarkan kesamaan kata antar dokumen salah satu contohnya yaitu mahasiswa sekarang sering melakukan *copy paste* dalam mengerjakan tugas yang diberikan, baik itu dari internet atau hasil tugas orang lain. Tindakan tersebut dapat memicu mahasiswa melakukan plagiasi salah satu contohnya pada penulisan proposal skripsi, dengan mengambil kutipan atau pernyataan tanpa menyertakan sumbernya.

Plagiarisme dapat dicegah dengan mendeteksi kemiripan kata antar dokumen, tetapi hal tersebut membutuhkan waktu lama jika dilakukan secara manual. Pendektsian bisa dilakukan dengan algoritma dimana peningkatan akurasi sistem dalam pendektsian menjadi faktor utama (Dillak, Laumal, Kadja, & S, 2016). Adanya sistem yang dapat membantu dalam mendekksi plagiarisme pada dokumen proposal dapat memudahkan dalam pengecekan terutama dari pada efisiensi waktu.

Ada beberapa kriteria dasar yang digunakan dalam mendeteksi plagiarisme pada kata diantaranya yaitu : (1) *whitespace intensitivity* dimana dalam mencari kesamaan kalimat tidak terpengaruh dengan spasi, tanda baca huruf kapital, (2) *Noise suppresion* dimana menghindari kata yang tidak relevan, (3) *Position Independence* dimana posisi kata yang berbeda tidak mempengaruhi pencarian kesamaan kalimat (Schleimer, Wilkerson, Aiken, & Berkeley, n.d.).

Algoritma *n-gram* merupakan salah satu algoritma yang digunakan untuk mendapatkan nilai kemiripan kata antar dokumen. Algoritma *n-gram* mengambil karakter pada teks sesuai dengan panjang *n* dimulai dari posisi tertentu, sehingga kemiripan dapat diperoleh meskipun posisi kata berbeda (Lisangan, 2015). Kelebihan Algoritma *n-gram* tidak memperhatikan kesalahan dalam penulisan yang ada di dokumen (Dillak et al., 2016). Panjang *n* mempengaruhi dalam proses mendeteksi kemiripan kata, sehingga masih belum diketahui nilai *n* yang paling efektif digunakan dalam mendeteksi kemiripan kata pada dokumen.

Selain algoritma *n-gram*, untuk mengecek kemiripan kata dapat menggunakan algoritma *winnowing* yang merupakan salah satu algoritma untuk pengecekan kesamaan kata yang disebut dengan *document fingerprinting* (Wibowo & Hastuti, 2016). *Fingerprint* kata digunakan untuk pengecekan kemiripan kata pada dokumen.

Berdasarkan penjelasan dari permasalahan diatas, dengan adanya sistem yang mengimplementasikan algoritma *n-grams* dan *document fingerprinting* dalam mendeteksi kesamaan kata diharapkan dapat diketahui nilai *n* paling efektif digunakan untuk mendeteksi kemiripan kata pada dokumen. Serta dengan adanya sistem deteksi indikasi plagiarisme diharapkan dapat membantu dosen dalam mencegah adanya plagiarisme pada penulisan proposal skripsi mahasiswa.

Berdasarkan penjelasan tersebut, peneliti melakukan penelitian yang berjudul "Implementasi Sistem Deteksi Indikasi Plagiarisme Menggunakan Metode *N-Grams* dan *Document Fingerprinting* berbasis web".

## 1.2 Perumusan Masalah

Berdasarkan latar belakang diatas, dapat diambil rumusan masalah yaitu :

1. Bagaimana mengolah teks pada basis data dokumen proposal skripsi ?
  2. Bagaimana menerapkan algoritma *document fingerprinting* ?
  3. Bagaimana mengimplementasikan algoritma *document fingerprinting* pada aplikasi berbasis web ?

### 1.3 Batasan Masalah

Berikut batasan masalah dalam tugas akhir ini yaitu :

1. Dokumen yang dibahas dalam penelitian ini adalah dokumen proposal skripsi mahasiswa yang menggunakan bahasa Indonesia.
  2. Data yang diambil sebagai bahan uji berasal dari proposal skripsi mahasiswa Program Studi Sistem Informasi dan Biologi Fakultas Sains dan Teknologi Univeristas Islam Negeri Sunan Ampel Surabaya yang sudah melakukan sidang proposal dengan format dokumen *docx*.
  3. Penelitian ini tidak mendeteksi jenis plagiarisme *translation*.
  4. Dokumen yang digunakan sebagai pengujian berukuran maksimal 5Mb.
  5. Sistem yang dibangun tidak bersifat *online*.

#### **1.4 Tujuan Penelitian**

Berikut tujuan dari penelitian tugas akhir ini, diantaranya yaitu :

1. Mampu mengolah teks pada basis data dokumen skripsi.
  2. Mengetahui penerapan algoritma *document fingerprinting* dalam deteksi *plagiarisme*.
  3. Mengimplementasikan algoritma *document fingerprinting* pada aplikasi berbasis *web*.

## **1.5 Manfaat Penelitian**

Berikut manfaat dari penelitian tugas akhir ini, diantaranya yaitu :

1. Untuk Penulis : Penulis ini dapat menambah ilmu dan pengalaman serta mengetahui metode-metode yang digunakan untuk mendeteksi indikasi plagiarisme.
  2. Untuk Akademisi : Penelitian ini dapat menjadi referensi dalam pengembangan ilmu berkaitan dengan metode-metode yang digunakan dalam penelitian ini dan sebagai rujukan untuk penelitian lain dalam mengembangkan sistem yang sama.
  3. Untuk Dosen : sistem yang dibuat dalam penelitian ini dapat membantu dosen dalam mendeteksi indikasi plagiarisme tugas-tugas dari mahasiswa.

## 1.6 Sistematika Penulisan

Sistematika penulisan skripsi setiap bab akan diuraikan sebagai berikut :

## 1. BAB I PENDAHULUAN

Pada bab ini akan menjelaskan latar belakang, rumusan masalah, batasan masaah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

## 2. BAB II TINJAUAN PUSTAKA

Pada bab ini menjelaskan landasan teori yang berkaitan dengan penelitian yang dilakukan dan penelitian sebelumnya yang berkaitan dengan penelitian skripsi.

### 3. BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan metodologi penelitian yang akan digunakan dalam penelitian skripsi.

#### 4. BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan hasil dan pembahasan dari penelitian yang dilakukan

## 5. BAB V PENUTUP

Pada bab ini berisi kesimpulan dan saran berkaitan dengan penelitian.

## 6. DAFTAR PUSTAKA

## 7. LAMPIRAN

## **BAB 2**

# **TINJAUAN PUSTAKA**

## **2.1 Tinjauan Penelitian Terdahulu**

Ada beberapa penelitian yang berhubungan dengan implementasi sistem plagiarisme dengan menggunakan algoritma *N-Grams* dan *document fingerprintint*, diantaranya yaitu:

Penelitian yang membahas sistem deteksi plagiarisme "Sistem Deteksi Dini Plagiarisme Tugas Akhir Mahasiswa Menggunakan Algoritma N-Grams dan Winnowing" (Dillak et al., 2016). Penelitian tersebut mengimplementasikan sistem deteksi plagiarisme pada tugas akhir mahasiswa dengan menggunakan algoritma *N-Grams* untuk menghitung bobot dari karakter dan dilakukan perhitungan nilai *hash* dengan menggunakan metode *winnowing* serta dilakukan pengecekan dengan *cosine simmilarity*. Dalam penelitian tersebut menggunakan abstrak dari dokumen skripsi sebagai bahan uji. Pembandingan dokumen dilakukan dengan mengunggah dokumen dalam sistem deteksi. Dari penelitian tersebut diketahui bahwa isi dokumen berpengaruh pada kecepatan proses deteksi.

Penelitian selanjutnya "Implementasi *N-Grams Technique* Dalam Deteksi Plagiarisme pada Tugas Mahasiswa" (Lisangan, 2015). Penelitian tersebut mengimplementasikan algoritma *N-Grams* yang digunakan untuk mendeteksi plagiarisme pada tugas mahasiswa berdasarkan kesamaan kata. Proses implementasi diawali dengan menghilangkan tanda baca dan spasi menggunakan proses *cleansing text* dan dilakukan konversi dalam bentuk *N-Grams*. Pengecekan nilai  $n$  dari karakter dalam teks dibandingkan dengan nilai  $n$  5,6,7, setelah didapatkan nilai  $n$  kemudian dihitung persentase untuk mengetahui tingkat kesamaan antar tugas mahasiswa dengan menggunakan *sorensen-dice coefficient*.

Penelitian selanjutnya "Deteksi Keberadaan Kalimat Sama sebagai Indikasi Penjiplakan dengan Algoritma Hashing Berbasis *N-Grams*" (Purwitasari, Kusmawan, & dkk, 2011). Penelitian tersebut membahas pendekripsi kalimat yang sama hasil *copy paste* antar file teks sebagai indikasi plagiarisme. Proses deteksi dilakukan dengan menggunakan algoritma *winnowing*, dimana nilai *N*-

*Grams* dari teks dirubah menjadi nilai *hash* (*fingerprint*). Parameter yang digunakan diantaranya nilai  $n$  dari  $n$ -grams, bilangan prima  $b$  untuk proses *hashing*, ukuran *window*  $w$ , parameter batas penentuan penjiplakan. Kalimat dianggap terindikasi plagiarisme jika muncul melebihi batas toleransi tertentu. Dalam penelitian tersebut proses pengujian dilakukan dengan membandingkan dua file teks, sehingga tidak bisa dilakukan perbandingan dengan banyak file.

Penelitian selanjutnya "Penerapan Algoritma Winnowing untuk Mendeteksi Kemiripan Teks pada Tugas Akhir Mahasiswa" (Wibowo & Hastuti, 2016). Penelitian tersebut membahas penerapan algoritma *winnowing* dalam pendekripsi kemiripan teks tugas akhir program studi Teknik Informatika Universitas Dian Nuswantoro. Hasil penelitian tersebut didapatkan prosentase kemiripan teks antara dua dokumen menggunakan perhitungan *Jaccard Coefficient*.

Penelitian selanjutnya "Implementasi Algoritma *Winnowing* untuk Deteksi Kemiripan Judul Skripsi Studi Kasus STMIK Budidarma" (Setiawan, 2017). Penelitian tersebut membahas pengimplementasian aplikasi berbasis web yang dapat mendeteksi kemiripan judul skripsi pada STMIK Budidarma. Algoritma *winnowing* digunakan dalam proses pendekripsi plagiarisme dengan judul dan abstrak sebagai bahan pengujian. Dalam penelitian tersebut hasil proses deteksi berupa prosentase kemiripan antar judul yang dibandingkan.

## 2.2 Teori dasar yang digunakan

### **2.2.1 Plagiarisme**

Menurut Kamus Besar Bahasa Indonesia plagiat adalah pengambilan karangan atau orang lain dan disiarkan sebagai karangan atau pendapat sendiri. Jiplak atau menjiplak adalah mencontoh atau meniru (tulisan, pekerjaan orang lain), menggambar atau menulis mengikuti garis-garis gambaran atau tulisan yang telah tersedia (dengan menempelkan pada gambar atau tulisan yang akan ditiru), mencuri karangan orang lain, mengutip karangan orang lain tanpa menyebutkan sumbernya atau mengaku sebagai karangannya sendiri (Bahasa, 2008).

Sedangkan menurut Peraturan Menteri Pendidikan Nasional tentang pencegahan dan penanggulangan plagiat di perguruan tinggi Bab I Ketentuan

Umum Pasal 1 ayat 1, Plagiat adalah perbuatan secara sengaja atau tidak sengaja dalam memperoleh atau mencoba memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya dan/atau karya ilmiah pihak lain yang diakui sebagai karya ilmiahnya, tanpa menyatakan sumber secara tepat dan memadai (“Permen17-2010.pdf,” n.d.).

Menurut (Alzahrani Salha. & Abraham, 2011) Plagiarisme dibagi menjadi 2 tipe yaitu *Literal Plagiarism* dan *Intelligent Plagiarism*.

### **1. Literal Plagiarism**

Tipe plagiarisme dimana dicontohkan dengan perbuatan *copy-paste* tulisan yang bersumber dari internet tanpa menyebutkan sumber referensi dokumen yang asli.

## 2. Intelligent Plagiarism

Salah satu tipe plagiarisme dimana tidak menyertakan sumber dari tulisan orang lain, dan mengakui sebagai tulisan sendiri.

a. *loatText Manipulation*

Manipulasi teks dimana digunakan sinonim untuk merubah penulisan dari suatu kalimat dan tidak menyertakan sumber asli.

b. *Translation*

Salah satu ciri-ciri plagiarisme dimana menggunakan tulisan hasil terjemahan tulisan orang lain dengan bahasa yang berbeda tanpa menyertakan sumber aslinya.

### c. Idea Adoption

Salah satu ciri-ciri plagiarisme dimana melakukan adopsi dari ide tulisan orang lain berupa hasil, referensi dan kesimpulan suatu tulisan tanpa menyertakan sumber asli.

### **2.2.2 Natural Language Processing (NLP)**

Menurut Jackson dan Moulinier *Natural Language Processing (NLP)* digunakan untuk menggambarkan fungsi dari software atau hardware sistem komputer yang dianalisa dengan bahasa alami (Jackson & Moulinier, 2007).

### **2.2.2.a Tokenizing**

*Tokenizing* merupakan suatu metode untuk mengubah kalimat menjadi sekumpulan kata (*token*) berdasarkan susunan kata (Retrieval, 2008). Contoh terdapat kalimat berikut.

"saya menjadi mahasiswa sistem informasi di surabaya"

Ketika dilakukan *tokenizing* menjadi :

"saya", "menjadi", "mahasiswa", "sistem", "informasi", "di", "surabaya".

### **2.2.2.b Stop word**

*Stop word* merupakan suatu metode yang digunakan untuk menghilangkan kata yang tidak relevan dalam suatu kalimat (Crc, n.d.). Contoh kata yang dihilangkan seperti : "dari", "di", "[spasi]" dan lain sebagainya. Contoh menghilangkan kata yang tidak relevan seperti berikut :

"saya", "menjadi", "mahasiswa", "sistem", "informasi", "di", "surabaya".

Ketika dilakukan proses *stopword* akan berubah menjadi seperti berikut :

"saya", "menjadi", "mahasiswa", "sistem", "informasi", "surabaya".

Kata 'di' menjadi hilang, karena kata tersebut termasuk kata yang tidak relevan.

### **2.2.2.c Stemming**

Menurut (Ramasubramanian & Ramya, 2013) *Stemming* merupakan suatu metode yang digunakan untuk mengubah kata menjadi bentuk dasar dengan menghilangkan imbuhan dan kata hubung.

Algoritma *Stemming* banyak digunakan dalam kata dengan bahasa berbeda-beda, diantaranya yaitu untuk *Stemming* kata dengan bahasa inggris menggunakan algoritma Porter dan Lovin. Untuk *Stemming* kata dengan bahasa indonesia menggunakan algoritma Nazief & Adriani, Vega, Arifin & Setiono. Disebutkan juga untuk *stemming* kata bahasa indonesia lebih kompleks dibandingkan dengan bahasa inggris, karena harus menghilangkan kata imbuhan untuk mendapatkan kata dasar (Williams & Tahaghoghi, 2005).

Menurut (Williams & Tahaghoghi, 2005) dijelaskan bahwa penelitian tersebut sudah membandingkan beberapa teknik dalam *stemming* kata bahasa indonesia, dan

menunjukkan bahwa teknik dari Nazief dan Adriani berhasil melakukan *stemming* kata bahasa indonesia dengan keberaran 93%.

Berikut langkah-langkah *stemming* menurut algoritma Nazief & Adriani dalam paper (Williams & Tahaghoghi, 2005), sebagai berikut :

1. Jika kata yang akan di *stemming* sama dengan kata yang ada di kamus, maka kata tersebut merupakan kata dasar dan tidak perlu dilakukan *stemming*. Jika kata tidak sama dengan yang ada di kamus, maka dilanjutkan ke tahap kedua.
  2. Menghilangkan imbuhan akhiran *particle* ("-lah, "-kah") serta akhiran *Possesive Pronoun* ("-ku", "-mu", "-nya"). Contoh kata "bukunya" ketika dilakukan *stemming* akan menjadi "buku". Jika kata yang sudah sama dengan kamus, maka algoritma berhenti.
  3. Menghilangkan imbuhan akhiran ("-i", "-an"), jika berhasil maka berlanjut pada tahap selanjutnya, jika tidak berhasil masuk kedalam tahapan berikut :
    - a. Jika kata imbuhan berakhiran "-an" dan sudah dihapus tetapi jika kata imbuhan akhiran setelahnya adalah "-k" maka kata "-k" dihapus, jika tidak berhasil lanjut pada proses selanjutnya.
    - b. Mengembalikan kata imbuhan akhiran yang sudah dihapus sebelumnya ("-i", "-an", "-kan")
  4. Menghilangkan imbuhan awalan, berikut langkah-langkah yang dilakukan :
    - a. Mengecek kata dengan daftar awalan yang tidak diperbolehkan berpasangan dengan akhiran yang dijelaskan pada Tabel 2.1 Kombinasi Awalan dan Ahiran yang tidak dibolehkan, sebagai berikut :

Tabel 2. 1 Awalan & Akhiran

<b>Awalan (<i>prefixes</i>)</b>	<b>Akhiran (<i>Suffixes</i>)</b>
be-	-i
di-	-an
ke-	-i-kan
me-	-an
se-	-i-kan
te-	-an

- b. Mengecek awalan yang ada jika sama dengan awalan sebelum dihapus, algoritma berhenti.
  - c. Mengecek kata sudah tidak mempunyai awalan, algoritma berhenti.
  - d. Mengetahui tipe awalan berdasarkan langkah-langkah berikut :
    - i. Jika kata awalan seperti "di-", "ke-", "se-" akan langsung dihapus.
    - ii. Jika kata awalan seperti "te-", "be-", "me-", "pe-" akan dihapus menggunakan aturan seperti Tabel 2.2 yang diambil dari algoritma Nazief & Adriani (Kurniawan, Effendi, & Sitompul, 2012), sebagai berikut :

*Tabel 2. 2 Peluruhan Awalan & Imbuhan*

Aturan	Awalan	Peluruhan
1	BerV...	Ber-V be-Rv...
2	Belajar...	Bel-ajar
3	beC <sub>1</sub> erC <sub>2</sub> ...	Be-C <sub>1</sub> erC <sub>2</sub> ... dimana C <sub>1</sub> !={r' l'}
4	terV...	ter-V te-rV...
5	terCer..	ter-Cer... dimana C!=r'
6	teC <sub>1</sub> erC <sub>2</sub> ...	te-C <sub>1</sub> erC <sub>2</sub> ... dimana C <sub>1</sub> !=r'
7	Me{l r w y}V...	Me-{l r w y}V...
8	Mem{b f v}...	Mem-{b f v}...
9	Mempe...	Mem-pe...
10	Mem{rV V}...	Me-m{rV V}... me-p{rV V}...
11	Men{c d j z}...	Men-{c d j z}...
12	MenV...	Me-nV me-Tv
13	Meng{g h q k}...	Meng-{g h q k}...
14	MengV...	Meng-V...  Meng-kV...
15	MengeC...	Menge-C...
16	menyV...	Me-ny... meny-sV...
17	mempV...	Mem-pV...
18	Pe{w y}V...	Pe-{w y}V...
19	perV...	Per-V... pe-rV...
20	Pem{b f v}...	Pem-{b f v}...
21	Pem{rV V}...	Pe-m{rV V}... pe-p{rV V}
22	Pen{c d j z}...	Pen-{c d j z}...
23	penV...	Pe-nV... pe-tV...
24	Peng{g h q}...	Peng-{g h q}...
25	pengV...	Peng-V peng-kV...
26	penyV...	Pe-nya peny-sV...
27	pelV...	pelV...; kecuali kata "pelajar"
28	peCP	Pe-CP... dimana C!= {r w y l m n} dan P!=er'
29	perCerV	Per-CerV... dimana C!= {r w y l m n}

#### **2.2.2.d N-Grams**

Menurut Cavnar & Trenkle yang dikutip (Hornik, Mair, Rauch, Geiger, & Buchta, 2013) algoritma *N-Grams* merupakan suatu algoritma untuk melakukan pemisahan text yang panjang menjadi *string* berdasarkan panjang tertentu ( $n$ ). Panjang  $n$  bermacam-macam diantaranya  $n=1$  (*unigrams*),  $n=2$  (*bigrams*), dan  $n=3$  (*trigrams*). Berikut contoh dari pemisahan kata menjadi *string* menggunakan kata "corpus", sebagai berikut:

$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
-	_c	_co	_cor	_corp
c	co	cor	corp	corpu
o	or	orp	orpu	opus
r	rp	rpu	rpus	rpus_
p	pu	pus	pus_	pus__
u	us	us_	us__	us___
s	s_	s__	s___	s____

*Gambar 2. 1 Contoh N-Grams*

Gambar 2.1 menggambarkan contoh rangkaian *n-grams* mulai dari  $n=1$  sampai  $n=7$ , banyaknya  $n$  mempengaruhi banyaknya rangkaian kata yang terbentuk. Berikut contoh hasil dari rangkaian *n-grams* jika  $n=5$ , seperti berikut :

```
{"sayaj", "ayaja", "yajad", "ajadi", "jadim", "adima", "dimah",
"imaha", "mahas", "ahasi", "hasis", "asisw", "siswa", "iswas", "swasi",
"wasis", "asist", "siste", "istem", "stemi", "temin", "eminf", "minfo",
"infor", "nform", "forma", "ormas", "rmasi", "masis", "asisu", "sisur",
"isura", "surab", "uraba", "rabay", "abaya"}
```

### 2.2.3 *Rolling Hash*

Rolling hash merupakan metode untuk menghitung nilai *hash* dari rangkaian *n-grams* yang digunakan pada algoritma *winnowing* dalam menghitung kesamaan kata. Berikut rumus untuk menghitung nilai *hash* dari rangkaian *n-grams* (Setiawan, 2017).

$$H_{(c_1 \dots c_n)} = c_1 * b^{(n-1)} + c_2 * b^{(n-2)} + \dots + c_{(n-1)} * b^{(n)} + c_n$$

*Gambar 2. 2 Rumus Rolling Hash 1*

Gambar 2.2 merupakan rumus *Rolling Hash* pertama, diketahui  $b$  merupakan nilai bilangan prima yang tidak ditentukan,  $c$  merupakan bilangan ASCII dari karakter, dan  $n$  merupakan panjang  $n$ -grams. Untuk rumus kedua digunakan untuk rangkaian  $n$ -grams kedua dan seterusnya, seperti pada Gambar 2.3 berikut:

$$H_{(c_1 \dots c_{n+1})} = (H_{(c_1 \dots c_n)} - c_1 * b^{(n-1)}) * b + c_{(n+1)}$$

*Gambar 2. 3 Rumus Rolling Hash 2*

Gambar 2.3 merupakan rumus kedua yang digunakan untuk rangkain  $n$ -grams kedua dan seterusnya, hal tersebut dikarenakan tidak diperlukan iterasi karakter pertama pada rangkaian  $n$ -grams.

#### 2.2.4 Algoritma Document Fingerprinting (Winnowing)

Algoritma *Winnowing* merupakan salah satu algoritma yang digunakan untuk menganalisa *document fingerprinting* dengan merubah kumpulan kata dalam bentuk *hash* (Schleimer et al., n.d.). Hasil Algoritma *winnowing* dapat memberikan informasi letak dari *fingerprint* meskipun antar dokumen berbeda posisi.

Langkah-langkah yang dilakukan dalam menerapkan algoritma *winnowing* untuk mendeteksi adanya plagiarisme dalam dokumen (Schleimer et al., n.d.) sebagai berikut:

1. Memasukkan dokumen yang akan di analisa.
  2. Melakukan *tokenizing* pada kalimat untuk memecahkan kalimat menjadi beberapa kata.
  3. Melakukan *stop word* pada kata yang tidak relevan.
  4. Mengubah kata menjadi bentuk *n-gram*.
  5. Melakukan proses algoritma *winnowing* untuk memilih *fingerprint*.

### 2.2.5 Jaccard Coefficient

Jaccard Coefficient merupakan suatu metode untuk menghitung kesamaan pada data set yang biasa digunakan pada data set bilangan biner (Aggarwal, 2015). Berikut rumus untuk menghitung kesamaan *similarity* berdasarkan *Jaccard Coefficient* secara umum seperti pada gambar 2.4 berikut:

$$J(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}.$$

### *Gambar 2. 4 Rumus Jaccard Coefficient*

Gambar 2.4 menggambarkan rumus *Jaccard Coefficient* dengan  $J(S_1, S_2)$  merupakan nilai kesamaan antara dataset  $S_1$  dan  $S_2$ ,  $S_1 \cap S_2$  merupakan irisan atau data yang sama antara kedua dataset, dan  $S_1 \cup S_2$  merupakan *union* atau gabungan data dari kedua dataset.

## 2.2.6 Bahasa Pemrograman *PHP*

Bahasa pemrograman PHP juga disebut sebagai bahasa *scripting* atau bahasa pemrograman yang *embedded* yang artinya hanya perlu membaca sintaks dalam memproses pemrograman. Secara singkat sejarah bahasa pemrograman *PHP* menurut (Sakur, 2010) dimulai ketika model antarmuka diciptakan oleh Rasmus Lerdorf untuk menampung informasi di situsnya. Rasmus menggunakan PERL untuk membuat antarmuka yang kemudian dikembangkan dengan bahasa C di antarmuka. Antarmuka tersebut dinamakan *Personal Home Page (PHP)* yang bisa mencatat semua informasi dari pengunjung di situsnya.

PHP kemudian dikembangkan sehingga muncullah PHP versi 2.0 pada November 1997 yang merupakan awal dari pengembangan bahasa pemrograman PHP hingga mencapai versi 5.x.x saat ini. Zend saat ini menjadi pengembang utama PHP dan sampai sekarang PHP banyak diminati pengguna karena memberikan kemudahan.

Berikut beberapa contoh penulisan *script* menggunakan bahasa pemrograman PHP.

```
<?php  
echo "Hello World!";
```

?>

Hasil dari *script* diatas ketika dijalankan yaitu : Hello World!

### **2.2.7 Basis Data (*Database*)**

Basis data atau *database* adalah suatu kumpulan informasi atau data yang berhubungan satu sama lain dan disimpan secara terstruktur (Garcia-Molina et al., 2010). *Database* dikelola menggunakan perangkat lunak yang disebut dengan *Database Management Systems* (DBMS). DBMS membantu pengguna dalam mengelola *database* yang ada, mulai dari mengontrol, melakukan aktivitas yang berhubungan dengan *database* (Solichin, 2010). Salah satu perangkat lunak DBMS yang sering digunakan yaitu *MySQL*.

*MySQL* merupakan salah satu DBMS yang familiar digunakan dengan total 6 juta instalasi di dunia. Kelebihan *MySQL* diantaranya bisa digunakan pada semua sistem operasi seperti *Linux*, *Mac OS X*, *Open BSD* dan *Windows* secara gratis. *MySQL* juga bisa digunakan pada beberapa bahasa pemrograman, diantaranya yaitu bahasa pemrograman C, C++, Java, Perl, PHP, dan Ruby (W.Jason Gilmore, 2010).

Menurut (W.Jason Gilmore, 2010) MySQL didukung beberapa tipe data yang dibagi menjadi tiga kategori, yaitu *date and time*, *numeric*, dan *string*. Berikut penjelasan dari masing-masing kategori, seperti berikut :

## 1. Tipe data *Date and Time*

Tipe data *date and time* memberikan informasi yang berhubungan dengan tanggal dan waktu. Seperti tipe data *DATE, DATETIME, TIME, TIMESTAMP, dan YEAR*.

## 2. Tipe data *Numeric*

Tipe data *numeric* merepresentasikan data yang berupa angka. Seperti tipe data *BOOLEAN*, *BIGINT*, *INT*, *MEDIUMINT*, *SMALLINT*, *TINYINT*, *DECIMAL*, *DOUBLE*, dan *FLOAT*.

### 3. Tipe data *String*

Tipe Data *String* merepresentasikan tipe data *string*, seperti tipe data ***CHAR***, ***VARCHAR***, DAN ***TEXT***.

## 2.3 Integrasi Keilmuan

Integrasi kelimuan antara penelitian yang dilakukan dengan Islam dilakukan dengan melakukan wawancara pada pakar, pada penelitian ini peneliti melakukan wawancara pada Bapak Dr. H. Suis, M. Fil. I, beliau sekarang menjabat sebagai kepala program studi Dhirosah Islamiyah Pascasarjana Universitas Islam Negeri Sunan Ampel Surabaya dan sebagai Dosen Fakultas Syariah Universitas Islam Negeri Sunan Ampel Surabaya. Riwayat pendidikan non-formal beliau di Pondok Wonorejo selama 7,5 tahun dan di Pondok Tebu Ireng Jombang selama 3 tahun. Pendidikan formal beliau dengan lulusan S1 Sastra Arab, S2 Pemikiran Islam, S3 Pemikiran Islam di IAIN Sunan Ampel Surabaya.

Menurut hasil wawancara yang dilakukan plagiat menurut Islam merupakan salah satu **كذب** (*Al kidzbu*) atau berbohong dan termasuk dosa yang sangat tinggi amal jariyah dosanya. Bapak Suis menjelaskan bahwa dulu para ulama tidak pernah mengarang kitab sehingga ketika mengajarkan pada santrinya dengan cara berceramah. Ketika santrinya mengarang kitab, ulamanya selalu disebut dan tidak pernah berbohong sehingga tidak ada yang plagiat. Beliau juga menjelaskan ada hadist bukhori yang menjelaskan tentang berbohong, salah satunya yaitu :

مَنْ كَذَبَ عَلَىٰ مُتَعَمِّدًا فَلْيَتَبُوأْ مَقْعَدَهُ مِنَ النَّارِ

"Barang siapa yang berbohong dengan sengaja maka siap-siap tempat duduknya adalah di neraka." (H.R Bukhori)

Berdasarkan hadist diatas dapat diketahui bahwa plagiat yang termasuk dalam kategori berbohong yang dimana nanti akan bertempat di neraka. Bapak Suis juga menjelaskan dalam Al-Quran ada ayat yang menjelaskan tentang plagiat yang termasuk dalam kategori berbohong. Berikut surat Al-Hujurat ayat 6 :

يَا أَيُّهَا الَّذِينَ آمَنُوا إِنْ جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَنْ تُصِيبُوا  
قَوْمًا بِجَهَالَةٍ فَلَا تُصِيبُوهُمْ عَلَى مَا فَعَلُتُمْ نَادِمِينَ

*"Hai orang-orang yang beriman, jika datang padamu orang fasik membawa suatu berita, maka periksalah dengan teliti agar kamu tidak menimpakan suatu*

*"musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu." (Q.S Al-Hujurat:6)*

Berdasarkan ayat diatas dapat disimpulkan bahwa akan lebih baik jika kita memeriksa terlebih dahulu apa yang sudah diberitakan, ketika dikaitkan dengan penelitian ini akan lebih baik jika memeriksa indikasi plagiarisme terlebih dahulu sehingga dapat mencegah adanya plagiarisme, dimana plagiarisme merupakan tindakan berbohong.

Penelitian ini akan sangat berguna bagi dosen dalam mendeteksi indikasi plagiarisme pada tugas-tugas yang dikumpulkan mahasiswa, agar dosen terhindar dari plagiasi yang dilakukan mahasiswa. Dosen tidak perlu mengecek secara manual karena membutuhkan waktu yang lama. Adanya sistem deteksi indikasi plagiarisme dapat meningkatkan efisiensi sehingga dosen dapat mengerjakan kegiatan lain. Sebagaimana dijelaskan dalam firman Allah yang tercantum pada Al-Quran surat Al-Insyirah ayat 7-8, sebagaimana berikut:

فَإِذَا فَرَغْتَ فَانْصَبْ (٧) وَإِلَيْ رَبِّكَ فَارْغَبْ (٨)

"Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain (7) dan hanya kepada Tuhanmu lah hendaknya kamu berharap(8)." (Q.S Al-Insyirah:7-8)

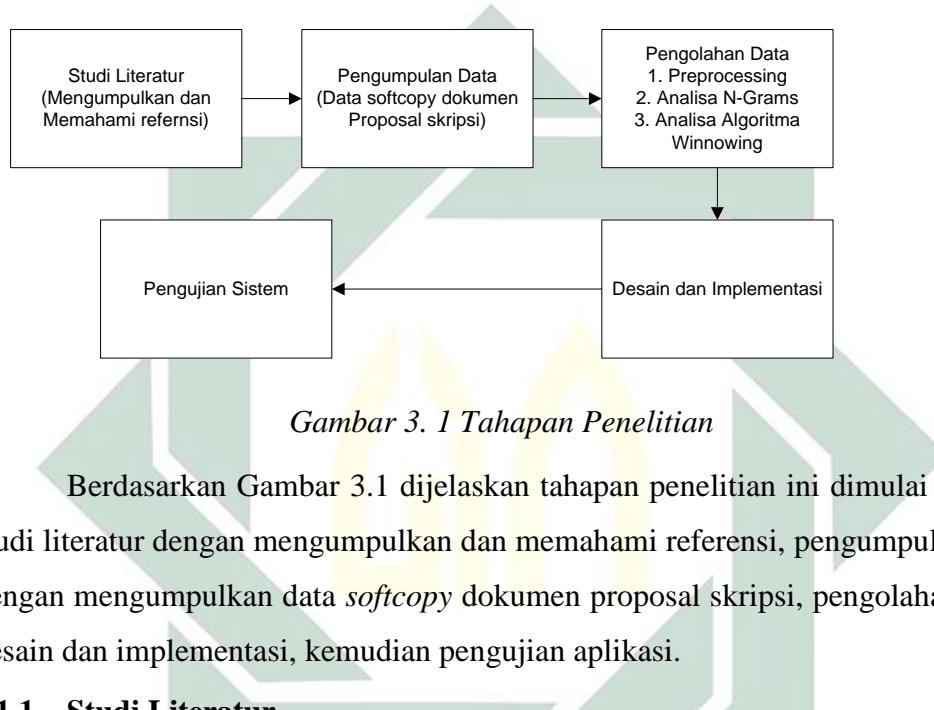
Berdasarkan ayat tersebut dapat disimpulkan bahwa hendaknya segera menyelesaikan pekerjaan yang sudah diberikan dan segera mengerjakan tugas yang lain sehingga semua pekerjaan dapat berjalan dengan lancar. Seperti adanya sistem deteksi indikasi plagiarisme dapat membantu dosen dalam mendeteksi indikasi plagiarisme yang dilakukan mahasiswa dan tidak membutuhkan waktu yang lama, sehingga dosen dapat mengerjakan pekerjaan yang lain.

## **BAB 3**

### **METODOLOGI PENELITIAN**

### **3.1 Tahapan Penelitian**

Ada beberapa tahapan yang dilakukan sebagai alur dalam menyelesaikan penelitian tugas akhir ini, seperti pada Gambar 3.1 berikut:



Berdasarkan Gambar 3.1 dijelaskan tahapan penelitian ini dimulai dengan studi literatur dengan mengumpulkan dan memahami referensi, pengumpulan data dengan mengumpulkan data *softcopy* dokumen proposal skripsi, pengolahan data, desain dan implementasi, kemudian pengujian aplikasi.

### **3.1.1 Studi Literatur**

Studi literatur dilakukan untuk mendapatkan informasi tentang permasalahan yang akan diteliti dengan mengumpulkan dan jurnal-jurnal penelitian sebelumnya dan buku dasar teori yang berkaitan tentang algoritma untuk mendeteksi plagiarisme. Dari studi literatur didapatkan metode yang digunakan untuk mendeteksi indikasi plagiarisme.

### 3.1.2 Pengumpulan Data

Pengumpulan data dilakukan untuk mengumpulkan dokumen-dokumen proposal skripsi mahasiswa untuk pengujian aplikasi deteksi indikasi plagiarisme. Proposal skripsi untuk pengujian menggunakan proposal skripsi mahasiswa Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Ampel Surabaya angkatan 2014, dari jurusan teknologi pada program studi sistem informasi

sebanyak 20 proposal, dan jurusan sains pada program studi biologi sebanyak 17 proposal. Jumlah pengumpulan data tersebut berdasarkan jumlah mahasiswa yang sudah melakukan sidang proposal angkatan pertama untuk program studi Sistem Informasi dan Biologi.

### **3.1.3 Pengolahan Data**

Pengolahan data dilakukan untuk menyelesaikan permasalahan yang akan diteliti dan diterapkan dalam sistem, berikut langkah-langkah yang dilakukan dalam mendeteksi plagiarisme menggunakan algoritma *N-gram* dan *document fingerprinting*:

## 1. Analisa Preprocessing

Dalam tahapan *Preprocessing* dilakukan proses *tokenizing*, *stop word*, dan *stemming*.

## 2. Analisa Algoritma N-Grams

Dalam tahapan ini dilakukan perubahan dari bentuk kata menjadi rangkaian *N-Grams* dengan urutan sesuai struktur kalimat.

### 3. Analisa Algoritma Winnowing

Dalam tahapan ini dilakukan proses perubahan dari bentuk rangkaian *N-Grams* menjadi *hash* untuk menentukan *fingerprinting*. Dengan menggunakan algoritma *winnowing* dapat menentukan tingkat kesamaan kata antar dokumen.

### 3.1.4 Desain dan Implementasi

Desain aplikasi dilakukan untuk mendesain struktur dan tampilan *interface* dari aplikasi deteksi indikasi plagiarisme. Dalam sistem deteksi indikasi plagiarisme yang dibangun berbasis *web*, menggunakan bahasa pemrograman *PHP*, dan *database* yang digunakan yaitu *MySQL* untuk menyimpan dokumen serta *fingerprint* dari dokumen proposal. Implementasi dilakukan dengan menerapkan hasil pengolahan data pada aplikasi deteksi plagiarisme yang sudah dirancang untuk mengetahui tingkat plagiarisme dari suatu dokumen.

### 3.1.5 Pengujian

Proses pengujian menggunakan 3 skenario pada dokumen Sistem Informasi dan Biologi. Berdasarkan jumlah dokumen Sistem Informasi sebanyak 20 dokumen

dan 17 dokumen Biologi, berikut pembagian dokumen yang digunakan dalam 3 skenario seperti berikut:

1. Pengujian Skenario 1 untuk dokumen Sistem Informasi menggunakan 7 dokumen dengan 6 dokumen pembanding dan 1 dokumen *testing*. Pengujian Skenario 1 untuk dokumen Biologi menggunakan 6 dokumen dengan 5 dokumen sebagai data pembanding dan 1 dokumen sebagai dokumen *testing*.
  2. Pengujian Skenario 2 untuk dokumen Sistem Informasi menggunakan 6 dokumen dengan 5 dokumen pembanding dan 1 dokumen *testing*. Pengujian Skenario 2 untuk dokumen Biologi menggunakan 6 dokumen dengan 5 dokumen pembanding dan 1 dokumen *testing*.
  3. Pengujian Skenario 3 untuk dokumen Sistem Informasi menggunakan 7 dokumen dengan 6 dokumen pembanding dan 1 dokumen *testing*. Pengujian Skenario 3 untuk dokumen Biologi menggunakan 6 dokumen dengan 5 dokumen pembanding dan 1 dokumen *testing*.

Waktu proses pengujian dokumen juga akan dimonitor untuk mengetahui kecepatan pemrosesan pengujian indikasi plagiarisme.

### **3.2 Jadwal Penelitian**

Berikut jadwal penelitian yang dilakukan tentang sistem deteksi indikasi plagiarism menggunakan algoritma *N-Grams* dan *Document Fingerprinting*, sebagai berikut :

*Tabel 3. 1 Jadwal Penelitian*

No	Tahapan	Februari	Maret	April	Mei	Juni	Juli
1.	Studi Literatur						
2.	Pengumpulan Data						
3.	Pengolahan Data						
4.	Desain Dan Implementasi						
5.	Pengujian						
6.	Pembuatan laporan						

## **BAB 4**

# **HASIL DAN PEMBAHASAN**

## 4.1 Pengumpulan Data

Pengumpulan data yang dilakukan dengan mengumpulkan proposal skripsi untuk pengujian menggunakan proposal skripsi mahasiswa Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Ampel Surabaya angkatan 2014, dari jurusan teknologi pada program studi sistem informasi sebanyak 20 proposal, dan jurusan sains pada program studi biologi sebanyak 17 proposal.

Berikut pengumpulan data untuk program studi Sistem Informasi sebanyak 20 proposal skripsi mahasiswa pada Tabel 4.1 berikut:

Tabel 4. 1 Data Proposal skripsi Prodi Sistem Informasi

No	Nama	Judul
1	Alfian Pradana Isanto	Pengembangan Integrasi <i>E-Learning</i> Berbasis Moodle Dengan
2	Aria Dwi Azzida	Mengukur Risiko Pemborosan Perangkat ICT ( <i>Information and Communication Technology</i> ) Di Laboratorium Terintegrasi UIN Sunan Ampel Surabaya Menggunakan ISO 31000 dan ISO 50002
3	Aslichatul Millah	Pengukuran Tingkat Layanan Teknologi Informasi pada PT.XYZ Menggunakan FRAMEWORK ISO 20000-1: 2011
4	Cahya Hari Sulistia	Rancang Bangun Dan Analisis Performa Ayoumroh.Com Dengan Implementasi Basis Data Terdistribusi
5	M. Fikri Nur Akbar	Integrasi Modul Tabungan Dengan Modul Keuangan Multi Departemen Di Yayasan Pondok Pesantren Al-Hidayah Mojosari Mojokerto
6	Frista Gifti Weddiningrum	Deteksi Konten Hoax Berbahasa Indonesia Pada Media Sosial Menggunakan Metode <i>Levenshtein Distance</i>
7	Ilmi Alfianti	Sistem Pemodelan Simulasi Pelayanan Pembayaran Pajak Kendaraan Lima Tahunan Menggunakan Metode <i>Next-Event Time Advance</i>
8	Laili Nur Qomary	Rancang Bangun Sistem Pendukung Keputusan Pemilihan Program Latihan Fitnes Metode <i>Case Based Reasoning</i>
9	Niar Ismi Lafirda AS	Rancangan Master Plan Sistem Teknologi Informasi Pada Dinas Komunikasi Dan Informatika Kabupaten Nganjuk Menggunakan Metode <i>Ward And Peppard</i>

No	Nama	Judul
10	Pipit Merit Biyanti	Perencanaan Strategis Sistem Informasi Pada Universitas Islam Negeri Sunan Ampel Surabaya Menggunakan Metode <i>Ward And Peppard</i>
11	M. Abdul Aziz	Sistem Informasi Penentuan Uang Kuliah Tunggal Dengan Metode <i>Fuzzy Sugeno</i> Di UIN Sunan Ampel Surabaya
12	Sayyidatul Eka Putri R.	Evaluasi <i>Usability</i> Sistem Informasi Manajemen Kepegawaian UIN Sunan Ampel Surabaya Berbasis ISO 9241-11
13	Fachrizal Hamzah	Rancang Bangun Media Pembelajaran Studi Hadits Berbasis Android
14	Siti Nurazizah	Rancangan Prototipe <i>Mockup</i> Sistem Informasi Manajemen Pada Biro Travel UINSA
15	Tsania Saraswati	Desain Tata Kelola Teknologi Informasi Ramah Lingkungan Berbasis ITIL Versi 3
16	Ahmad Mustain Marzuki	Rancang Bangun Sistem Informasi Manajemen Sekolah dengan Server Terpusat
17	Taufiqurrochman	Rancang Bangun <i>Dashboard</i> Untuk Pemantauan Kinerja Program Studi Di Fakultas Sains Dan Teknologi
18	Lutfiana Fatmawati	Sistem Pendukung Keputusan Pemilihan Pasangan Hidup Perspektif Islam Nusantara Menggunakan Metode <i>Fuzzy Inference System</i>
19	Yudha Tri Satya	Implementasi <i>Honeypot</i> untuk Meningkatkan Keamanan server sebagai Pendekripsi dan Pencegah dari Serangan Peretas
20	Arif Iswahyudi	Rancang Bangun Sistem Prediksi Cuaca Maritim Menggunakan <i>Fuzzy Mamdani Inference System</i> di perairan Gresik-bawean

Tabel 4.1 menjelaskan daftar proposal skripsi untuk program studi Sistem Informasi dari 20 mahasiswa yang sudah melakukan sidang proposal skripsi untuk angkatan 2014. Pengumpulan data untuk program studi Biologi sebanyak 15 proposal skripsi mahasiswa seperti Tabel 4.2 berikut:

Tabel 4. 2 Data Proposal skripsi Prodi Biologi

No	Nama	Judul
1	Andita Febrianti	Pengaruh Pemberian Berbagai Dosis Ekstrak Daging Buah Kurma Ajwa ( <i>Phoenix Dactylifera</i> ) Terhadap Kadar Glukosa Darah Mencit ( <i>Mus Musculus</i> ) Bunting
2	Anis Mukaromatul Ula	Pengaruh Pemberian Ekstrak Daging Buah Kurma Ajwa terhadap Hitung Jenis Leukosit Embrio Mencit

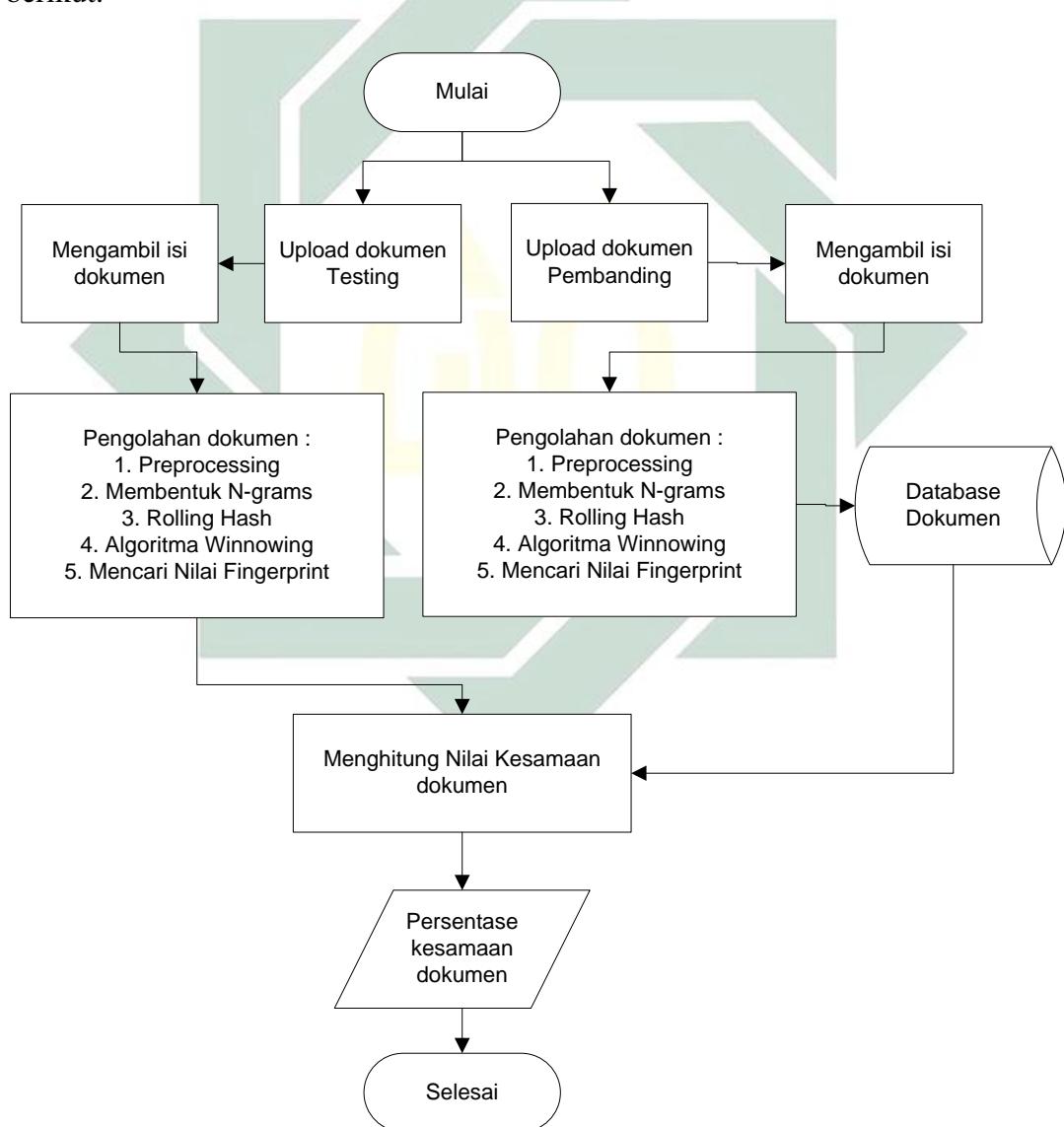
No	Nama	Judul
3	Arika Wahyuningsih	Identifikasi Varian Genetik Gajah SumaterA ( <i>Elephas maximus sumatranus</i> ) di Kawasan Hutan Ulu Masen
4	Devi Silvia	Uji Aktifitas Antifungi Ekstrak Kulit Jeruk Nipis ( <i>Citrus Aurantifolia</i> ) terhadap Jamur Candida Albicans
5	Eni Rahmawati	Uji Aktivitas Antibakteri Ekstrak Biji Kelor ( <i>Moringa Oleifera</i> ) Terhadap Bakteri Shigella Dysenteriae
6	Ika Sayyidatul Khumairoh	Uji Aktivitas Antifungi Lengkuas Merah (Alpinia Purpurata), Jahe ( <i>Zingiber Officinale</i> ) Dan Kunyit ( <i>Curcuma Longa</i> ) Terhadap Candida Albicans
7	Nurul Nahdiyah	Aktivitas Hepatoprotektif Dari Ekstrak Kurma Ruthab ( <i>Phoenix Dactylifera</i> ) Pada Histologi Hepar Mencit ( <i>Mus Musculus</i> ) Betina Yang Diinduksi Paracetamol
8	Mashita Andiana	Perbedaan Efek Pemberian Getah Tanaman Yodium ( <i>Jatropha Multifida</i> ), Jarak Pagar ( <i>Jatropha Curcas</i> ) Dan Povidone Iodine 10% Terhadap Penyembuhan Luka Sayat Pada Mencit ( <i>Mus Musculus</i> )
9	Ustatik	Pengaruh Ekstrak Buah Jambu Biji Merah ( <i>Psidium Guajava L.</i> ) Terhadap Morfologi Spermatozoa Mencit Jantan ( <i>Mus Musculus</i> ) Yang Terpapar Asap Rokok
10	Vida Nofrianinda	Morfogenesis Batang Bawah Jeruk ( <i>Citrus Limonia Osbeck</i> ) Kultivar Japansche Citroen Pada Kombinasi Media Kultur Yang Mengandung Metionin
11	Tri Jayanti	Karakterisasi Fenotip Kepuh ( <i>Sterculia Foetida L.</i> ) Di Kabupaten Lamongan
12	Nur Setianingsih	Pengaruh Pemberian Ekstrak Daging Buah Kurma Ajwa ( <i>Phoenix Dactylifera L.</i> ) Terhadap Kadar Glukosa Darah Embrio Mencit ( <i>Mus Musculus</i> )
13	Ulfa Wafirotul Bahiyyah	Pengaruh Pemberian Ekstrak Minyak Biji Kelor ( <i>Moringa Oleifera</i> ) Terhadap Kadar Glukosa Tikus ( <i>Rattus Norvegicus</i> ) Yang Diinduksi Aloksan
14	Dita Umi Rosidah Larasati	<i>Limit of Detection (LOD)</i> Fragmen DNA Pengkode Gen Sitokrom B ( <i>cyt b</i> ) pada Babi ( <i>Sus scrofa</i> )
15	Syahidatul Ulya	Pengaruh Pemberian Ekstrak Daging Buah Kurma ajwa ( <i>Phoenix dactylifera L.</i> ) Terhadap Kadar Hemoglobin Pada Mencit ( <i>Mus musculus</i> ) Bunting
16	Muhibbudin	Diversitas Odonata dan Peranannya sebagai Indikator Kualitas air di Sumber Clangap dan Sumber Mangli Desa Puncu Kecamatan Puncu Kabupaten Kediri
17	Khoirun Nihayati	Deteksi Bakteri <i>Escherichia Coli</i> O157:H7 Pada Sampel Air Di Lingkungan UIN Sunan Ampel Surabaya

Tabel 4.2 menjelaskan daftar proposal skripsi untuk program studi Biologi dari 17 mahasiswa yang sudah melakukan sidang proposal skripsi untuk angkatan 2014.

## 4.2 Proses Sistem

#### 4.2.1 Alur Proses Sistem

Secara garis besar sistem deteksi indikasi plagiarisme menggunakan algoritma *n-grams* dan *document fingerprinting* memiliki alur proses sistem seperti berikut.



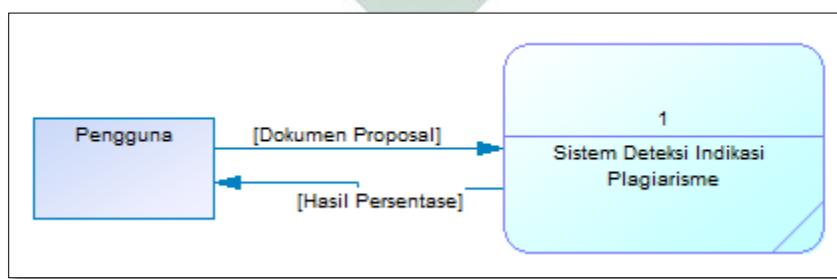
Gambar 4. 1 Alur Sistem

Berdasarkan alur proses sistem pada Gambar 4.1 tahapan mendeteksi indikasi plagiarisme pada dokumen dimulai dengan mengunggah dokumen pembanding dan mengambil isi teks dari dokumen, kemudian dilakukan proses preprocessing, pembentukan *n-grams*, merubah rangkaian *n-grams* menjadi *hash*, pembentukan *windowing* dari nilai *hash* dan mencari nilai *fingerprint* dokumen dari nilai hash terkecil berdasarkan rangkaian nilai *hash* yang dibentuk dalam *windowing* lalu disimpan di database. *fingerprint* yang disimpan dalam database digunakan sebagai pembanding dengan dokumen testing. Dokumen yang diunggah harus berekstensi .docx.

Langkah selanjutnya mengunggah dokumen *testing* yang merupakan dokumen yang akan dibandingkan dengan dokumen pembanding. Selanjutnya dilakukan proses preprocessing sampai mendapatkan nilai *fingerprint* dokumen. Setelah didapatkan nilai *fingerprint* selanjutnya membandingkan antara *fingerprint* dokumen pembanding yang sudah disimpan dalam database dengan *fingerprint* dokumen *testing* sehingga didapatkan prosentase kesamaan dokumen antara dokumen *testing* dan dokumen pembanding.

#### 4.2.2 Gambaran Pemodelan *Data Flow Diagram*

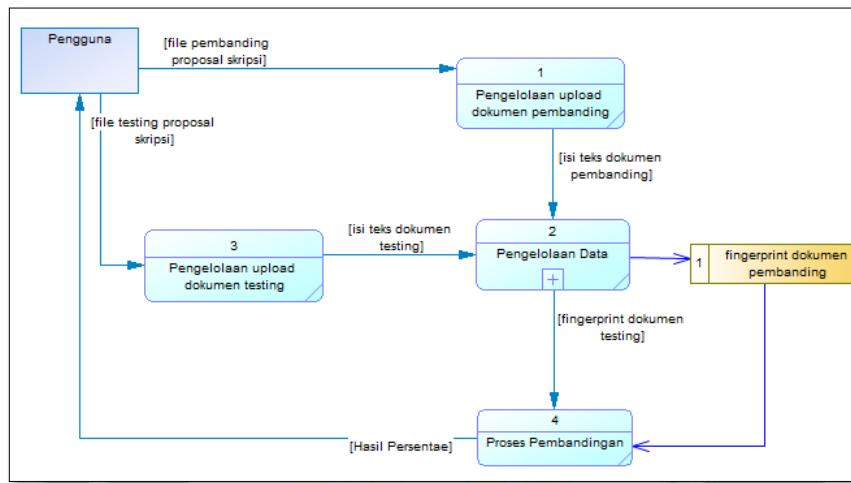
Pada tahapan ini menjelaskan gambaran alir data di sistem yang dibuat. Penggambaran alir data pada sistem digambarkan dalam *Data Flow Diagram* mulai dari *context diagram* sampai diagram level 1. Berikut gambaran dari Data Flow Diagram untuk *context diagram* seperti pada gambar 4.5 seperti berikut :



Gambar 4. 2 DFD Context diagram

Berdasarkan Gambar 4.2 dapat dijelaskan bahwa pengguna memberikan data berupa dokumen proposal pada Sistem Deteksi Indikasi Plagiarisme dan sebaliknya sistem deteksi indikasi plariarisme memberikan hasil prosentase pada

pengguna. Berdasarkan gambar diatas dapat diketahui bahwa pengguna memasukkan dokumen pada sistem, lalu sistem memberikan hasil prosentase kesamaan dokumen. Selanjutnya gambaran dari alir data sistem secara keseluruhan pada diagram level 0 seperti pada Gambar 4.3 berikut:



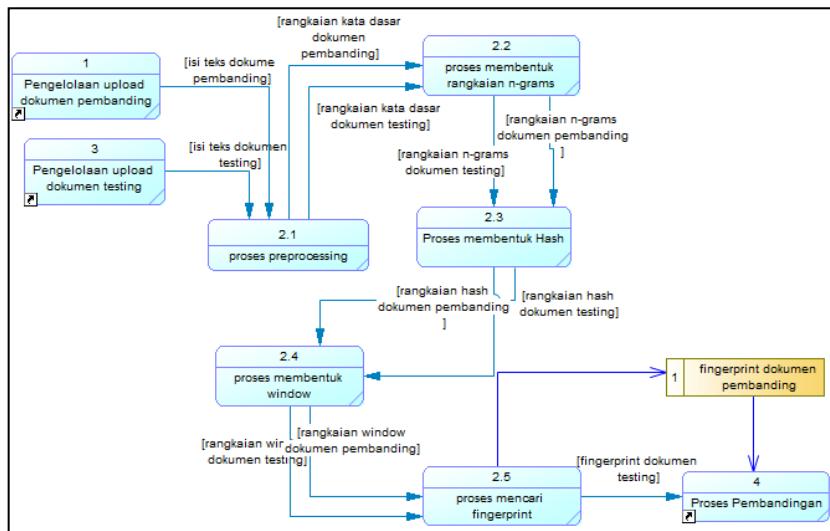
Gambar 4. 3 DFD Level 0

Berdasarkan Gambar 4.3 dapat dijelaskan bahwa terdapat proses-proses diantaranya :

1. Pengolahan upload dokumen pembanding : Proses ini menggambarkan fungsi unggah dokumen pembanding, yang berisi pemrograman untuk mengunggah dokumen.
  2. Pengolahan Data : Proses ini menggambarkan fungsi dalam memproses data yang berisi pemrograman dalam mengolah teks sampai mendapatkan *fingerprint* dokumen.
  3. Pengelolaan Upload dokumen *Testing* : Proses ini menggambarkan fungsi unggah dokumen *testing*, yang berisi pemrograman untuk mengunggah dokumen.
  4. Proses Pembandingan : Proses ini menggambarkan fungsi yang berisi pemrograman dalam membandingkan *fingerprint* dokumen pembanding dari database yang menyimpan *fingerprint* dokumen pembanding dan *fingerprint* dokumen *testing*.

Pada gambar diatas juga menggambarkan bahwa sistem membutuhkan database untuk menyimpan *fingerprint* dokumen pembanding. Selanjutnya menggambarkan

penjelasan proses-proses yang ada pada proses pengolahan data dan di gambarkan dalam diagram level 1, seperti pada Gambar 4.4 berikut:



Gambar 4. 4 DFD Level 1

Berdasarkan Gambar 4.4 dapat dijelaskan bahwa terdapat proses-proses diantaranya :

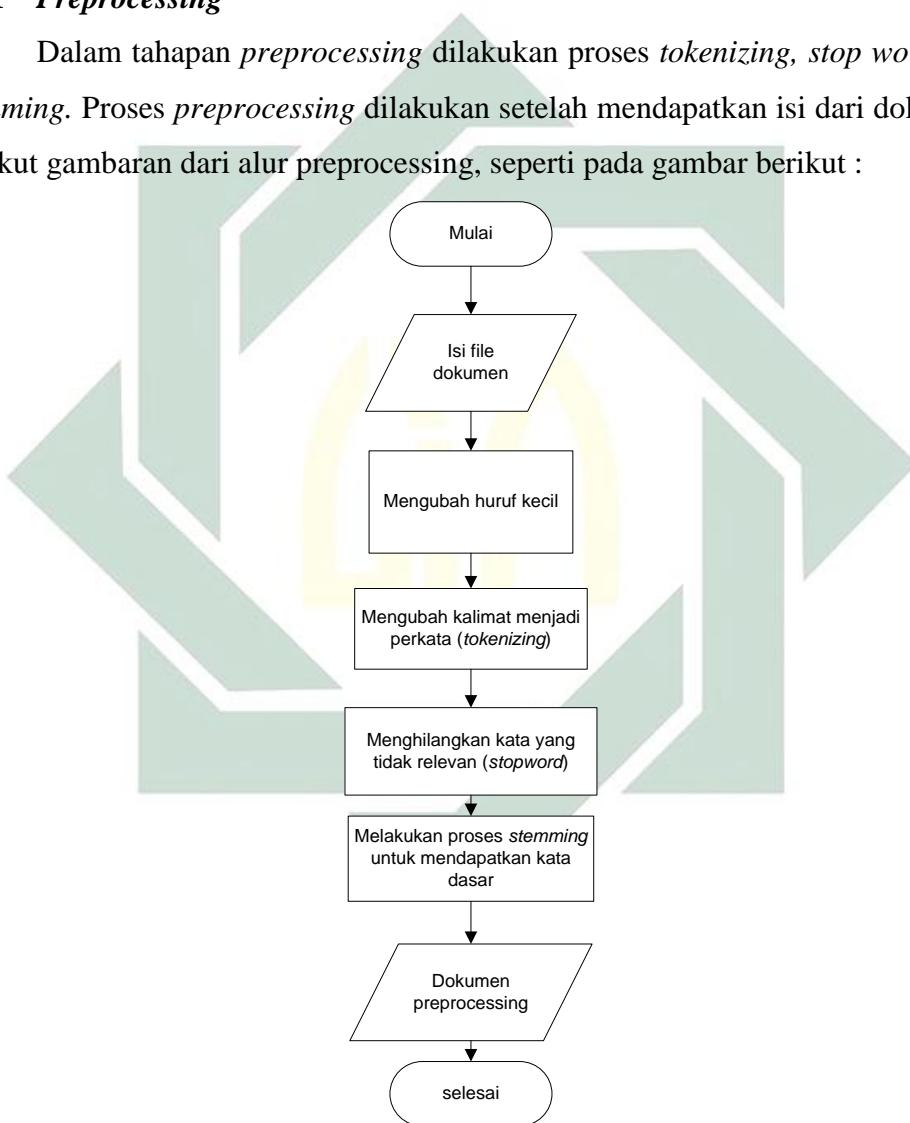
1. 2.1 Proses Preprocessing : Proses ini menggambarkan fungsi preprocessing dalam melakukan pemrograman sistem, berisi fungsi *tokenizing*, *stopword*, *stemming*.
  2. 2.2 Proses membentuk rangkaian *n-grams* : Proses ini menggambarkan fungsi membentuk rangkaian *n-grams* dalam memprogram sistem. Hasil dari fungsi ini yaitu rangkaian *n-grams*.
  3. 2.3 Proses membentuk *Hash* : Proses ini menggambarkan fungsi merubah rangkaian *n-grams* menjadi nilai *hash* dalam memprogram sistem. Hasil dari fungsi ini yaitu rangkaian *hash* dari rangkaian *n-grams*.
  4. 2.4 Proses membentuk *window* : Proses ini menggambarkan fungsi merubah rangkaian *hash* menjadi rangkaian *window* dalam memprogram sistem. Hasil dari fungsi ini yaitu rangkaian *hash* dalam sktruktur *window*.
  5. 2.5 Proses mencari *fingerprint* : Proses ini menggambarkan fungsi mencari nilai *hash* terkecil pada rangkaian *window* dalam memprogram sistem. Hasil dari fungsi ini yaitu nilai *fingerprint* dokumen.

### 4.3 Pengolahan Data

Pengolahan data dilakukan untuk menyelesaikan permasalahan yang akan diteliti dan diterapkan dalam sistem, berikut langkah-langkah yang dilakukan dalam mendeteksi plagiarisme menggunakan algoritma *n-gram* dan *document fingerprinting*, sebagai berikut:

### 4.3.1 *Preprocessing*

Dalam tahapan *preprocessing* dilakukan proses *tokenizing*, *stop word*, dan *stemming*. Proses *preprocessing* dilakukan setelah mendapatkan isi dari dokumen. Berikut gambaran dari alur *preprocessing*, seperti pada gambar berikut :



Gambar 4. 5 Alur Preprocessing

Gambar 4.5 menjelaskan alur *preprocessing* pada sistem deteksi indikasi plagiarisme menggunakan algoritma *n-grams* dan *document fingerprinting* dimulai dari mengambil teks isi file dokumen baik dokumen pembanding maupun dokumen

*testing* kemudian melakukan proses *preprocessing* (*tokenizing*, *stopword*, dan *stemming*). Tujuan dari *preprocessing* adalah mendapatkan kata dasar untuk pengecekan kesamaan dokumen.

Proses *preprocessing* dimulai dari isi file dokumen dirubah menjadi huruf kecil semua kemudian dilakukan *tokenizing* yaitu proses merubah isi file dokumen yang terdiri dari banyak kalimat menjadi perkata. Setelah itu menghilangkan kata yang tidak relevan seperti (dan, pada, di, dan lain-lain). Kemudian melakukan *stemming* untuk merubah kata dari hasil *stopword* menjadi kata dasar. Kata dasar yang didapatkan dari proses *stemming* digunakan untuk proses selanjutnya yaitu merubah menjadi rangkaian *n-gram*.

a. *Tokenizing*

Proses *Tokenizing* dilakukan untuk merubah dari bentuk kalimat yang tersusun dalam paragraf-paragraf menjadi bentuk perkata. Dalam pembuatan sistem penulis menggunakan bantuan *Library* dalam menjalankan fungsi proses tokenizing, dimana menggunakan *Library* Sastrawi untuk pemrograman PHP. Berikut sintaks yang digunakan untuk memanggil *Library* Tokenizing.

```
//tokenizing

$tokenizerFactory = new \Sastrawi\Tokenizer\TokenizerFactory();
$tokenizer = $tokenizerFactory->createDefaultTokenizer();

$tokens1 = $tokenizer->tokenize($teks1);
```

Berdasarkan sintaks diatas, `$tokenizerFactory` berfungsi untuk mendeklarasikan file yang akan dibaca pada saat proses *tokenizing* yang ada pada library sastrawi, yang kemudian `$tokenizer` untuk mengambil fungsi `createDefaultTokenizer()` yang ada pada file yang sudah dideklarasikan di `$tokenizerFactory`. Selanjutnya memproses *tokenizing* sesuai fungsi yang sudah dijelaskan sebelumnya dan didapatkan rangkaian kata, lalu disimpan dalam `$tokens1`.

b. *Stopword*

Proses *Stopword* dilakukan untuk menghilangkan kata-kata yang tidak relevan. Dalam pembuatan sistem penulis menggunakan bantuan *Library* dalam

menjalankan fungsi proses *stopword*, dimana menggunakan *Library Sastrawi*. Kata-kata Bahasa Indonesia yang tidak relevan yang ada dalam *Library Sastrawi* terdapat seperti berikut:

```
public function getStopWords() {
    return array(
        'yang', 'untuk', 'pada', 'ke', 'para', 'namun', 'menurut',
        'antara', 'dia', 'dua', 'ia', 'seperti', 'jika', 'jika', 'sehingga',
        'kembali', 'dan', 'tidak', 'ini', 'karena', 'kepada', 'oleh', 'saat',
        'harus', 'sementara', 'setelah', 'belum', 'kami', 'sekitar', 'bagi',
        'serta', 'di', 'dari', 'telah', 'sebagai', 'masih', 'hal', 'ketika',
        'adalah', 'itu', 'dalam', 'bisa', 'bahwa', 'atau', 'hanya', 'kita', 'dengan',
        'akan', 'juga', 'ada', 'mereka', 'sudah', 'terhadap', 'secara', 'agar',
        'lain', 'anda', 'begitu', 'mengapa', 'kenapa', 'yaitu', 'yakni', 'daripada',
        'itulah', 'lagi', 'maka', 'tentang', 'demi', 'dimana', 'kemana', 'pula',
        'ambil', 'sebelum', 'sesudah', 'supaya', 'guna', 'kah', 'pun', 'sampai',
        'sedangkan', 'selagi', 'sementara', 'tetapi', 'apakah', 'kecuali', 'sebab',
        'selain', 'seolah', 'seraya', 'seterusnya', 'tanpa', 'agak', 'boleh',
        'dapat', 'dsb', 'dst', 'dll', 'dahulu', 'dulunya', 'anu', 'demikian', 'tapi',
        'ingin', 'juga', 'nggak', 'mari', 'nanti', 'melainkan', 'oh', 'ok',
        'seharusnya', 'sebetulnya', 'setiap', 'setidaknya', 'sesuatu', 'pasti',
        'saja', 'toh', 'ya', 'walau', 'tolong', 'tentu', 'amat', 'apalagi',
        'bagaimanapun',
    );
}
```

Untuk menjalankan proses *stopword remover* menggunakan sintaks untuk memanggil *library Stopword*, seperti gambar berikut.

```
//stopword
$StopWordRemover = new \Sastrawi\StopWordRemover\StopWordRemoverFactory();
$stopword = $StopWordRemover->createStopWordRemover();
$hasil      = $stopword->remove($teks1);
```

Berdasarkan sintaks diatas, sama seperti proses *tokenizing*, *stopword remover* juga menggunakan library sastrawi dengan bahasa pemrograman PHP. Dimulai dari sintaks `$StopWordRemover` berfungsi untuk mendeklarasikan file yang akan dibaca pada saat proses *stopword remover* yang ada pada library sastrawi, yang kemudian `$stopword` untuk mengambil fungsi `createStopWordRemover()` yang ada pada file yang sudah dideklarasikan di `$StopWordRemover`. Selanjutnya memproses *stopword remover* sesuai fungsi yang sudah dijelaskan sebelumnya untuk menghilangkan kata yang tidak relevan

dan didapatkan rangkaian kata tanpa kata yang tidak relevan, lalu disimpan dalam \$hasil.

### c. Stemming

Proses *Stemming* dilakukan untuk merubah kata yang mempunyai imbuhan menjadi kata dasar, berdasarkan sistem hasil dari *stopword remover* kemudian dilakukan *stemming* sehingga didapatkan kata dasar yang akan digunakan dalam mendeteksi kesamaan kata antar dokumen. Berikut contoh proses *stemming* seperti berikut :

"saya", "menjadi", "mahasiswa", "sistem", "informasi", "surabaya".

Setelah dilakukan *stemming* berubah menjadi :

"saya", "jadi", "mahasiswa", "sistem", "informasi", "surabaya".

Dalam pembuatan sistem penulis juga menggunakan bantuan *library* dalam menjalankan fungsi proses *stemming*, dimana menggunakan *library* Sastrawi. Berikut sintaks yang digunakan untuk memanggil *Library stemming*.

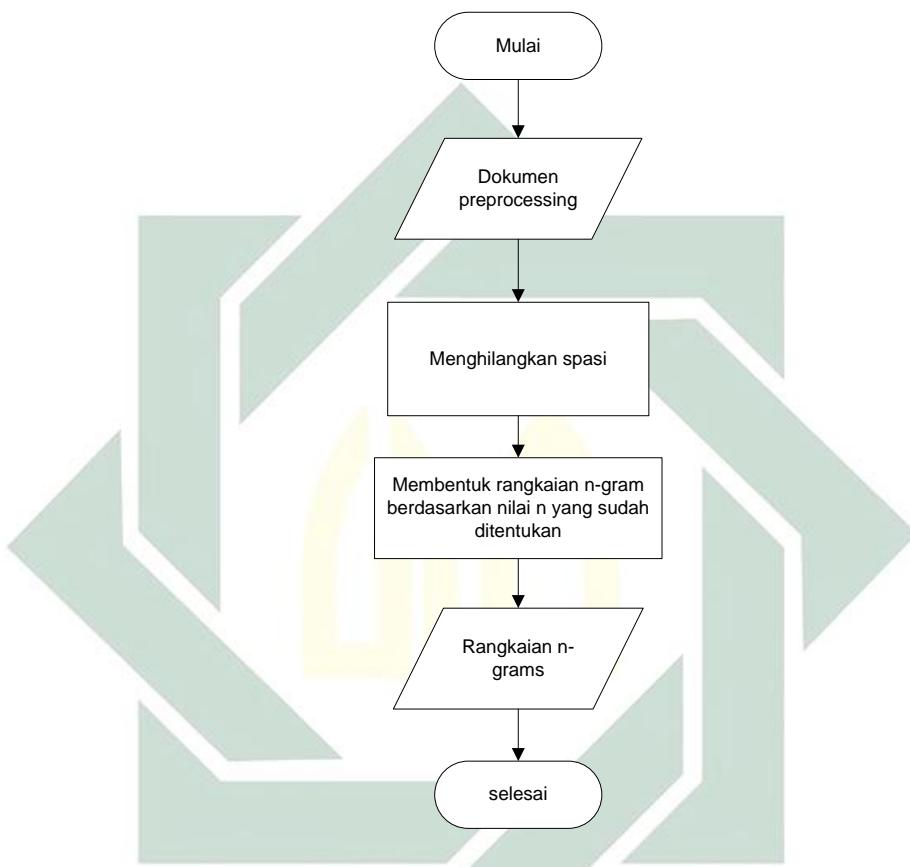
```
//stemming
$stemmerFactory = new \Sastrawi\Stemmer\StemmerFactory();
$stemmer      = $stemmerFactory->createStemmer();

$stemming     = $stemmer->stem($hasil);
```

Berdasarkan sintaks diatas, sama seperti proses *tokenizing* dan *stopword remover*, *stemming* juga menggunakan library sastrawi dengan bahasa pemrograman PHP. Dimulai dari sintaks `$StemmerFactory` berfungsi untuk mendeklarasikan file yang akan dibaca pada saat proses *stemming* yang ada pada library sastrawi, yang kemudian `$stemmer` untuk mengambil fungsi `createStemmer()` yang ada pada file yang sudah dideklarasikan di `$StemmerFactory`. Selanjutnya memproses *stemming* sesuai fungsi yang sudah dijelaskan sebelumnya untuk mendapatkan kata dasar dengan menghilangkan imbuhan dalam kata sehingga didapatkan kata dasar, lalu disimpan dalam `$stemming`.

### 4.3.2 N-Grams

Dalam tahapan ini dilakukan perubahan dari bentuk kata menjadi rangkaian *n-grams* dengan urutan sesuai struktur kalimat. Berikut gambaran alur pembentukan rangkaian *n-grams* seperti pada gambar dibawah ini :



Gambar 4. 6 Alur Rangkaian N-Grams

Gambar 4.6 menjelaskan alur pembentukan rangkaian *n-grams* pada sistem deteksi indikasi plagiarisme menggunakan algoritma *n-grams* dan *document fingerprinting* dimulai dari mengambil teks hasil proses *preprocessing* kemudian menghilangkan spasi sehingga didapatkan rangkaian kata tanpa spasi, selanjutnya membentuk rangkaian *n-grams* berdasarkan panjang *n*.

Pada sistem ini panjang  $n$  menggunakan  $n=3$  sampai  $n=7$  sehingga dapat diketahui nilai  $n$  yang lebih efektif untuk diterapkan dalam mendekripsi kesamaan kata pada dokumen. Hasil dari alur ini didapatkan rangkaian  $n$ -grams. Dalam penerapan algoritma  $n$ -grams dalam sistem menggunakan nilai  $n$  mulai dari 3

sampai 7 untuk mendapatkan nilai  $n$  yang paling efektif dalam mendeteksi kesamaan kata pada dokumen.

Berikut sintaks fungsi *n-grams* yang digunakan dalam mendeteksi kesamaan kata dalam dokumen menggunakan bahasa pemrograman *PHP*. Hal yang dilakukan sebelum mulai mendeteksi yaitu menghilangkan semua spasi, sehingga tersisa barisan kata. Untuk  $n = 3$  berikut sintaks yang digunakan seperti berikut.

```

public function n_gram($n_gram){
    //menghilangkan spasi
    $hilangspasi = $n_gram;
    $hilangspasi = str_replace(" ", "", $hilangspasi);

    //ngram
    $word = $hilangspasi;

    $ngr = array();
    $lengt = strlen($word);
    for($i = 0; $i < $lengt; $i++) {
        if($i > (5 - 2)) {
            $ng = '';
            for($j = 5-1; $j >= 0; $j--) {
                $ng .= $word[$i-$j];
            }
            $ngr[] = $ng;
        }
    }
    return $ngr;
}

```

Pada sintaks diatas, proses membentuk rangkaian *n-grams* dimulai dari menghilangkan spasi pada hasil *preprocessing*, sehingga didapatkan rangkaian kata yang saling berhubungan tanpa adanya spasi. Berikut contoh menghilangkan spasi pada rangkaian kata, seperti berikut:

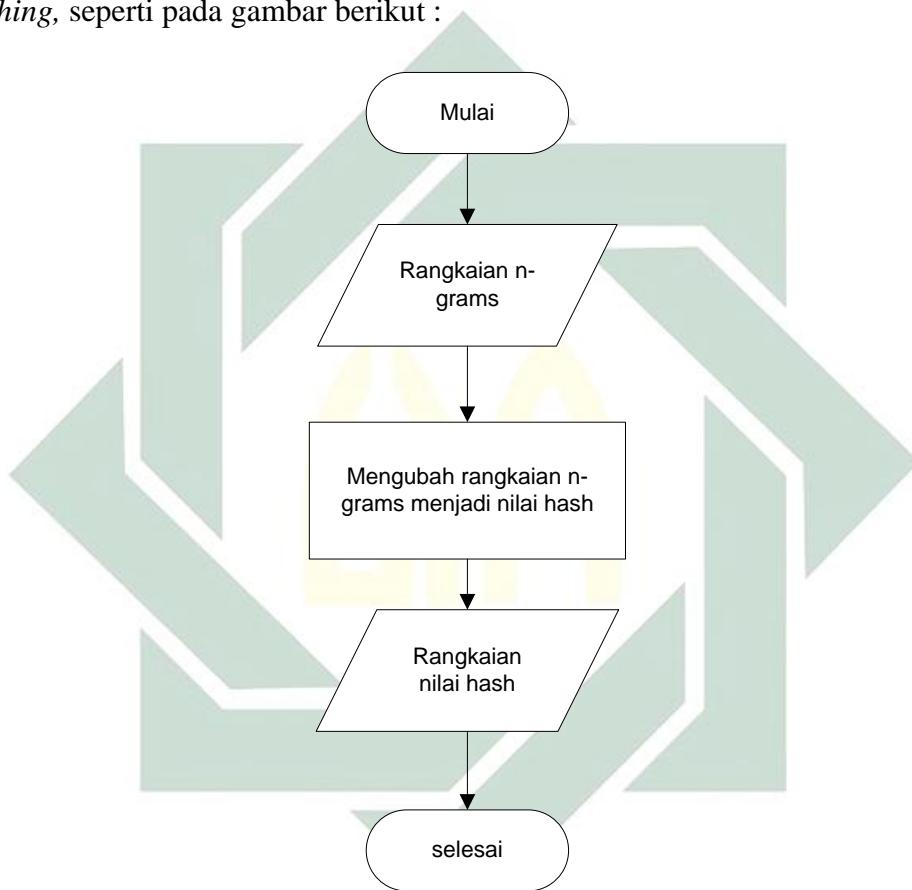
"sayajadimahasiswasisteminformasisurabaya"

Selanjutnya dilakukan proses *n-grams*, dimulai dari menghitung jumlah karakter pada rangkaian kata yang sudah dihilangkan spasinya dengan perintah *STRLEN*. Pada sintaks diatas juga menggunakan fungsi *FOR* untuk *looping* pembentukan rangkaian *n-grams*. Untuk  $n = 4$  sampai  $n = 7$  sintaks yang digunakan

sama hanya mengganti nilai  $n$  saja. Sintaks untuk fungsi  $n$ -grams dalam sistem dimasukkan dalam *function* sehingga tidak memanggil fungsi berkali-kali.

### 4.3.3 Hashing

Tahapan selanjutnya setelah mendapatkan rangkaian *n-grams* adalah merubah rangkaian *n-grams* menjadi nilai *hash* (*Hashing*). Berikut alur dari proses *Hashing*, seperti pada gambar berikut :



#### Gambar 4.7 Alur Hashing

Gambar 4.7 menjelaskan alur *hashing* pada sistem deteksi indikasi plagiarisme menggunakan algoritma *n-grams* dan *document fingerprinting* menggunakan rumus *rolling hash*, dimulai dari mengambil rangkaian *n-grams* selanjutnya mengubahnya menjadi rangkaian hash sesuai rumus *rolling hash* dengan nilai bilangan prima ditentukan langsung dan mencari bilangan ASCII setiap karakter. Dari langkah diatas didapatkan nilai *hash* dari setiap rangkaian *n-grams*.

Untuk mendapatkan nilai hash menggunakan rumus *Rolling Hash* dengan menentukan nilai bilangan prima = 3. Berikut rumus yang digunakan dan diterapkan dalam bahasa pemrograman *PHP*, seperti gambar berikut.

```
public function hashing($gram) {
    $hashing = array();
    for($x=0; $x < count($gram); $x++) {
        $panjang = strlen($gram[$x]);
        $hash=0;
        for ($i = 0; $i < $panjang; $i++) {
            $hash += ord(substr($gram[$x], $i, 1))*pow(3, ($panjang-($i+1)));
        }
        $hashing[] = $hash;
    }
    return $hashing;
}
```

Sintaks untuk fungsi *Hashing* dalam sistem juga dimasukkan dalam *function* sehingga tidak memanggil fungsi berkali-kali.

Berikut contoh mengubah rangkain *n-grams* menjadi nilai *hash* dengan rumus *rolling hash*, dengan rangkaian *n-grams* berikut :

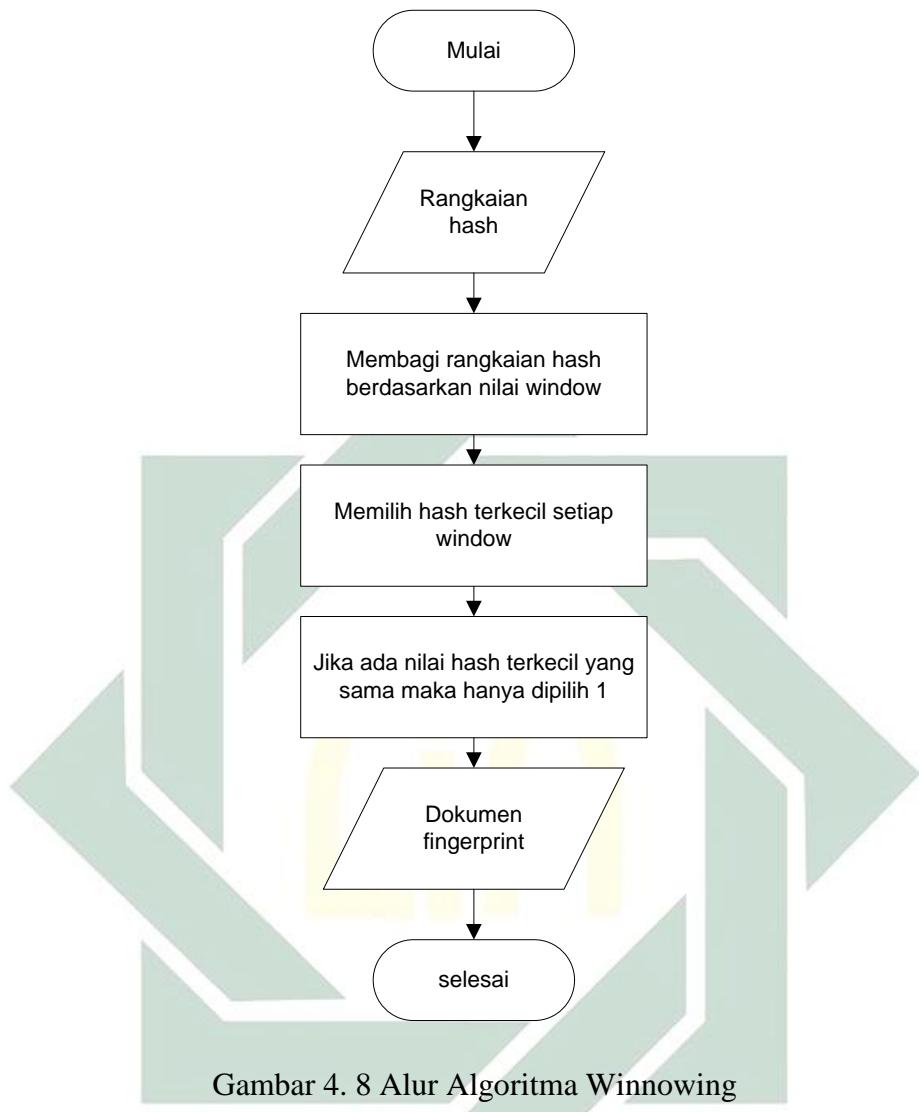
sed edi dia iaa aai air

ketika dirubah menjadi nilai *hash* dengan bilangan prima = 1 menjadi rangkaian seperti berikut :

431 407 402 404 396 413

#### 4.3.4 Algoritma Winnowing

Tahapan ini dilakukan proses perubahan dari bentuk rangkaian *n-grams* yang sudah menjadi nilai *hash* untuk menentukan *fingerprinting* dengan menentukan rangkaian *window* (*w*) kemudian mengambil nilai hash terkecil dari setiap *window*. *Fingerprint* hasil dari algoritma *winnowing* dapat digunakan untuk menentukan tingkat kesamaan kata antar dokumen. Berikut gambaran dari alur proses algoritma *winnowing* seperti pada gambar berikut :



Gambar 4. 8 Alur Algoritma Winnowing

Gambar 4.8 dijelaskan alur algoritma *winnowing* pada sistem deteksi indikasi plagiarisme menggunakan algoritma *n-grams* dan *document fingerprinting* dimulai dari mengambil rangkaian *hash* yang sudah di dapatkan dari proses *Hashing*, kemudian membagi rangkaian *hash* berdasarkan nilai *window* ( $w$ ). Kemudian dari rangkaian *window* dipilih nilai *hash* terkecil disetiap *window*. Nilai *hash* terkecil yang sudah di dapat kemudian di cek apabila ada yang sama maka hanya di pilih satu saja. Rangkaian nilai *hash* terkecil tersebut adalah nilai *fingerprint* dari dokumen.

Tujuan dari alur proses algoritma *winnowing* adalah mendapatkan nilai *fingerprint* dari isi dokumen yang nantinya akan dibandingkan dengan nilai *fingerprint* dari dokumen lain. Dengan menggunakan *fingerprint* maka bisa

diketahui kesamaan antar dokumen meskipun letak kata diurutan yang berbeda. Berikut rumus *windowing* yang diterapkan dalam bahasa pemrograman PHP seperti gambar berikut.

```

public function windowing($hash) {
    //windowing
    $window = array();
    $pjng = count($hash);
    $x = 0;
    for($c = 0; $c < $pjng; $c++) {
        if($c > (5 - 2)) {
            $window[$x] = array();
            $y = 0;
            for($d = 5-1; $d >= 0; $d--) {
                $window[$x][$y] = $hash[$c-$d];
                $y++;
            }
            $x++;
        }
    }
    return $window;
}

```

Pada sistem ini menggunakan nilai  $w = 5$ , berikut contoh rangkaian dari *window* berdasarkan hasil yang didapatkan dari nilai *hash* rangkaian *n-grams* dengan  $w = 2$ , seperti berikut :

[431 407] [407 402] [402 404] [404 396] [396 413]

Dari contoh diatas, dapat didapatkan 5 rangkaian *window*. Sintaks untuk fungsi *winnowing* dalam sistem juga dimasukkan dalam *function* sehingga tidak memanggil fungsi berkali-kali. Setelah mendapatkan rangkaian *window* selanjutnya mencari nilai *fingerprint* dari setiap *window*, berikut sintaks yang digunakan dalam mendapatkan *fingerprint* dokumen.

```
public function fingerprint($window) {
    $fingers = array();
    for($e = 0; $e < count($window); $e++) {
        $min = $window[$e][0];
        for($f = 1; $f < 4; $f++) {
            if($min > $window[$e][$f]) {
                $min = $window[$e][$f];
            }
        }
    }
}
```

```
    $fingers[] = $min;  
}  
  
return $fingers;  
}
```

Nilai *fingerprint* yang sudah didapatkan dari setiap dokumen yang akan dijadikan sebagai bahan perbandingan untuk mendapatkan tingkat kesamaan antar dokumen. *Fingerprint* dipilih dari nilai *hash* terkecil dari rangkaian *window* yang didapat sebelumnya, berikut contoh memilih *fingerprint* dari nilai *hash* dalam *window*, seperti berikut :

[431 407] [407 402] [402 404] [404 396] [396 413]

Dari contoh diatas, didapatkan nilai *fingerprint* yaitu :

[407, 402, 396]

#### **4.4 Desain dan Implementasi**

#### 4.4.1 Desain Basis Data

Sistem deteksi indikasi plagiarisme yang dibuat menggunakan basis data untuk menyimpan nilai *fingerprint* dokumen pembanding. Tujuan dari penyimpanan nilai *fingerprint* dokumen pembanding untuk digunakan dalam perulangan perbandingan dengan *fingerprint* dokumen *testing* untuk mendapatkan prosentase kesamaan kata.

Berikut gambaran atribut dari basis data yang digunakan, seperti pada tabel 4.3 berikut :

Tabel 4. 3 Atribut Basis Data

No.	Nama	Tipe Data
1	id_dok (Primary Key)	int(11) Auto_Increment
2	nama_dok	varchar(50)
3	type_dok	varchar(50)
4	size_dok	11
5	fingerprint_dok	varchar(10000)
6	tgl_upload	timestamp

Berdasarkan Tabel 4.3 dijelaskan atribut-atribut basis data yang digunakan, berikut penjelasan dari masing-masing atribut, diantaranya:

1. Id\_dok : untuk ID setiap dokumen dengan penomoran secara otomatis.
  2. Nama\_dok : untuk nama dokumen, contoh "Proposal Skripsi 01.docx".

3. Type\_dok : untuk tipe dokumen contoh tipe dokumen .docx.
  4. Size\_dok : untuk ukuran dokumen dengan satuan *Kilobyte* (Kb).
  5. Fingerprint\_dok : untuk menyimpan *fingerprint* dokumen.
  6. Tgl\_upload : untuk menyimpan tanggal dan waktu upload yang secara otomatis disesuaikan dengan waktu asli.

#### **4.4.2 Implementasi**

#### **4.4.2.1 Spesifikasi sistem Perangkat Lunak**

Sistem Deteksi Indikasi Plagiarisme ini dikembangkan dengan menggunakan beberapa perangkat lunak, sebagai berikut :

1. Sistem Operasi Windows 10 64bit
  2. XAMPP 5.6.11 2013
  3. Atom Version 1.28.0 X64
  4. HeidiSQL
  5. Google Chrome Version 67.0.3396.9 64bit

#### **4.4.2.2 Spesifikasi Perangkat Keras**

Perancangan dan pengembangan Sistem Deteksi Indikasi Plagiarisme ini menggunakan laptop penulis dengan spesifikasi sebagai berikut.

Tabel 4. 4 Spesifikasi Perangkat Keras

Perangkat Keras	Spesifikasi
Processor	AMD E1-6010
RAM	4 Gb
Hardisk	500 Gb

Berdasarkan Tabel 4.4 dijelaskan bahwa spesifikasi perangkat keras yang digunakan dalam pembuatan sistem menggunakan *Processor*, RAM, dan Hardisk seperti yang tercantum pada Tabel 4.4. Sistem dapat dijalankan pada perangkat keras dengan spesifikasi lebih rendah dibandingkan yang tercantum pada Tabel 4.4.

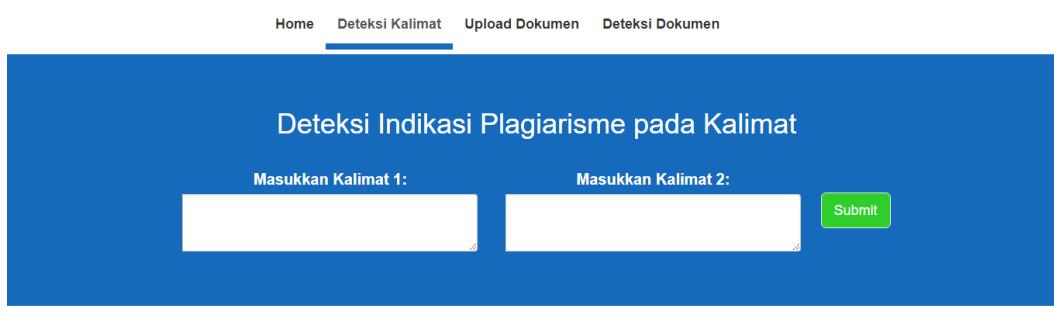
#### **4.4.2.3 Tampilan Antarmuka (User Interfaces)**

Sistem deteksi indikasi plagiarisme didukung dengan tampilan antarmuka untuk membantu dalam menjalankan fungsi sistem. Halaman antar muka terdapat menu-menu, diantaranya halaman Beranda, halaman Deteksi Kalimat, halaman Unggah Dokumen, dan halaman Deteksi Indikasi Plagiarisme. Berikut tampilan untuk halaman Beranda untuk *User Interfaces* terdapat pada gambar 4.9 berikut.



Gambar 4, 9 Halaman Utama

Selanjutnya untuk *User Interfaces* halaman Deteksi Kalimat, pada halaman ini digunakan untuk menguji tingkat kesamaan kata pada kalimat, dengan memasukkan dua kalimat. Ketika memasukkan dua kalimat masing-masing pada kalimat 1 dan kalimat 2, selanjutnya menekan tombol 'Submit' maka sistem akan memproses dan hasil prosentase akan muncul. Berikut tampilan dari halaman Deteksi Kalimat pada Gambar 4.10:



## Deteksi Indikasi Plagiarisme pada Kalimat

Gambar 4. 10 Halaman Deteksi Kalimat  
Pada halaman menu Unggah Dokumen, ditampilkan form untuk mengunggah dokumen pembanding yang akan disimpan dalam database dan digunakan sebagai dokumen pembanding untuk dokumen pengujian (*testing*). Hasil dari mengunggah dokumen pembanding adalah untuk mendapatkan *fingerprint* dokumen dan disimpan dalam database. Pada halaman ini juga di tampilkan dokumen-dokumen pembanding yang sudah diunggah dalam bentuk tabel. Berikut tampilan dari halaman Unggah Dokumen pada Gambar 4.11 berikut:

[Home](#)   [Deteksi Kalimat](#)   [Upload Dokumen](#)   [Deteksi Dokumen](#)

## Unggah Dokumen Pembanding

No.	Dokumen yang Telah di Unggah	Ukuran Dokumen yang Telah di Unggah (Kb)
1	PROPOSAL SKRIPSI SI- Aslitchatul Millah BAB I SK1.d	30280
2	PROPOSAL SKRIPSI SI- Frista Gifti Weddiningrum BAB	26790
3	PROPOSAL SKRIPSI SI- Pipit Merit Biyanti BAB I SK1	37070

Gambar 4. 11 Halaman Unggah Dokumen Pembanding  
Pada halaman menu Deteksi Dokumen, ditampilkan form untuk  
unggah dokumen pengujian (*testing*), yaitu dokumen yang akan diuji untuk  
getahui prosentase tingkat indikasi plagiarisme. Berikut tampilan dari halaman  
ksi dokumen pada Gambar 4.12 berikut :

Home Deteksi Kalimat Upload Dokumen Deteksi Dokumen

Gambar 4. 12 Halaman Deteksi Dokumen

Setelah dokumen diunggah, sistem akan memproses untuk mendapatkan *fingerprint* dokumen kemudian dibandingkan dengan *fingerprint* dokumen pembanding dari database sehingga didapatkan hasil prosentase tingkat kesamaan antar dokumen.

## 4.5 Pengujian

#### **4.5.1 Pengujian Sistem**

Pengujian deteksi indikasi plagiarisme menggunakan sistem dilakukan berdasarkan skenario yang sudah dibuat. Berikut daftar skenario yang sudah dibuat sebagai berikut:

1. Skenario 1 : Pengujian untuk dokumen BAB I yang tidak sama.
  2. Skenario 2 : Pengujian untuk dokumen BAB I yang 100% sama.
  3. Skenario 3 : Pengujian untuk dokumen BAB I yang sebagian sama.

Berdasarkan daftar skenario diatas, berikut penjelasan dari masing-masing skenario, sebagai berikut:

#### **4.5.1.1 Skenario 1**

Pada skenario 1 dimulai dengan menentukan data testing dan data pembanding menggunakan BAB I dari masing-masing dokumen. Pada skenario 1 pengujian antara program studi sistem informasi dan program studi biologi diuji secara terpisah. Untuk dokumen *testing* skenario 1, dilakukan skenario dengan menyalin 1 paragraf dari masing-masing dokumen pembanding ke dokumen *testing*, didimana juga digunakan sebagai pembuktian bahwa sistem bisa mendeteksi indikasi plagiasi dengan dokumen yang tidak sama atau memiliki sedikit kesamaan antara dokumen *testing* dan dokumen pembanding.

#### **4.5.1.1.1 Dokumen Program Studi Sistem Informasi**

Untuk program studi sistem informasi, berikut daftar data *testing* dan data pembanding yang digunakan.

Tabel 4. 5 skenario 1 dokumen pembanding sistem informasi

<b>Nama Dokumen</b>	<b>Ukuran File</b>
PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	29kb

Berdasarkan Tabel 4.5 dapat dijelaskan dokumen *testing* untuk skenario 1 menggunakan dokumen proposal M. Fikri Nur akbar bab I. Untuk dokumen pembanding Skenario 1 menggunakan 6 dokumen proposal skripsi prodi sistem informasi. Berikut daftar dokumen pembanding prodi sistem informasi, sebagai berikut :

Tabel 4. 6 skenario 1 dokumen testing sistem informasi

<b>Dokumen Pembanding</b>		
<b>No.</b>	<b>Nama Dokumen</b>	<b>Ukuran File</b>
1	PROPOSAL SKRIPSI SI- Aslichatul Millah BAB I SK1.docx	30kb
2	PROPOSAL SKRIPSI SI- Frista Gifti Weddiningrum BAB I SK1.docx	27kb
3	PROPOSAL SKRIPSI SI- Pipit Merit Biyanti BAB I SK1.docx	37kb
4	PROPOSAL SKRIPSI SI- Arif Iswahyudi BAB I SK1.docx	25kb
5	PROPOSAL SKRIPSI SI- Fachrizal Hamzah BAB I SK1.docx	19kb
6	PROPOSAL SKRIPSI SI- M. Abdul Aziz BAB I SK1.docx	41kb

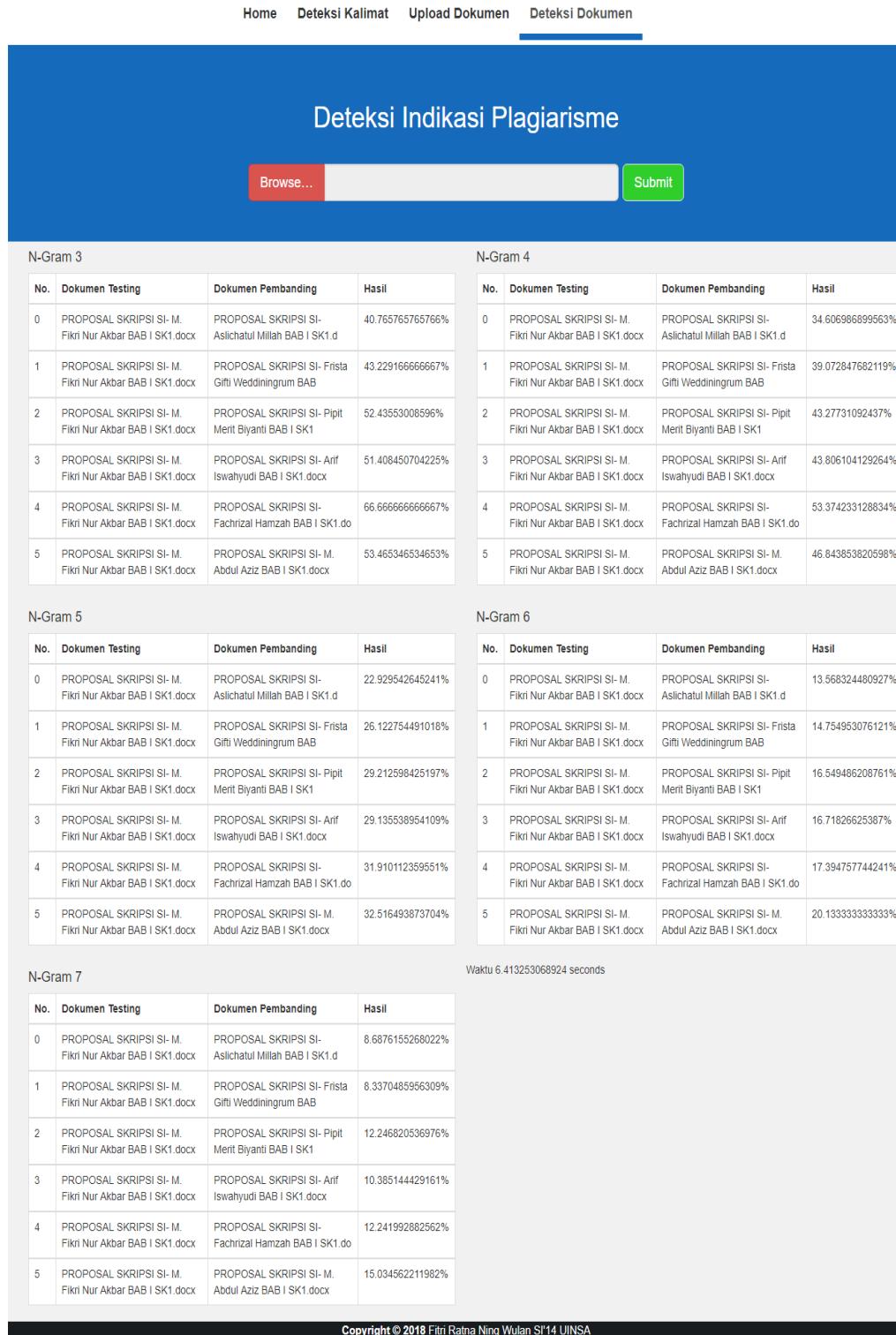
Tabel 4.6 menjelaskan dokumen-dokumen Sistem Informasi yang digunakan sebagai data pembanding pada skenario 1 sebanyak 6 dokumen. Setelah ditentukan dokumen *testing* dan dokumen pembandingnya, selanjutnya menginputkan dokumen pembanding ke dalam database menggunakan sistem deteksi indikasi plagiarisme dengan membuka menu unggah dokumen, selanjutnya menekan tombol *Browse* untuk memilih dokumen pembanding, berikut gambaran daftar dokumen pembanding setelah diunggah pada Gambar 4.13 berikut:

**Unggah Dokumen Pembanding**

No.	Dokumen yang Telah di Unggah	Ukuran Dokumen (Kb)	Aksi
1	PROPOSAL SKRIPSI SI- Arif Iswahyudi BAB I SK1.docx	24608	<a href="#" style="background-color: red; color: white; padding: 2px 5px; border-radius: 3px;">Hapus</a>
2	PROPOSAL SKRIPSI SI- Aslitchatul Millah BAB I SK1.d	30280	<a href="#" style="background-color: red; color: white; padding: 2px 5px; border-radius: 3px;">Hapus</a>
3	PROPOSAL SKRIPSI SI- Fachrizal Hamzah BAB I SK1.do	18673	<a href="#" style="background-color: red; color: white; padding: 2px 5px; border-radius: 3px;">Hapus</a>
4	PROPOSAL SKRIPSI SI- Frista Gifti Weddiningrum BAB	26790	<a href="#" style="background-color: red; color: white; padding: 2px 5px; border-radius: 3px;">Hapus</a>
5	PROPOSAL SKRIPSI SI- M. Abdul Aziz BAB I SK1.docx	41973	<a href="#" style="background-color: red; color: white; padding: 2px 5px; border-radius: 3px;">Hapus</a>
6	PROPOSAL SKRIPSI SI- Pipit Merit Biyanti BAB I SK1	37070	<a href="#" style="background-color: red; color: white; padding: 2px 5px; border-radius: 3px;">Hapus</a>

Gambar 4. 13 Unggah Dokumen Pembanding Sistem Informasi

Setelah selesai mengunggah 6 dokumen pembanding, selanjutnya masuk pada menu Deteksi Dokumen untuk memulai mendeteksi indikasi plagiarisme berdasarkan kesamaan kata dengan mengunggah dokumen *testing*. Berikut hasil prosentase dari pengujian deteksi indikasi plagiarisme, seperti berikut :



Gambar 4. 14 Hasil Pengujian Skenario 1

Gambar 4.14 menggambarkan hasil prosentase pengujian skenario 1 mulai dari  $n=3$  sampai  $n=7$  dengan kecepatan waktu eksekusi 6 detik, berdasarkan hasil tersebut diketahui hasil prosentase setiap  $n$  berbeda-beda. Untuk lebih detail berikut tabel hasil pengujian sistem dengan skenario 1 untuk  $n=3$  sebagai berikut:

Tabel 4. 7 Hasil Pengujian Sistem dokumen sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	n	Prosentase
1	PROPOSAL SKRIPSI SI-M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI-Aslitchatul Millah BAB I SK1.docx	3	40,77%
2	PROPOSAL SKRIPSI SI-M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI-Frista Gifti Weddiningrum BAB I SK1.docx	3	43,23%
3	PROPOSAL SKRIPSI SI-M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Pipit Merit Biyanti BAB I SK1.docx	3	52,44%
4	PROPOSAL SKRIPSI SI-M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Arif Iswahyudi BAB I SK1.docx	3	51,41%
5	PROPOSAL SKRIPSI SI-M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Fachrizal Hamzah BAB I SK1.docx	3	66,67%
6	PROPOSAL SKRIPSI SI-M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- M. Abdul Aziz BAB I SK1.docx	3	53,47%

Berdasarkan Tabel 4.7 Hasil pengujian skenario 1 untuk  $n = 3$  didapatkan prosentase berbeda-beda dengan prosentase paling tinggi 66,67%. Berdasarkan hasil pengujian dengan setiap  $n$  berbeda-beda, hal tersebut belum bisa diketahui nilai  $n$  yang paling efektif diantara  $n = 3$  sampai  $n = 7$ . Oleh karena itu dilakukan pengujian secara manual untuk membandingkan antara pengujian sistem dan pengujian secara manual untuk mengetahui prosentase sistem yang mendekati prosentase hasil manual. Prosentase sistem dari setiap  $n$  yang mendekati prosentase pengujian manual membuktikan bahwa  $n$  tersebut paling efektif digunakan untuk mendeteksi plagiasi pada skenario 1.

Setelah didapatkan hasil pengujian menggunakan sistem, selanjutnya hasil prosentase tersebut dibandingkan dengan hasil prosentase perhitungan kesamaan kata secara manual. Berikut hasil perhitungan kesamaan kata yang dilakukan secara manual pada Tabel 4.8 berikut:

Tabel 4. 8 Hasil Pengujian Manual skenario 1 dokumen sistem informasi

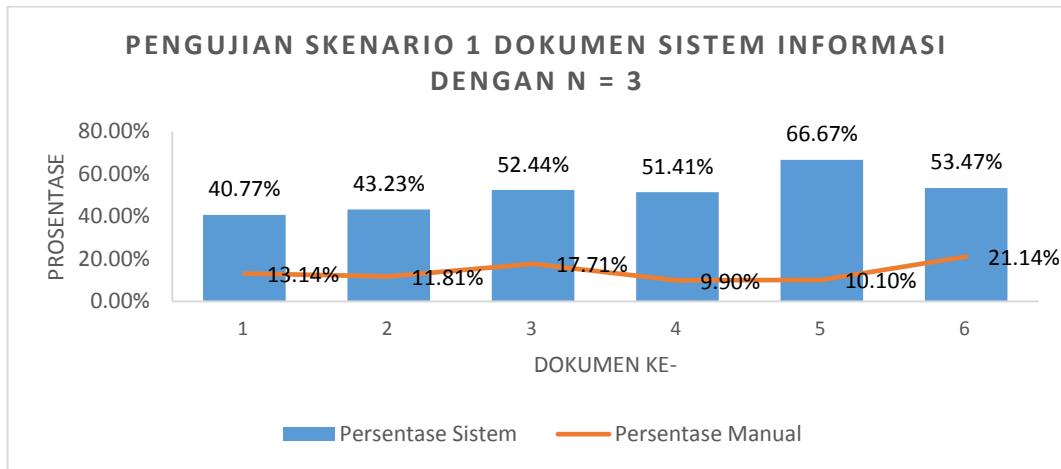
No.	Dokumen Testing	Dokumen Pembanding	Jumlah Kata	Prosentase
1	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI-Aslichatul Millah BAB I SK1.docx	69 dari 525	13,14%
2	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI-Frista Gifti Weddiningrum BAB I SK1.docx	62 dari 525	11,81%
3	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI-Pipit Merit Biyanti BAB I SK1.docx	93 dari 525	17,71%
4	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI-Arif Iswahyudi BAB I SK1.docx	52 dari 525	9,90%
5	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI-Fachrizal Hamzah BAB I SK1.docx	53 dari 525	10,10%
6	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI-M. Abdul Aziz BAB I SK1.docx	111 dari 525	21,14%

Tabel 4.8 menjelaskan hasil perhitungan secara manual dengan menghitung jumlah kata yang sama antara dokumen *testing* dan dokumen pembanding, kemudian membagi dengan jumlah kata yang ada pada dokumen *testing*.

Berikut hasil dari perbandingan hasil prosentase secara manual dan menggunakan sistem, sehingga dapat diketahui nilai  $n$  yang paling efektif untuk digunakan mendekripsi indikasi plagiarisme.

- #### 1. Pengujian Skenario 1 dokumen Sistem Informasi dengan $N = 3$

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=3$ , dan pengujian secara manual, seperti pada Gambar 4.15 berikut :



Gambar 4. 15 Grafik pengujian n = 3 skenario 1 prodi sistem informasi

Berdasarkan grafik pengujian pada Gambar 4.15 dapat dikatakan bahwa pengujian sistem dengan  $n = 3$  mempunyai jarak yang besar dengan pengujian secara manual, untuk mencari akurasi pengujian, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual untuk mencari tingkat akurasi dari pengujian, seperti pada Tabel 4.9 berikut:

Tabel 4. 9 Hasil Pengujian secara sistem dan manual N=3 Skenario 1

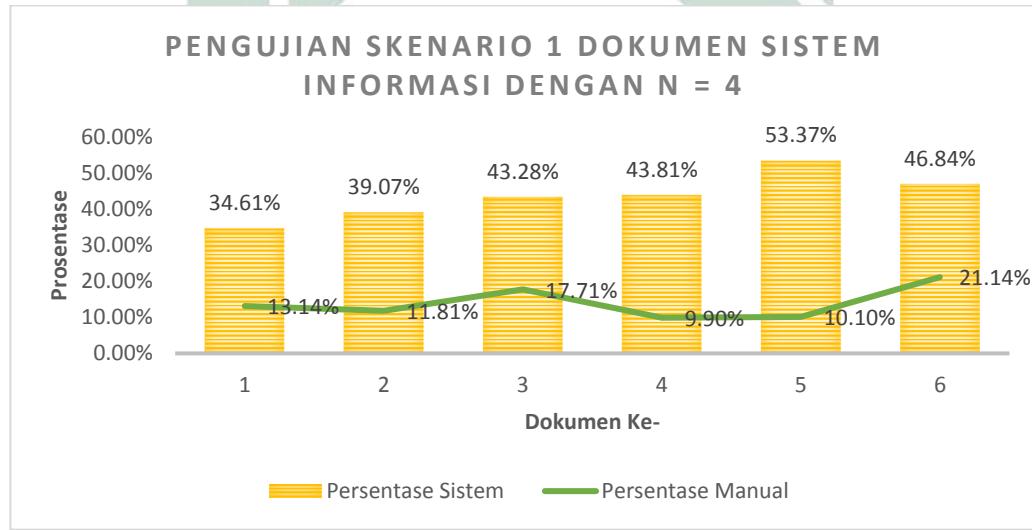
No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Aslichatul Millah BAB I SK1.docx	40,77%	13,14%	27,63%
2	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Frista Gifti Weddiningrum BAB I SK1.docx	43,23%	11,81%	31,42%
3	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Pipit Merit Biyanti BAB I SK1.docx	52,44%	17,71%	34,73%
4	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Arif Iswahyudi BAB I SK1.docx	51,41%	9,90%	41,51%
5	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Fachrizal Hamzah BAB I SK1.docx	66,67%	10,10%	56,57%

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
6	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- M. Abdul Aziz BAB I SK1.docx	53,47%	21,14%	32,33%
Jumlah					224,19%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					37,37%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.9, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 1 dokumen produksi sistem informasi  $n = 3$  didapatkan tingkat akurasi 37,37%.

## 2. Pengujian Skenario 1 dokumen Sistem Informasi dengan N = 4

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=4$ , dan pengujian secara manual, seperti pada Gambar 4.16 berikut:



Gambar 4. 16 Grafik pengujian skenario 1 n=4 prodi sistem informasi

Berdasarkan grafik diatas untuk mencari akurasi pengujian pada  $n = 4$  skenario 1 dokumen prodi sistem informasi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual pada Tabel 4.10 sebagai berikut :

Tabel 4. 10 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=4 prodi sistem informasi

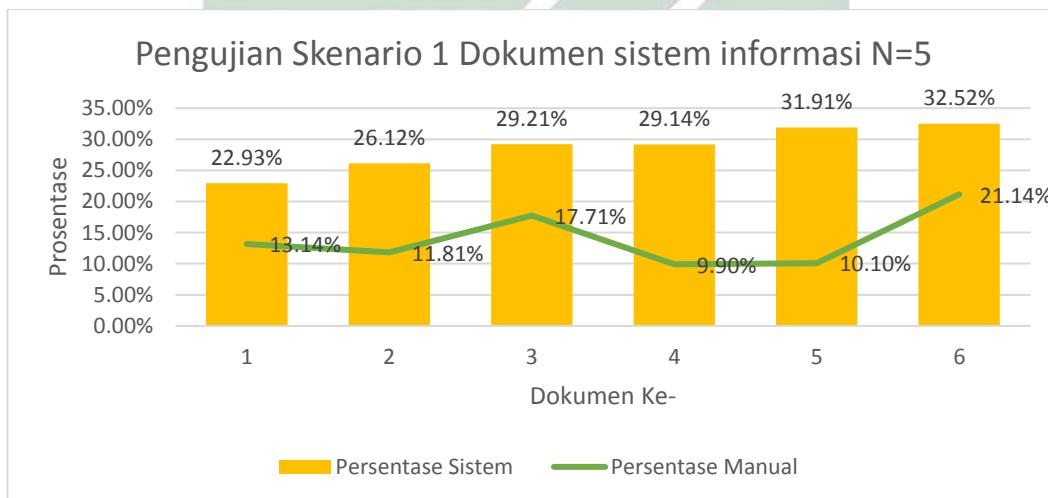
No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Aslitchatul Millah BAB I SK1.docx	34,61%	13,14%	21,47%

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
2	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Frista Gifti Weddiningrum BAB I SK1.docx	39,07%	11,81%	27,26%
3	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Pipit Merit Biyanti BAB I SK1.docx	43,28%	17,71%	25,57%
4	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Arif Iswahyudi BAB I SK1.docx	43,81%	9,90%	33,91%
5	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Fachrizal Hamzah BAB I SK1.docx	53,37%	10,10%	43,27%
6	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- M. Abdul Aziz BAB I SK1.docx	46,84%	21,14%	25,70%
Jumlah					177,18%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					29,53%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.10, didapatkan tingkat akurasi (rata-rata selisih) pada skenario 1 dokumen prodi sistem informasi  $n = 4$  didapatkan tingkat akurasi 29,53%.

### 3. Pengujian Skenario 1 dokumen Sistem Informasi dengan N = 5

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=5$ , dan pengujian secara manual, sebagai berikut :



Gambar 4. 17 Grafik pengujian n=5 skenario 1 prodi sistem informasi

Berdasarkan Grafik pengujian Gambar 4.17 untuk mencari akurasi pengujian pada  $n = 4$  skenario 1 dokumen prodi sistem informasi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase pada Tabel 4.11 berikut:

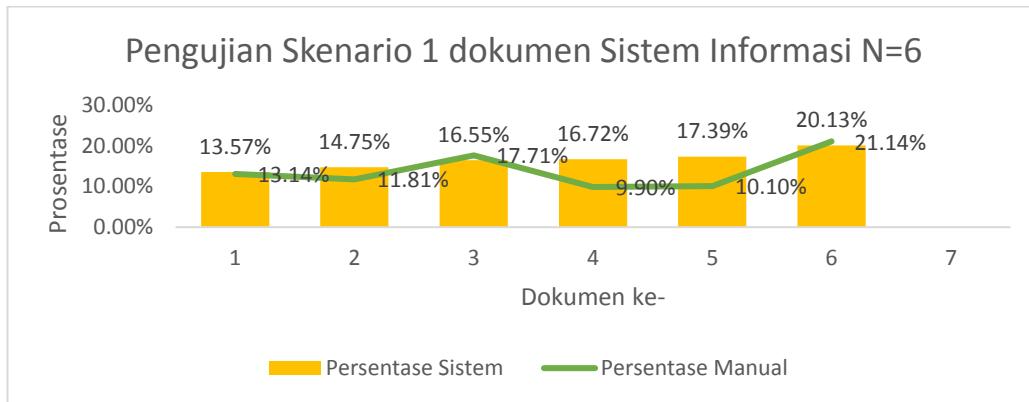
Tabel 4. 11 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=5 prodi sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Aslichatul Millah BAB I SK1.docx	22,93%	13,14%	9,79%
2	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Frista Gifti Weddiningrum BAB I SK1.docx	26,12%	11,81%	14,31%
3	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Pipit Merit Biyanti BAB I SK1.docx	29,21%	17,71%	11,50%
4	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Arif Iswahyudi BAB I SK1.docx	29,14%	9,90%	19,24%
5	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Fachrizal Hamzah BAB I SK1.docx	31,91%	10,10%	21,81%
6	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- M. Abdul Aziz BAB I SK1.docx	32,52%	21,14%	11,38%
Jumlah					88,02%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					14,67%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.11, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 1 dokumen prodi sistem informasi  $n = 5$  didapatkan tingkat akurasi 14,67%.

#### 4. Pengujian Skenario 1 dokumen sistem informasi dengan N = 6

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=6$ , dan pengujian secara manual, sebagai berikut :



Gambar 4. 18 Grafik pengujian n=6 skenario 1 prodi sistem informasi

Berdasarkan grafik pengujian Gambar 4.18 untuk mencari akurasi pengujian pada  $n = 6$  skenario 1 dokumen prodi sistem informasi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual untuk mencari tingkat akurasi dari pengujian, seperti pada Tabel 4.12 berikut :

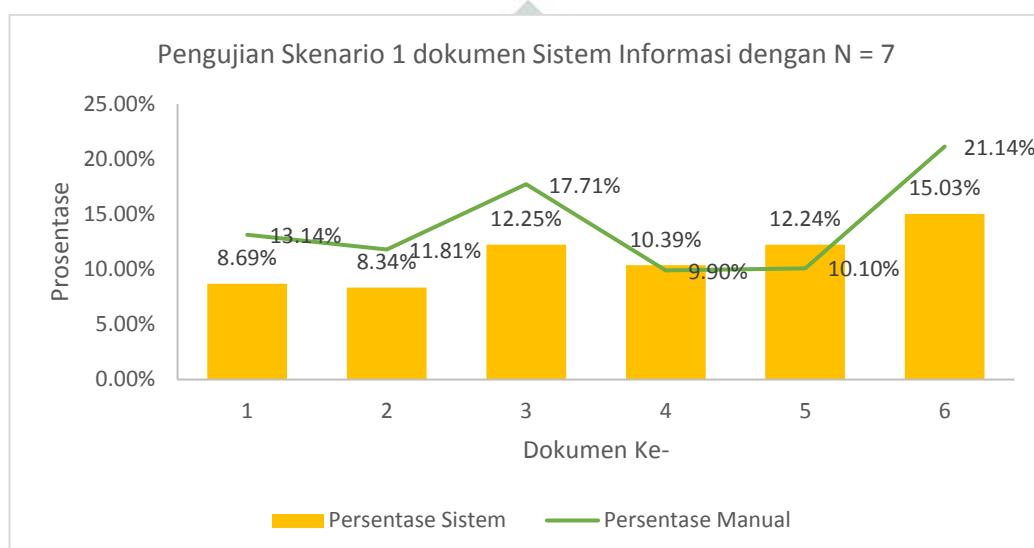
Tabel 4. 12 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=6 prodi sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Aslitchatul Millah BAB I SK1.docx	13,57%	13,14%	0,43%
2	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Frista Gifti Weddiningrum BAB I SK1.docx	14,75%	11,81%	2,94%
3	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Pipit Merit Biyanti BAB I SK1.docx	16,55%	17,71%	1,16%
4	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Arif Iswahyudi BAB I SK1.docx	16,72%	9,90%	6,28%
5	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Fachrizal Hamzah BAB I SK1.docx	17,39%	10,10%	7,29%
6	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- M. Abdul Aziz BAB I SK1.docx	20,13%	21,14%	1,01%
Jumlah					19,11%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					3,19%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.12, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 1 dokumen produksi sistem informasi  $n = 6$  didapatkan tingkat akurasi 3,19%.

## 5. Pengujian Skenario 1 dokumen Sistem Informasi dengan N = 7

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=7$ , dan pengujian secara manual, seperti pada Gambar 4.19 berikut :



Gambar 4. 19 Grafik pengujian n=7 prodi sistem informasi

Berdasarkan grafik pengujian pada Gambar 4.19 untuk mencari akurasi pengujian pada  $n = 7$  skenario 1 dokumen ini, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual seperti pada Tabel 4.13 berikut:

Tabel 4. 13 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=7 prodi sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Aslitchatul Millah BAB I SK1.docx	8,69%	13,14%	4,45%
2	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Frista Gifti Weddiningrum BAB I SK1.docx	8,34%	11,81%	3,47%
3	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Pipit Merit Biyanti BAB I SK1.docx	12,25%	17,71%	5,46%

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
4	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Arif Iswahyudi BAB I SK1.docx	10,39%	9,90%	0,49%
5	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- Fachrizal Hamzah BAB I SK1.docx	12,24%	10,10%	2,14%
6	PROPOSAL SKRIPSI SI- M. Fikri Nur Akbar BAB I SK1.docx	PROPOSAL SKRIPSI SI- M. Abdul Aziz BAB I SK1.docx	15,03%	21,14%	6,11%
Jumlah					22,12%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					3,69%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.13, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 1 dokumen produksi sistem informasi  $n = 7$  didapatkan tingkat akurasi 3,69%.

Berdasarkan hasil perbandingan diatas, dapat disimpulkan skenario 1 untuk pengujian dokumen proposal program studi sistem informasi nilai  $n$  yang paling efektif berdasarkan rata-rata selisih prosentase paling sedikit yaitu  $n = 6$  dengan 3,19%.

#### **4.5.1.1.2 Dokumen Program Studi Biologi**

Untuk pengujian dokumen program studi biologi, berikut daftar dokumen *testing* dan dokumen pembanding, untuk dokumen pembanding seperti pada Tabel 4.14 berikut:

Tabel 4. 14 Dokumen Testing skenario 1 Prodi Biologi

<b>Dokumen Testing Skenario 1 Prodi Biologi</b>	
<b>Nama Dokumen</b>	<b>Ukuran File</b>
PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	20kb

Berdasarkan Tabel 4.14 Dokumen *testing* untuk skenario 1 menggunakan dokumen proposal Nur Setianingsih bab I. Untuk dokumen pembanding Skenario 1 menggunakan 5 dokumen proposal skripsi prodi biologi pada Tabel 4.15 berikut :

Tabel 4. 15 Dokumen Pembanding Skenario 1 Prodi Biologi

<b>Dokumen Pembanding Skenario 1 Prodi Biologi</b>		
<b>No.</b>	<b>Nama Dokumen</b>	<b>Ukuran File</b>
1	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	33kb
2	PROPOSAL SKRIPSI BIO- Anis Mukaromatal Ula SK1.docx	19kb

No.	Nama Dokumen	Ukuran File
3	PROPOSAL SKRIPSI BIO- Arika Wahyuningsih SK1.docx	18kb
4	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	25kb
5	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	30kb

Berdasarkan Tabel 4.15 dokumen *testing* dan dokumen pembanding seperti yang ada pada tabel diatas, selanjutnya mengunggah dokumen pembanding dan mendapatkan *fingerprint* dokumen dan disimpan dalam. Setelah selesai mengunggah 5 dokumen pembanding, selanjutnya masuk pada menu Deteksi Dokumen untuk memulai mendeteksi indikasi plagiarisme berdasarkan kesamaan kata dengan mengunggah dokumen *testing*. Berikut hasil prosentase dari pengujian deteksi indikasi plagiarisme, seperti pada Gambar 4.20 berikut :

Home Deteksi Kalimat Upload Dokumen Deteksi Dokumen

Deteksi Indikasi Plagiarisme

[Browse...](#)

[Submit](#)

N-Gram 3

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	44.53781512605%
1	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Anis Mukaromatul Ula SK1.doc	49.847094801223%
2	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Arika Wahyuningisih SK1.docx	57.894736842105%
3	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	57.241379310345%
4	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	52.076677316294%

N-Gram 4

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	37.921727395412%
1	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Anis Mukaromatul Ula SK1.doc	40.348330914369%
2	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Arika Wahyuningisih SK1.docx	47.184466019417%
3	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	48.310810810811%
4	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	42.040185471406%

N-Gram 5

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	26.490066225166%
1	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Anis Mukaromatul Ula SK1.doc	24.483550114767%
2	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Arika Wahyuningisih SK1.docx	27.987082884822%
3	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	28.663003663004%
4	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	26.81281618887%

N-Gram 6

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	14.34405308831%
1	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Anis Mukaromatul Ula SK1.doc	12.665606783254%
2	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Arika Wahyuningisih SK1.docx	15.793528505393%
3	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	17.627567925779%
4	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	13.279908414425%

N-Gram 7

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	8.396133748056%
1	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Anis Mukaromatul Ula SK1.doc	7.1806157400881%
2	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Arika Wahyuningisih SK1.docx	9.539052496799%
3	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	12.450704225352%
4	PROPOSAL SKRIPSI BIO- Nur Setaningsih SK1.docx	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	7.725527831094%

Waktu 13.327412128448 seconds

Gambar 4. 20 Hasil Pengujian sistem skenario 1 prodi Biologi

Berdasarkan hasil pengujian skenario 1 pada Gambar 4.20 untuk dokumen program studi biologi diatas, didapatkan prosentase indikasi plagiarisme mulai dari  $n=3$  sampai  $n=7$  dengan kecepatan waktu eksekusi 13 detik. Berikut tabel hasil pengujian sistem dengan skenario 1 untuk  $n=3$  pada Tabel 4.16 berikut:

Tabel 4. 16 Hasil Pengujian sistem skenario 1 prodi Biologi

No.	Dokumen Testing	Dokumen Pembanding	n	Prosentase
1	PROPOSAL SKRIPSI BIO-Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO-Andita Febrianti SK1.docx	3	44,54%
2	PROPOSAL SKRIPSI BIO-Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO-Anis Mukaromatul Ula SK1.docx	3	49,85%
3	PROPOSAL SKRIPSI BIO-Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO-Arika Wahyuningsih SK1.docx	3	57,89%
4	PROPOSAL SKRIPSI BIO-Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO-Eni Rahmawati SK1.docx	3	57,24%
5	PROPOSAL SKRIPSI BIO-Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO-Nurul Nahdiyah SK1.docx	3	52,08%

Tabel 4.16 menjelaskan hasil dari pengujian secara sistem untuk skenario 1 program studi Biologi dengan  $n=3$ . Setelah didapatkan hasil pengujian menggunakan sistem, selanjutnya hasil prosentase tersebut dibandingkan dengan hasil prosentase perhitungan kesamaan kata secara manual. Perhitungan secara manual dilakukan dengan menghitung jumlah kata yang sama antara dokumen *testing* dan dokumen pembanding, kemudian membagi dengan jumlah kata yang ada pada dokumen *testing*. Berikut hasil perhitungan kesamaan kata yang dilakukan secara manual.

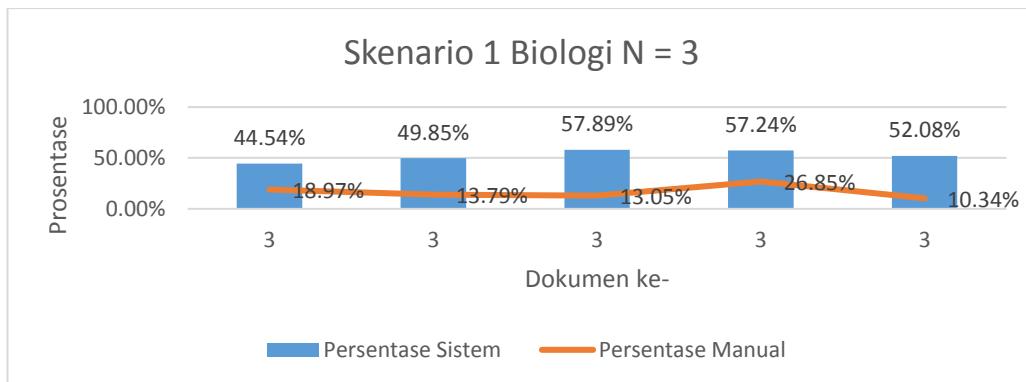
Tabel 4. 17 Hasil Pengujian Manual Skenario 1 Prodi Biologi

No.	Dokumen Testing	Dokumen Pembanding	Jumlah Kata	Prosentase
1	PROPOSAL SKRIPSI BIO-Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO-Andita Febrianti SK1.docx	77 dari 406	18,97%
2	PROPOSAL SKRIPSI BIO-Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO-Anis Mukaromatul Ula SK1.docx	56 dari 406	13,79%
3	PROPOSAL SKRIPSI BIO-Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO-Arika Wahyuningsih SK1.docx	53 dari 406	13,05%
4	PROPOSAL SKRIPSI BIO-Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO-Eni Rahmawati SK1.docx	109 dari 406	26,85%
5	PROPOSAL SKRIPSI BIO-Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO-Nurul Nahdiyah SK1.docx	42 dari 406	10,34%

Berikut hasil dari perbandingan hasil prosentase secara manual dan menggunakan sistem, sehingga dapat diketahui nilai  $n$  yang paling efektif untuk digunakan mendeteksi indikasi plagiarisme.

#### 1. Pengujian Skenario 1 dokumen Biologi dengan N = 3

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=3$ , dan pengujian secara manual, sebagai berikut:



Gambar 4. 21 Grafik pengujian n=3 skenario 1 prodi biologi

Berdasarkan grafik Gambar 4.21 untuk mencari akurasi pengujian pada  $n = 3$  skenario 1 dokumen prodi biologi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual seperti pada Tabel 4.18 berikut :

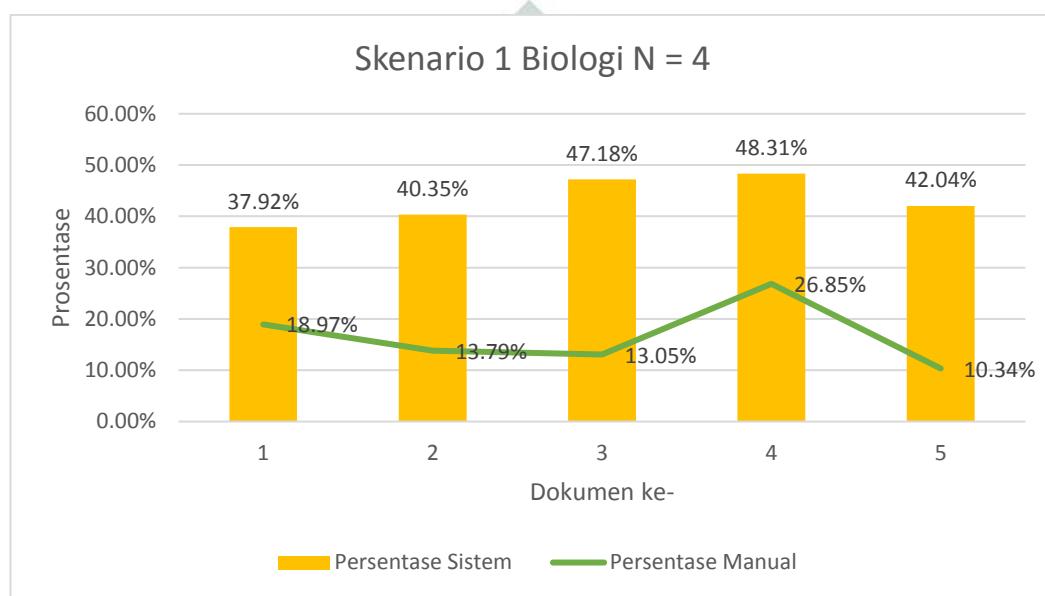
Tabel 4. 18 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=3 prodi Biologi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	44,54%	18,97%	25,57%
2	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Anis Mukaromatul Ula SK1.docx	49,85%	13,79%	36,06%
3	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Arika Wahyuningsih SK1.docx	57,89%	13,05%	44,84%
4	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	57,24%	26,85%	30,39%
5	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	52,08%	10,34%	41,74%
Jumlah					178,60%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					35,72%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.18, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 1 dokumen produksi biologi  $n = 3$  didapatkan tingkat akurasi 35,72%.

## 2. Pengujian Skenario 1 dokumen Biologi dengan N = 4

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=4$ , dan pengujian secara manual, seperti pada Gambar 4.22 berikut :



Gambar 4. 22 Grafik Pengujian n=4 skenario 1 prodi Biologi

Berdasarkan grafik pada Gambar 4.22 untuk mencari akurasi pengujian pada  $n = 4$  skenario 1 dokumen prodi biologi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase, seperti pada Tabel 4.19 berikut:

Tabel 4. 19 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=4 prodi Biologi

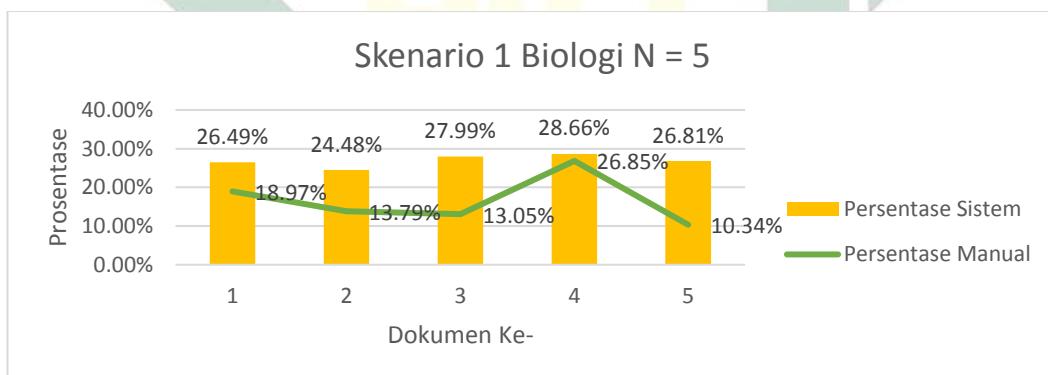
No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	37,92%	18,97%	18,95%
2	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Anis Mukaromatal Ula SK1.docx	40,35%	13,79%	26,56%

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
3	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Arika Wahyuningsih SK1.docx	47,18%	13,05%	34,13%
4	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	48,31%	26,85%	21,46%
5	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	42,04%	10,34%	31,70%
Jumlah					132,80%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					26,56%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.19, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 1 dokumen produksi biologi  $n = 4$  didapatkan tingkat akurasi 26,56%.

### 3. Pengujian Skenario 1 dokumen Biologi dengan N = 5

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=5$ , dan pengujian secara manual, sebagai berikut :



Gambar 4. 23 Grafik Pengujian n=5 skenario 1 prodi Biologi

Berdasarkan grafik pada Gambar 4.23 untuk mencari akurasi pengujian pada  $n = 5$  skenario 1 dokumen prodi biologi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual untuk mencari tingkat akurasi dari pengujian, pada Tabel 4.20 berikut:

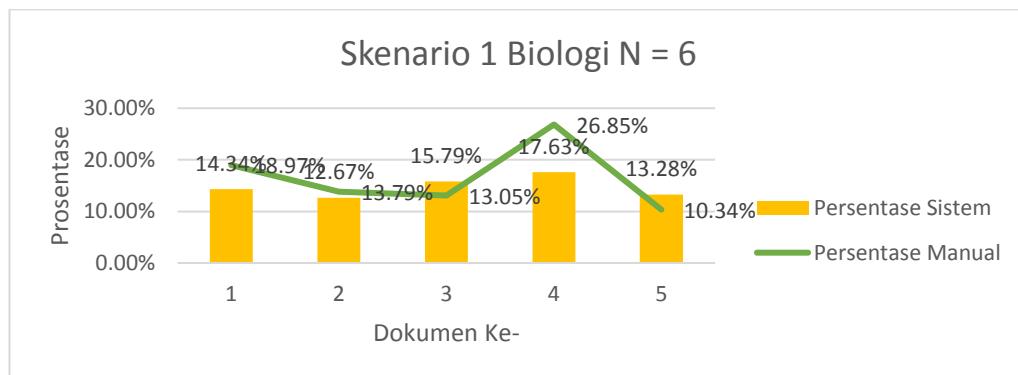
Tabel 4. 20 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=5 prodi Biologi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	26,49%	18,97%	7,52%
2	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Anis Mukaromatul Ula SK1.docx	24,48%	13,79%	10,69%
3	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Arika Wahyuningsih SK1.docx	27,99%	13,05%	14,94%
4	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	28,66%	26,85%	1,81%
5	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	26,81%	10,34%	16,47%
Jumlah					51,43%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					10,29%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.20, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 1 dokumen produksi biologi  $n = 5$  didapatkan tingkat akurasi 26,56%.

#### 4. Pengujian Skenario 1 dokumen Biologi dengan N = 6

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=6$ , dan pengujian secara manual, seperti Gambar 4.24 berikut :



Gambar 4. 24 Hasil Pengujian N=6 Skenario 1 prodi Biologi

Berdasarkan grafik pada Gambar 4.24 untuk mencari akurasi pengujian pada  $n = 6$  skenario 1 dokumen prodi biologi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual pada Tabel 4.21 berikut :

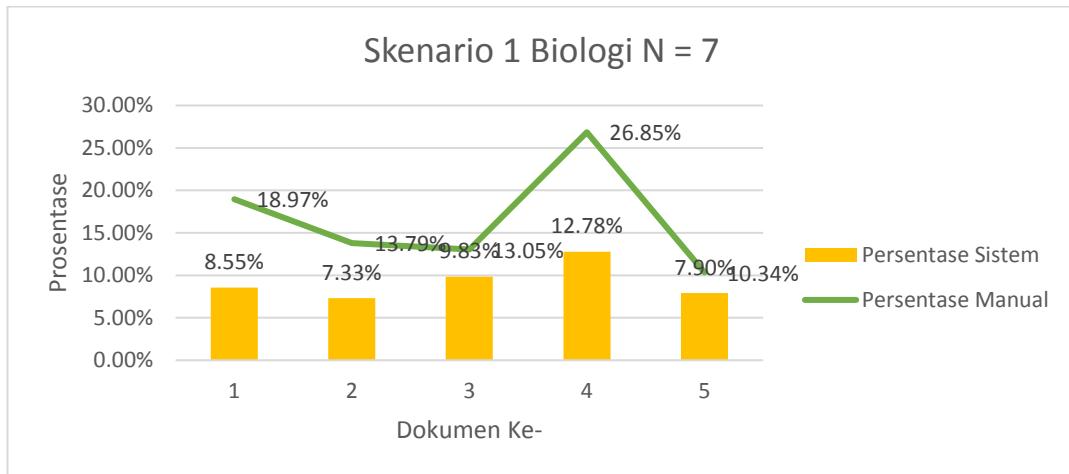
Tabel 4. 21 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=6 prodi Biologi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	14,34%	18,97%	4,63%
2	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Anis Mukaromatul Ula SK1.docx	12,67%	13,79%	1,12%
3	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Arika Wahyuningsih SK1.docx	15,79%	13,05%	2,74%
4	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	17,63%	26,85%	9,22%
5	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	13,28%	10,34%	2,94%
Jumlah					20,65%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					4,13%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.21, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 1 dokumen produksi biologi  $n = 6$  didapatkan tingkat akurasi 4,13%.

## 5. Pengujian Skenario 1 dokumen Biologi dengan N = 7

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=7$ , dan pengujian secara manual, seperti pada Gambar 4.25 berikut :



Gambar 4. 25 Grafik Pengujian n=7 skenario 1 prodi Biologi

Berdasarkan grafik pada Gambar 4.25 untuk mencari akurasi pengujian pada  $n = 7$  skenario 1 dokumen prodi biologi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual seperti pada Tabel 4.22 berikut :

Tabel 4. 22 Hasil Perbandingan pengujian sistem dan manual skenario 1 n=7 prodi Biologi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Andita Febrianti SK1.docx	8,55%	18,97%	10,42%
2	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Anis Mukaromatul Ula SK1.docx	7,33%	13,79%	6,46%
3	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Arika Wahyuningih SK1.docx	9,83%	13,05%	3,22%
4	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Eni Rahmawati SK1.docx	12,78%	26,85%	14,07%
5	PROPOSAL SKRIPSI BIO- Nur Setianingsih SK1.docx	PROPOSAL SKRIPSI BIO- Nurul Nahdiyah SK1.docx	7,90%	10,34%	2,44%
Jumlah					36,61%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					7,32%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.22, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 1 dokumen produksi biologi  $n = 7$  didapatkan tingkat akurasi 7,32%.

Berdasarkan hasil perbandingan dari dokumen Sistem Informasi dan Biologi, dapat disimpulkan skenario 1 untuk pengujian dokumen proposal program studi biologi nilai  $n$  yang paling efektif berdasarkan rata-rata selisih persentase paling sedikit yaitu  $n = 6$  dengan 4,13%.

#### 4.5.1.2 Skenario 2

Skenario 2 untuk pengujian dokumens BAB I yang 100% sama dimulai dengan menentukan data testing dan data pembanding menggunakan BAB I dari masing-masing dokumen. Pada skenario 2 pengujian antara program studi sistem informasi dan program studi biologi juga diuji secara terpisah. Pengujian Skenario 2 dilakukan dengan menyalin semua isi dari salah satu dokumen pembanding menjadi dokumen *testing*, hal tersebut untuk menguji sistem apakah dapat mendeteksi kesamaan dengan 100% sama atau tidak.

- a. Untuk program studi sistem informasi, berikut daftar data *testing* dan data pembanding yang digunakan pada Tabel 4.23 seperti berikut:

Tabel 4. 23 Dokumen Testing skenario 2 prodi sistem informasi

<b>Dokumen Testing</b>	
<b>Nama Dokumen</b>	<b>Ukuran File</b>
PROPOSAL SKRIPSI SI- Aria Dwi Azzida BAB I SK2.docx	27kb

Dokumen *testing* untuk skenario 2 menggunakan dokumen proposal Aria Dwi Azzida bab I. Untuk dokumen pembanding Skenario 2 menggunakan 5 dokumen proposal skripsi prodi sistem informasi. Berikut daftar dokumen pembanding prodi sistem informasi, seperti pada Tabel 4.24 berikut :

Tabel 4. 24 Dokumen pembanding skenario 2 prodi sistem informasi

Dokumen Pembanding		
No.	Nama Dokumen	Ukuran File
1	PROPOSAL SKRIPSI SI- Alfian Pradana Isanto BAB I SK2.docx	28kb
2	PROPOSAL SKRIPSI SI- Cahya Hari Sulistia BAB I SK2.docx	30kb
3	PROPOSAL SKRIPSI SI- Niar Ismi Lafirda As BAB I SK2.docx	33kb
4	PROPOSAL SKRIPSI SI- Sayyidatul Eka Putri Rosalinda BAB I SK2.docx	18kb
5	PROPOSAL SKRIPSI SI- Yudha Tri Satya BAB I SK2.docx	27kb

Setelah ditentukan dokumen *testing* dan dokumen pembandingnya, selanjutnya menginputkan dokumen pembanding ke dalam database menggunakan sistem deteksi indikasi plagiarisme dengan membuka menu Unggah Dokumen, selanjutnya menekan tombol Browse untuk memilih dokumen pembanding. Setelah selesai mengunggah 6 dokumen pembanding, berikut gambaran daftar dokumen pembanding setelah diunggah pada Gambar 4.26 berikut:

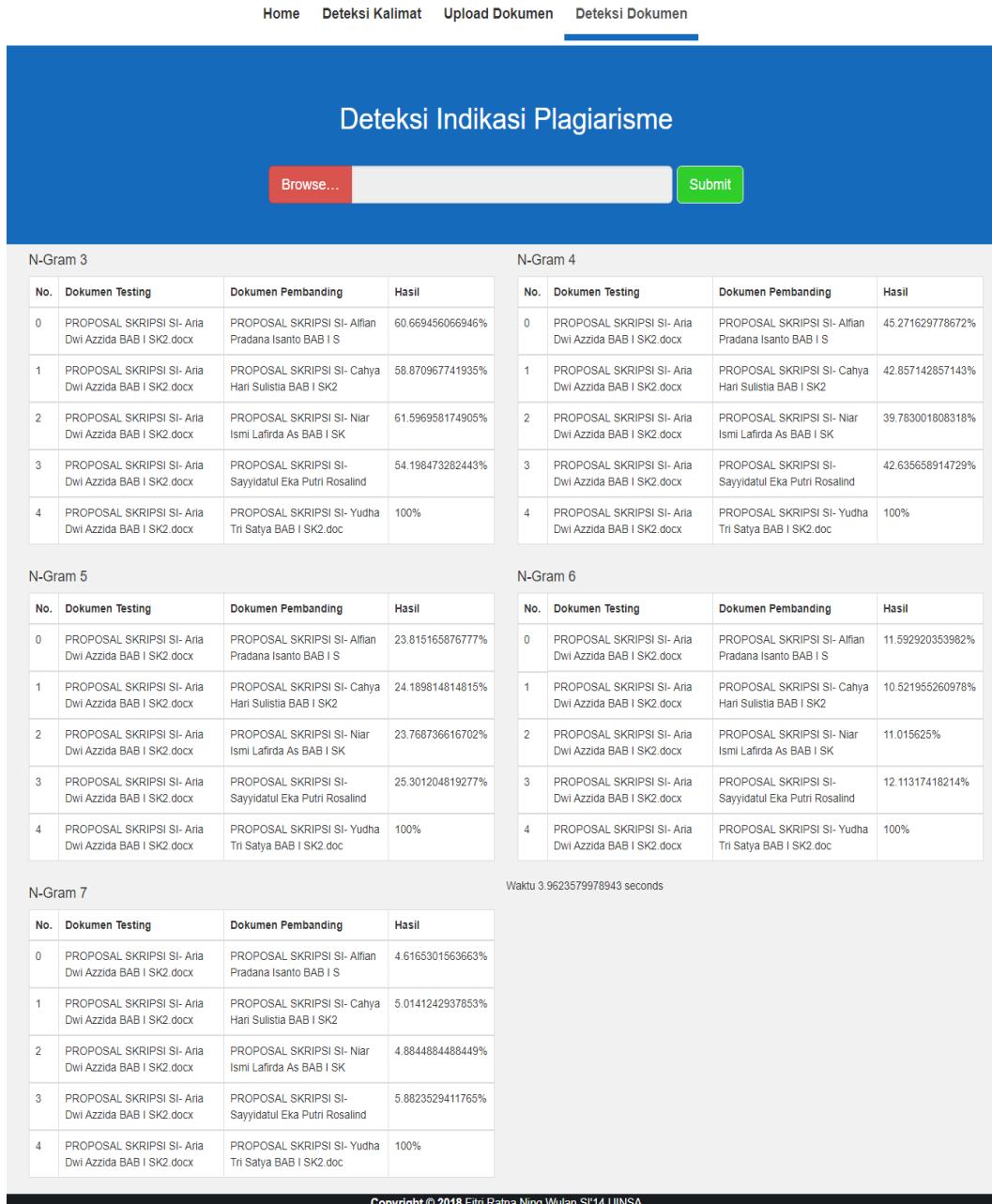
[Home](#)   [Deteksi Kalimat](#)   [Upload Dokumen](#)   [Deteksi Dokumen](#)

## Unggah Dokumen Pembanding

No.	Dokumen yang Telah di Unggah	Ukuran Dokumen (Kb)	Aksi
1	PROPOSAL SKRIPSI SI- Alfian Pradana Isanto BAB I S	28232	<a href="#">Hapus</a>
2	PROPOSAL SKRIPSI SI- Cahya Hari Sulistia BAB I SK2	30564	<a href="#">Hapus</a>
3	PROPOSAL SKRIPSI SI- Niar Ismi Lafirda As BAB I SK	33271	<a href="#">Hapus</a>
4	PROPOSAL SKRIPSI SI- Sayyidatul Eka Putri Rosalind	17819	<a href="#">Hapus</a>
5	PROPOSAL SKRIPSI SI- Yudha Tri Satya BAB I SK2.doc	27226	<a href="#">Hapus</a>

Gambar 4. 26 Tampilan Upload Dokumen Sistem Informasi Skenario 2

Selanjutnya masuk pada menu Deteksi Dokumen untuk memulai mendeteksi indikasi plagiarisme berdasarkan kesamaan kata dengan mengunggah dokumen *testing*. Berikut hasil prosentase dari pengujian deteksi indikasi plagiarisme, seperti pada Gambar 4.27 berikut :



Gambar 4. 27 Hasil Pengujian sistem skenario 2 prodi Sistem Informasi

Berdasarkan hasil pengujian pada Gambar 4.27, dapat didapatkan prosentase indikasi plagiarisme mulai dari  $n=3$  sampai  $n=7$  dapat mendeteksi kesamaan kata 100% sama dengan kecepatan waktu eksekusi 3 detik. Tabel 4.25 berisi hasil pengujian sistem dengan skenario 1 untuk  $n=3$  sepperti berikut:

Tabel 4. 25 Hasil Pengujian skenario 2 prodi sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	n	Prosentase
1	PROPOSAL SKRIPSI SI-Aria Dwi Azzida BAB I SK2.docx	PROPOSAL SKRIPSI SI-Alfian Pradana Isanto BAB I SK2.docx	3	60,67%
2	PROPOSAL SKRIPSI SI-Aria Dwi Azzida BAB I SK2.docx	PROPOSAL SKRIPSI SI-Cahya Hari Sulistia BAB I SK2.docx	3	58,87%
3	PROPOSAL SKRIPSI SI-Aria Dwi Azzida BAB I SK2.docx	PROPOSAL SKRIPSI SI- Niar Ismi Lafirda As BAB I SK2.docx	3	61,60%
4	PROPOSAL SKRIPSI SI-Aria Dwi Azzida BAB I SK2.docx	PROPOSAL SKRIPSI SI-Sayyidatul Eka Putri Rosalinda BAB I SK2.docx	3	54,20%
5	PROPOSAL SKRIPSI SI-Aria Dwi Azzida BAB I SK2.docx	PROPOSAL SKRIPSI SI-Yudha Tri Satya BAB I SK2.docx	3	100%

Berdasarkan Tabel 4.25 dapat diketahui bahwa sistem mampu mendekripsi kesamaan kata dengan 100% sama.

- b. Untuk program studi biologi, berikut daftar data testing dan data pembanding yang digunakan.

Tabel 4. 26 Dokumen Testing skenario 2 prodi Biologi

<b>Dokumen Testing</b>	
<b>Nama Dokumen</b>	<b>Ukuran File</b>
PROPOSAL SKRIPSI BIO- udin SK2.docx	27kb

Dokumen *testing* untuk skenario 2 menggunakan dokume proposal Muhibbudin bab I. Untuk dokumen pembanding Skenario 2 prodi biologi menggunakan 5 dokume proposal skripsi prodi sistem informai. Berikut daftar dokumen pembanding prodi biologi, sebagai berikut :

Tabel 4. 27 Dokumen Pembanding skenario 2 prodi Biologi

<b>Dokumen Pembanding</b>		
<b>No.</b>	<b>Nama Dokumen</b>	<b>Ukuran File</b>
1	PROPOSAL SKRIPSI BIO- Devi Silvia SK2.docx	28kb
2	PROPOSAL SKRIPSI BIO- Dita Umi BAB I SK2.docx	30kb
3	PROPOSAL SKRIPSI BIO- Ika Sayyidatul Khumairoh SK2.docx	33kb
4	PROPOSAL SKRIPSI BIO- Khoirun Nihayati SK2.docx	18kb
5	PROPOSAL SKRIPSI BIO- Ulfa BAB I SK2.docx	27kb

Setelah ditentukan dokumen testing dan dokumen pembandingnya, selanjutnya menginputkan dokumen pembanding ke dalam database menggunakan sistem

deteksi indikasi plagiarisme dengan membuka menu **unggah dokumen**, selanjutnya menekan tombol **Browse** untuk memilih dokumen pembanding. Setelah selesai mengunggah 6 dokumen pembanding, selanjutnya masuk pada menu **Deteksi Dokumen** untuk memulai mendeteksi indikasi plagiarisme berdasarkan kesamaan kata dengan mengunggah dokumen *testing*. Berikut hasil prosentase dari pengujian deteksi indikasi plagiarisme, seperti berikut :

Home Deteksi Kalimat Upload Dokumen Deteksi Dokumen

## Deteksi Indikasi Plagiarisme

**N-Gram 3**

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Devi Silvia SK2.docx	59.530791788856%
1	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Dita Umi BAB I SK2.docx	62.893081761006%
2	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Ika Sayyidatul Khumairoh SK2	100%
3	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Khoirun Nihayati SK2.docx	49.777777777778%
4	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Ulfa BAB I SK2.docx	59.20245398773%

**N-Gram 5**

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Devi Silvia SK2.docx	39.691444600281%
1	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Dita Umi BAB I SK2.docx	37.373737373737%
2	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Ika Sayyidatul Khumairoh SK2	100%
3	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Khoirun Nihayati SK2.docx	34.948096885813%
4	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Ulfa BAB I SK2.docx	35.285396590067%

**N-Gram 7**

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Devi Silvia SK2.docx	10.873983739837%
1	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Dita Umi BAB I SK2.docx	8.5842852197992%
2	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Ika Sayyidatul Khumairoh SK2	100%
3	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Khoirun Nihayati SK2.docx	7.9423403052572%
4	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Ulfa BAB I SK2.docx	8.0893682586598%

**N-Gram 4**

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Devi Silvia SK2.docx	57.28021978022%
1	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Dita Umi BAB I SK2.docx	56.458635703919%
2	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Ika Sayyidatul Khumairoh SK2	100%
3	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Khoirun Nihayati SK2.docx	46.702702702703%
4	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Ulfa BAB I SK2.docx	52.789699570815%

**N-Gram 6**

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Devi Silvia SK2.docx	21.233480176211%
1	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Dita Umi BAB I SK2.docx	17.852408307556%
2	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Ika Sayyidatul Khumairoh SK2	100%
3	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Khoirun Nihayati SK2.docx	16.60777385159%
4	PROPOSAL SKRIPSI BIO-udin SK2.docx	PROPOSAL SKRIPSI BIO- Ulfa BAB I SK2.docx	17.536231884058%

Waktu 13.922622919083 seconds

© 2024 Sistem Pengidentifikasi Plagiarisme

Gambar 4. 28 Hasil Pengujian sistem skenario 2 prodi Biologi

Berdasarkan hasil pengujian pada Gambar 4.28, dapat didapatkan persentase indikasi plagiarisme mulai dari  $n=3$  sampai  $n=7$  dengan kecepatan waktu eksekusi 12 detik, berikut tabel hasil pengujian sistem dengan skenario 1 untuk  $n=3$  sebagai berikut:

Tabel 4. 28 Hasil Pengujian sistem skenario 2 prodi Biologi

No.	Dokumen Testing	Dokumen Pembanding	n	Prosentase
1	PROPOSAL SKRIPSI BIO- udin SK2.docx	PROPOSAL SKRIPSI BIO- Devi Silvia SK2.docx	3	59,53%
2	PROPOSAL SKRIPSI BIO- udin SK2.docx	PROPOSAL SKRIPSI BIO- Dita Umi BAB I SK2.docx	3	62,89%
3	PROPOSAL SKRIPSI BIO- udin SK2.docx	PROPOSAL SKRIPSI BIO- Ika Sayyidatul Khumairoh SK2.docx	3	100%
4	PROPOSAL SKRIPSI BIO- udin SK2.docx	PROPOSAL SKRIPSI BIO- Khoirun Nihayati SK2.docx	3	49,78%
5	PROPOSAL SKRIPSI BIO- udin SK2.docx	PROPOSAL SKRIPSI BIO- Ulfa BAB I SK2.docx	3	59,20%

Berdasarkan Tabel 4.28 dapat diketahui bahwa sistem mampu mendekripsi kesamaan dengan 100% sama.

Dari kedua pengujian yang dilakukan pada dokumen prodi sistem informasi dan biologi, dapat disimpulkan bahwa hasil dari skenario 2 membuktikan bahwa sistem bisa mendeteksi kesamaan kata 100%.

#### **4.5.1.3 Skenario 3**

Pengujian untuk dokumen BAB I yang sebagian sama, dilakukan dengan menguji dokumen proposal prodi sistem informasi dan biologi secara terpisah. Hasil dari skenario 3 diharapkan dapat membuktikan bahwa sistem bisa mendeteksi adanya indikasi plagiasi dengan sebagian sama pada dokumen *testing* dan dokumen pembanding. Pengujian dilakukan dengan menyalin satu paragraf dari masing-masing dokumen pembanding ke dokumen *testing*, sehingga dapat dihitung jumlah kesamaan antara dokumen *testing* dan dokumen pembanding secara manual untuk membuktikan pengujian antara sistem dan manual.

#### **4.5.1.3.1 Dokumen Program Studi Sistem Informasi**

Untuk program studi sistem informasi, berikut daftar data *testing* dan data pembanding yang digunakan pada Tabel 4.29 berikut:

Tabel 4. 29 Dokumen Testing skenario 3 prodi Sistem Informasi

<b>Dokumen <i>Testing</i></b>	
<b>Nama Dokumen</b>	<b>Ukuran File</b>
PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	19kb

Dokumen *testing* untuk skenario 3 menggunakan dokume proposal Taufiqurrochman bab I. Untuk dokumen pembanding Skenario 3 menggunakan 6 dokume proposal skripsi prodi sistem informai. Berikut daftar dokumen pembanding prodi sistem informasi, pada Tabel 4.30 berikut :

Tabel 4. 30 Dokumen Pembanding skenario 3 prodi sistem informasi

Dokumen Pembanding		
No.	Nama Dokumen	Ukuran File
1	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	16kb
2	PROPOSAL SKRIPSI SI- Laili Nur Qomary BAB I SK3.docx	26kb
3	PROPOSAL SKRIPSI SI- Lutfiana Fatmawati BAB I SK3.docx	34kb
4	PROPOSAL SKRIPSI SI- Ahmad Mustain Marzuki BAB I SK3	17kb
5	PROPOSAL SKRIPSI SI- Siti Nurazizah BAB I SK3.docx	31kb
6	PROPOSAL SKRIPSI SI- Tsania Saraswati BAB I SK3.docx	26kb

Setelah ditentukan dokumen *testing* dan dokumen pembandingnya, selanjutnya menginputkan dokumen pembanding ke dalam database menggunakan sistem deteksi indikasi plagiarisme dengan membuka menu **unggah dokumen**. Setelah selesai mengunggah 6 dokumen pembanding, berikut gambaran daftar dokumen pembanding setelah di unggah pada Gambar 4.29 berikut:

[Home](#)
[Deteksi Kalimat](#)
[Upload Dokumen](#)
[Deteksi Dokumen](#)

## Unggah Dokumen Pembanding

No.	Dokumen yang Telah di Unggah	Ukuran Dokumen (Kb)	Aksi
1	PROPOSAL SKRIPSI SI- Ahmad Mustain Marzuki BAB I S	16927	Hapus
3	PROPOSAL SKRIPSI SI- Laili Nur Qomary BAB I SK3.do	25861	Hapus
4	PROPOSAL SKRIPSI SI- Lutfiani Fatmawati BAB I SK3.	34065	Hapus
5	PROPOSAL SKRIPSI SI- Siti Nurazizah BAB I SK3.docx	31172	Hapus

Gambar 4. 29 Tampilan Upload Dokumen Sistem Informasi Skenario 3

selanjutnya masuk pada menu **Deteksi Dokumen** untuk memulai mendeteksi indikasi plagiarisme berdasarkan kesamaan kata dengan mengunggah dokumen *testing*. Berikut hasil prosentase dari pengujian deteksi indikasi plagiarisme, pada Gambar 2.30 berikut :

Home Deteksi Kalimat Upload Dokumen Deteksi Dokumen

## Deteksi Indikasi Plagiarisme

### N-Gram 3

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	54.125412541254%
1	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Laili Nur Qomary BAB I SK3.doc	62.686567164179%
2	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Lutfiana Fatmawati BAB I SK3.	68.421052631579%
3	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Ahmad Mustian Marzuki BAB I S	60.687022900763%
4	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Siti Nurazizah BAB I SK3.docx	66.019417475728%
5	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Tsania Saraswati BAB I SK3 do	59.006211180124%

### N-Gram 4

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	41.35702747463651%
1	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Laili Nur Qomary BAB I SK3.do	49.916247906198%
2	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Lutfiana Fatmawati BAB I SK3.	62.593144560358%
3	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Ahmad Mustian Marzuki BAB I S	51.601423487544%
4	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Siti Nurazizah BAB I SK3.docx	57.838660578387%
5	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Tsania Saraswati BAB I SK3 do	52.529761904762%

### N-Gram 5

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	30.513051305131%
1	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Laili Nur Qomary BAB I SK3 do	34.16149068323%
2	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Lutfiana Fatmawati BAB I SK3.	51.079136690647%
3	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Ahmad Mustian Marzuki BAB I S	38%
4	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Siti Nurazizah BAB I SK3.docx	44.223107569721%
5	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Tsania Saraswati BAB I SK3 do	39.018302828619%

### N-Gram 6

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	17.78992106861%
1	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Laili Nur Qomary BAB I SK3 do	17.121300058038%
2	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Lutfiana Fatmawati BAB I SK3.	36.075949367089%
3	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Ahmad Mustian Marzuki BAB I S	28.327868852459%
4	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Siti Nurazizah BAB I SK3.docx	29.132439659755%
5	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Tsania Saraswati BAB I SK3 do	24.189114898296%

### N-Gram 7

Waktu 19.407456874847 seconds

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	12.20958971624%
1	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Laili Nur Qomary BAB I SK3 do	9.6685082872928%
2	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Lutfiana Fatmawati BAB I SK3.	28.409090909091%
3	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Ahmad Mustian Marzuki BAB I S	21.553090332805%
4	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Siti Nurazizah BAB I SK3.docx	19.60049937578%
5	PROPOSAL SKRIPSI SI-Taufiqurochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Tsania Saraswati BAB I SK3 do	14.864864864865%

Copyright © 2018 Fitri Ratna Ning Wuhan ST'14 UINSA

Gambar 4. 30 Hasil Pengujian sistem skenario 3 prodi Sistem Informasi

Berdasarkan hasil pengujian pada Gambar 4.30, dapat didapatkan prosentase indikasi plagiarisme mulai dari  $n=3$  sampai  $n=7$  dengan kecepatan waktu eksekusi 19 detik, berikut tabel hasil pengujian sistem dengan skenario 1 untuk  $n=3$  sebagai berikut:

Tabel 4. 31 Hasil Pengujian sistem skenario 3 prodi Sistem Informasi

No.	Dokumen Testing	Dokumen Pembanding	n	Prosentase
1	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Ilmi Alfianti BAB I SK3.docx	3	54,13%
2	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Laili Nur Qomary BAB I SK3.docx	3	62,69%
3	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Lutfiana Fatmawati BAB I SK3.docx	3	68,42%
4	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Ahmad Mustain Marzuki BAB I SK3	3	68,42%
5	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Siti Nurazizah BAB I SK3.docx	3	66,02%
6	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI-Tsania Saraswati BAB I SK3.docx	3	59,01%

Berdasarkan Tabel 4.31 didapatkan prosentase pengujian secara sistem. Selanjutnya hasil prosentase tersebut dibandingkan dengan hasil prosentase perhitungan kesamaan kata secara manual yang perhitungannya sama seperti perhitungan manual pada skenario 1. Berikut hasil perhitungan secara manual untuk skenario 3 dokumen sistem informasi untuk mencari persentasi akurasi terdapat pada Tabel 4.32:

Tabel 4. 32 Hasil pengujian manual skenario 3 prodi sistem informasi

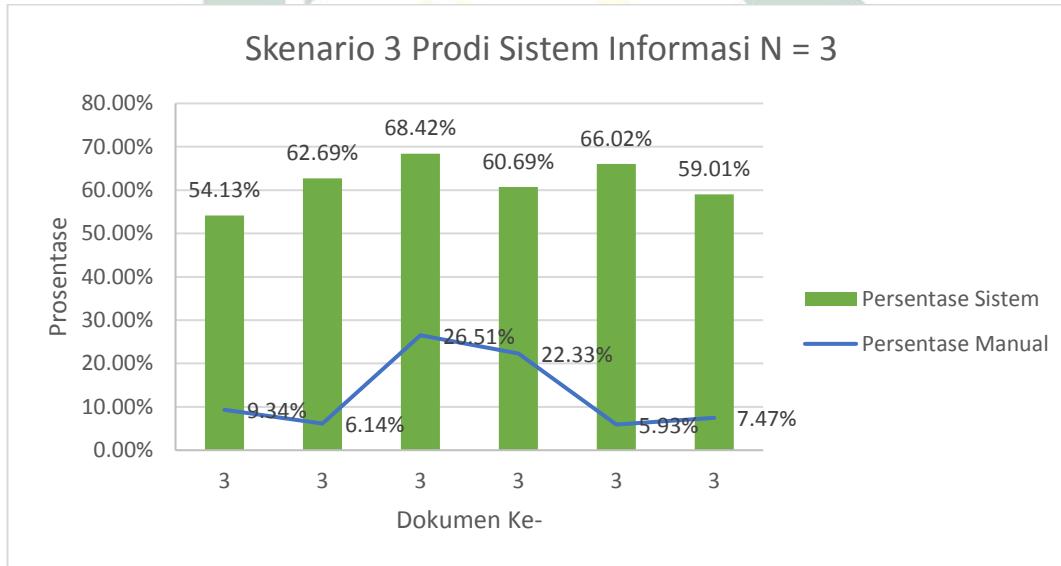
No.	Dokumen Testing	Dokumen Pembanding	Jumlah Kata	Prosentase
1	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	105 dari 1124	9,34%
2	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Laili Nur Qomary BAB I SK3.docx	69 dari 1124	6,14%
3	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Lutfiana Fatmawati BAB I SK3.docx	298 dari 611	26,51%

No.	Dokumen Testing	Dokumen Pembanding	Jumlah Kata	Prosentase
4	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ahmad Mustain Marzuki BAB I SK3	251 dari 1124	22,33%
5	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Siti Nurazizah BAB I SK3.docx	179 dari 1124	5,93%
6	PROPOSAL SKRIPSI SI-Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Tsania Saraswati BAB I SK3.docx	84 dari 1124	7,47%

Berikut hasil dari perbandingan hasil prosentase secara manual dan menggunakan sistem:

- ### 1. Pengujian Skenario 3 dokumen Sistem Informasi dengan $N = 3$

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=3$ , dan pengujian secara manual, terdapat pada Gambar 4.31 berikut:



Gambar 4. 31 Grafik Pengujian n=3 skenario 3 prodi sistem informasi

Berdasarkan grafik pada Gambar 4.31 untuk mencari akurasi pengujian pada  $n = 3$  skenario 3 dokumen prodi sistem informasi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual untuk mencari tingkat akurasi dari pengujian, pada Tabel 4.33 berikut :

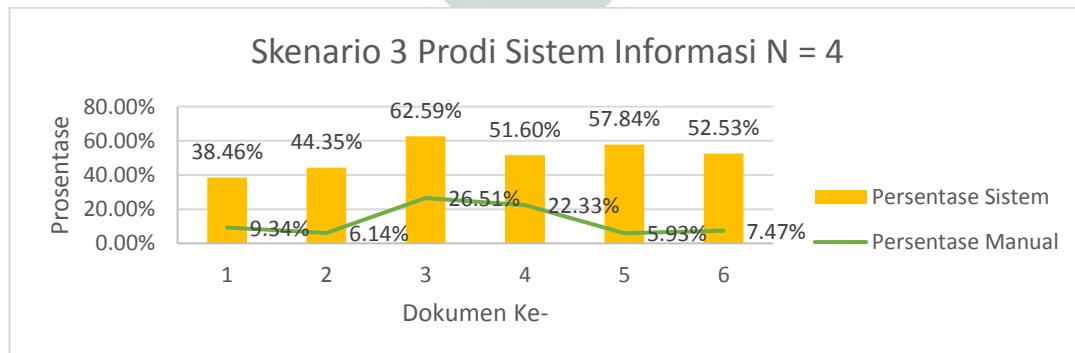
Tabel 4. 33 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=3 prodi sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	54,13%	9,34%	44,79%
2	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Laili Nur Qomary BAB I SK3.docx	62,69%	6,14%	56,55%
3	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Lutfiana Fatmawati BAB I SK3.docx	68,42%	26,51%	41,91%
4	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ahmad Mustain Marzuki BAB I SK3	68,42%	22,33%	38,36%
5	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Siti Nurazizah BAB I SK3.docx	66,02%	5,93%	60,09%
Jumlah					293,23%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					48,87%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.33, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 3 dokumen produksi sistem informasi  $n = 3$  didapatkan tingkat akurasi 48,87%.

## 2. Pengujian Skenario 3 dokumen Sistem Informasi dengan N = 4

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=4$ , dan pengujian secara manual, pada Gambar 4.32 berikut :



Gambar 4. 32 Grafik pengujian n=4 skenario 3 prodi sistem informasi

Berdasarkan grafik pada Gambar 4.32 untuk mencari akurasi pengujian pada  $n = 4$  skenario 3 dokumen prodi sistem informasi, dilakukan perhitungan selisih

antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual, pada Tabel 4.34:

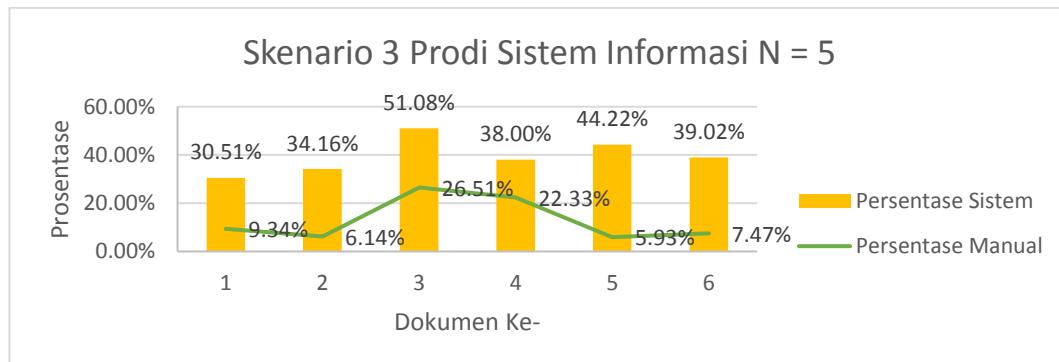
Tabel 4. 34 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=4 prodi sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	38,46%	9,34%	29,12%
2	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Laili Nur Qomary BAB I SK3.docx	44,35%	6,14%	38,21%
3	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Lutfiana Fatmawati BAB I SK3.docx	62,59%	26,51%	36,08%
4	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ahmad Mustain Marzuki BAB I SK3	51,60%	22,33%	29,27%
5	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Siti Nurazizah BAB I SK3.docx	57,84%	5,93%	51,91%
Jumlah					229,65%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					38,27%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.34, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 3 dokumen produksi sistem informasi  $n = 4$  didapatkan tingkat akurasi 38,27%.

### 3. Pengujian Skenario 3 dokumen Sistem Informasi dengan N = 5

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=5$ , dan pengujian secara manual, sebagai berikut :



Gambar 4. 33 Grafik pengujian n=5 skenario 3 prodi Sistem Informasi

Berdasarkan grafik pengujian pada Gambar 4.33 untuk mencari akurasi pengujian pada  $n = 5$  skenario 3 dokumen prodi sistem informasi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual untuk mencari tingkat akurasi dari pengujian, pada Tabel 4.35 berikut :

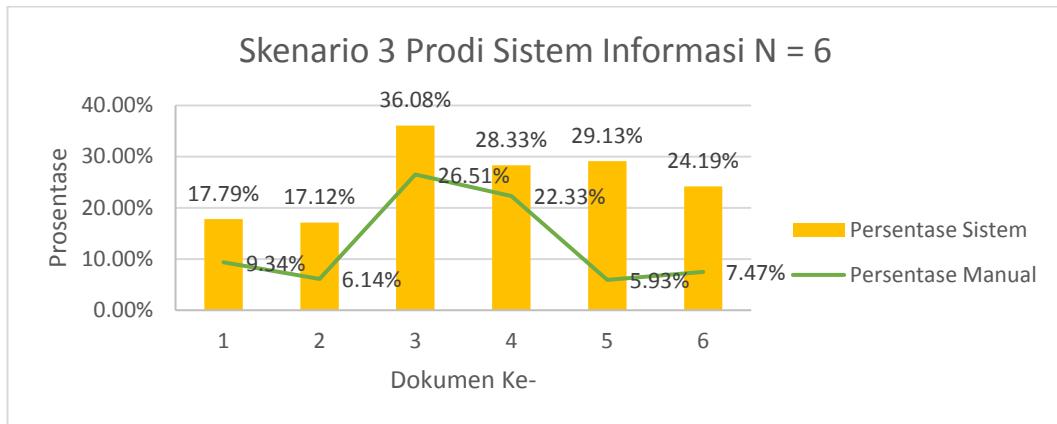
Tabel 4. 35 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=5  
prodi sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	30,51%	9,34%	21,17%
2	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Laili Nur Qomary BAB I SK3.docx	34,16%	6,14%	28,02%
3	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Lutfiana Fatmawati BAB I SK3.docx	51,08%	26,51%	24,57%
4	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ahmad Mustain Marzuki BAB I SK3	38,00%	22,33%	15,67%
5	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Siti Nurazizah BAB I SK3.docx	44,22%	5,93%	38,29%
Jumlah					159,27%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					26,54%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.35, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 3 dokumen produksi sistem informasi  $n = 5$  didapatkan tingkat akurasi 26,54%.

#### 4. Pengujian Skenario 3 dokumen Sistem Informasi dengan N = 6

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=6$ , dan pengujian secara manual, sebagai berikut :



Gambar 4. 34 Grafik Pengujian n=6 skenario 3 prodi Sistem Informasi

Berdasarkan grafik pada Gambar 4.34 untuk mencari akurasi pengujian pada  $n = 6$  skenario 3 dokumen prodi sistem informasi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual pada Tabel 4.36 berikut :

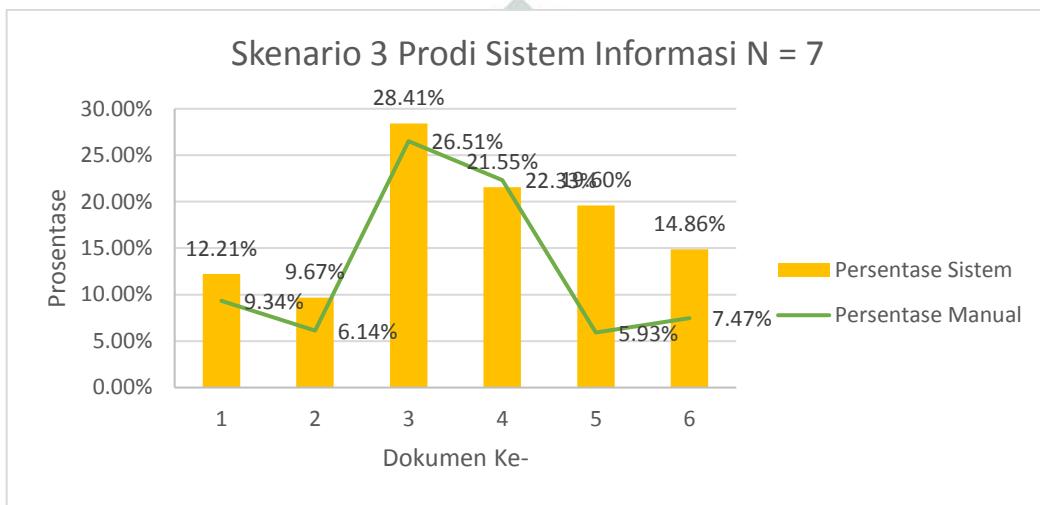
Tabel 4. 36 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=6 prodi sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	17,79%	9,34%	8,45%
2	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Laili Nur Qomary BAB I SK3.docx	17,12%	6,14%	10,98%
3	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Lutfiana Fatmawati BAB I SK3.docx	36,08%	26,51%	9,57%
4	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ahmad Mustain Marzuki BAB I SK3	28,33%	22,33%	6,00%
5	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Siti Nurazizah BAB I SK3.docx	29,13%	5,93%	23,20%
Jumlah					74,92%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					12,49%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.36, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 3 dokumen produksi sistem informasi  $n = 6$  didapatkan tingkat akurasi 12,49%.

## 5. Pengujian Skenario 3 dokumen Sistem Informasi dengan N = 7

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=7$ , dan pengujian secara manual, seperti pada Gambar 4.35 berikut :



Gambar 4. 35 Grafik Pengujian n=7 skenario 3 prodi Sistem Informasi

Berdasarkan grafik pada Gambar 4.35 untuk mencari akurasi pengujian pada  $n = 7$  skenario 3 dokumen prodi sistem informasi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual pada Tabel 4.37 berikut :

Tabel 4. 37 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=7 prodi sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ilmi Alfianti BAB I SK3.docx	12,21%	9,34%	2,87%
2	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Laili Nur Qomary BAB I SK3.docx	9,67%	6,14%	3,53%
3	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Lutfiana Fatmawati BAB I SK3.docx	28,41%	26,51%	1,90%

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
4	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Ahmad Mustain Marzuki BAB I SK3	21,55%	22,33%	0,78%
5	PROPOSAL SKRIPSI SI- Taufiqurrochman BAB I SK3.docx	PROPOSAL SKRIPSI SI- Siti Nurazizah BAB I SK3.docx	19,60%	5,93%	13,67%
Jumlah					30,14%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					5,02%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.37, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 3 dokumen produksi sistem informasi  $n = 7$  didapatkan tingkat akurasi 5,02%..

#### 4.5.1.3.2 Dokumen Program Studi Biologi

Untuk program studi biologi, berikut daftar data *testing* dan data pembanding yang digunakan.

Tabel 4. 38 Dokumen Testing Skenario 3 Prodi Biologi

<b>Dokumen Testing</b>	
<b>Nama Dokumen</b>	<b>Ukuran File</b>
PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	29kb

Dokumen *testing* untuk skenario 3 menggunakan dokume proposal Vida Nofrianinda bab I. Untuk dokumen pembanding Skenario 1 menggunakan 5 dokume proposal skripsi prodi sistem informai. Berikut daftar dokumen pembanding prodi sistem informasi, pada Tabel 4.39 berikut :

Tabel 4. 39 Dokumen Pembanding Skenario 3 Prodi Biologi

<b>Dokumen Pembanding</b>		
<b>No.</b>	<b>Nama Dokumen</b>	<b>Ukuran File</b>
1	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	56kb
2	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	29kb
3	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	27kb
4	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	197kb
5	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	27kb

Setelah ditentukan dokumen *testing* dan dokumen pembandingnya, selanjutnya menginputkan dokumen pembanding ke dalam database menggunakan sistem deteksi indikasi plagiarisme dengan membuka menu **unggah dokumen**, selanjutnya menekan tombol Browse untuk memilih dokumen pembanding.

Setelah selesai mengunggah 6 dokumen pembanding, selanjutnya masuk pada menu Deteksi Dokumen untuk memulai mendeteksi indikasi plagiarisme berdasarkan kesamaan kata dengan mengunggah dokumen *testing*. Berikut hasil prosentase dari pengujian deteksi indikasi plagiarisme, seperti berikut :

Home Deteksi Kalimat Upload Dokumen Deteksi Dokumen

## Deteksi Indikasi Plagiarisme

N-Gram 3

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	74.122807017544%
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	46.582278481013%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	52.786885245902%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	43.811881188119%
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	52.459016393443%

N-Gram 4

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	63.238512035011%
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	39.949431099874%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	45.46925566343%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	35.512820512821%
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	44.844517184943%

N-Gram 5

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	53.581267217631%
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	26.080793763288%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	29.802513464991%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	22.868217054264%
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	28.676470588235%

N-Gram 6

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	46.695095948827%
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	16.273229532898%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	16.232586311326%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	12.365894974591%
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	15.977084659453%

N-Gram 7

No.	Dokumen Testing	Dokumen Pembanding	Hasil
0	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	46.370967741935%
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	10.012259910094%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	8.1466395112016%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	7.1005917159763%
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	8.0347448425624%

Waktu 5.0386879444122 seconds

Copyright © 2018 Fitri Putra Ning Wulan S'14 UINSA

Gambar 4. 36 Hasil Pengujian sistem skenario 3 prodi Biologi

Berdasarkan hasil pengujian pada Gambar 4.36, dapat didapatkan prosentase indikasi plagiarisme mulai dari  $n=3$  sampai  $n=7$  dengan kecepatan waktu eksekusi 5 detik, berikut tabel hasil pengujian sistem dengan skenario 1 untuk  $n=3$  pada Tabel 4.40 berikut:

Tabel 4. 40 Hasil Pengujian sistem skenario 3 prodi Biologi

No.	Dokumen Testing	Dokumen Pembanding	n	Prosentase
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	3	74,12%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	3	46,58%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	3	52,79%
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	3	43,81%
5	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	3	52,46%

Setelah didapatkan hasil pengujian menggunakan sistem, selanjutnya hasil prosentase tersebut dibandingkan dengan hasil prosentase perhitungan kesamaan kata secara manual. Perhitungan secara manual dilakukan dengan menghitung jumlah kata yang sama antara dokumen *testing* dan dokumen pembanding, kemudian membagi dengan jumlah kata yang ada pada dokumen *testing*. Berikut hasil perhitungan kesamaan kata yang dilakukan secara manual untuk skenario 3 dokumen proposal skripsi prodi biologi pada Tabel 4.41 berikut:

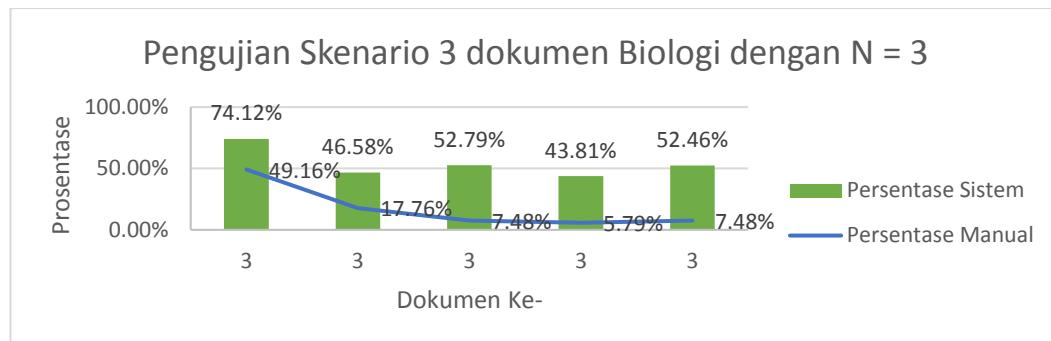
Tabel 4.41 Hasil pengujian manual skenario 3 prodi sistem informasi

No.	Dokumen Testing	Dokumen Pembanding	Jumlah Kata	Prosentase
1	PROPOSAL SKRIPSI BIO-Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO-Mashita Andiana SK3.docx	263 dari 525	49,16%
2	PROPOSAL SKRIPSI BIO-Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO-Syahidatul Ulya SK3.docx	95 dari 525	17,76%
3	PROPOSAL SKRIPSI BIO-Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO-Tri Jayanti SK3.docx	40 dari 525	7,48%
4	PROPOSAL SKRIPSI BIO-Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO-Ustatik SK3.docx	31 dari 525	5,79%
5	PROPOSAL SKRIPSI BIO-Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO-Yanti SK3.docx	40 dari 525	7,48%

Berikut hasil dari perbandingan hasil prosentase secara manual dan menggunakan sistem, sehingga dapat diketahui nilai  $n$  yang paling efektif untuk digunakan mendeteksi indikasi plagiarisme.

### 1. Pengujian Skenario 3 dokumen Biologi dengan N = 3

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=3$ , dan pengujian secara manual, seperti pada Gambar 4.37 berikut :



Gambar 4. 37 Grafik pengujian n=3 Skenario 3 Prodi Biologi

Berdasarkan grafik pada Gambar 4.37 untuk mencari akurasi pengujian pada  $n = 3$  skenario 3 dokumen prodi biologi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual untuk mencari tingkat akurasi dari pengujian, seperti pada Tabel 4.42 berikut :

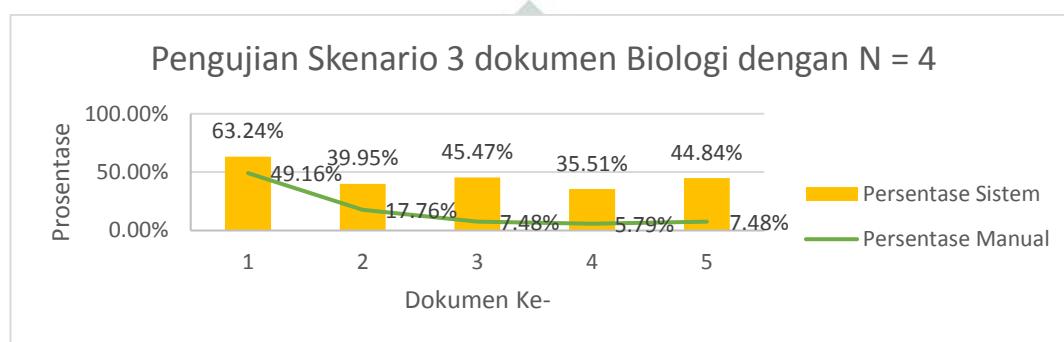
Tabel 4. 42 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=3 prodi Biologi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	74,12%	49,16%	24,96%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	46,58%	17,76%	28,82%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	52,79%	7,48%	45,31%
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	43,81%	5,79%	38,02%
5	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	52,46%	7,48%	44,98%
Jumlah					87,66%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					17,53%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.42, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 3 dokumen prodi biologi  $n = 3$  didapatkan tingkat akurasi 17,53%.

## 2. Pengujian Skenario 3 dokumen Biologi dengan N = 4

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=4$ , dan pengujian secara manual, sebagai berikut :

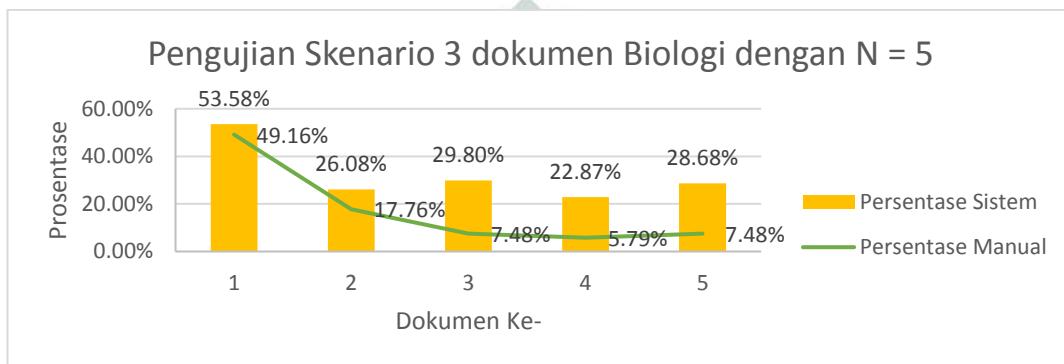


No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	63,24%	49,16%	14,08%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulyya SK3.docx	39,95%	17,76%	22,19%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	45,47%	7,48%	37,99%
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	35,51%	5,79%	29,72%
5	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	44,84%	7,48%	37,37%
Jumlah					141,35%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					28,27%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.43, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 3 dokumen prodi biologi  $n = 4$  didapatkan tingkat akurasi 28,27%.

### 3. Pengujian Skenario 3 dokumen Biologi dengan N = 5

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=5$ , dan pengujian secara manual, seperti pada Gambar 4.39 berikut :



Gambar 4. 39 Grafik Pengujian n=5 skenario prodi Biologi

Berdasarkan grafik pada Gambar 4.39 untuk mencari akurasi pengujian pada  $n = 5$  skenario 3 dokumen prodi biologi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual untuk mencari tingkat akurasi dari pengujian, pada Tabel 4.44 berikut :

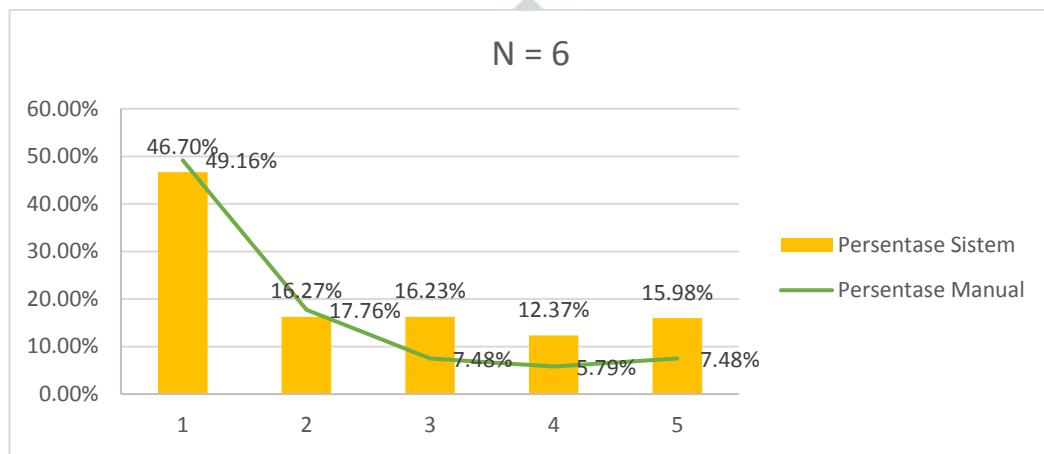
Tabel 4. 44 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=5 prodi Biologi

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	53,58%	49,16%	4,42%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	26,08%	17,76%	8,32%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	29,80%	7,48%	22,33%
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	22,87%	5,79%	17,07%
5	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	28,68%	7,48%	21,20%
Jumlah					73,35%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					14,67%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.44, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 3 dokumen prodi biologi  $n = 5$  didapatkan tingkat akurasi 14,67%.

#### 4. Pengujian Skenario 3 dokumen Biologi dengan N = 6

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=6$ , dan pengujian secara manual, pada Gambar 4.40 berikut :



Gambar 4. 40 Grafik Pengujian n=6 skenario 3 prodi Biologi

Berdasarkan grafik pada Gambar 4.40 untuk mencari akurasi pengujian pada  $n = 6$  skenario 3 dokumen prodi biologi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual untuk mencari tingkat akurasi dari pengujian, pada Gambar 4.45 berikut :

Tabel 4. 45 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=6 prodi Biologi

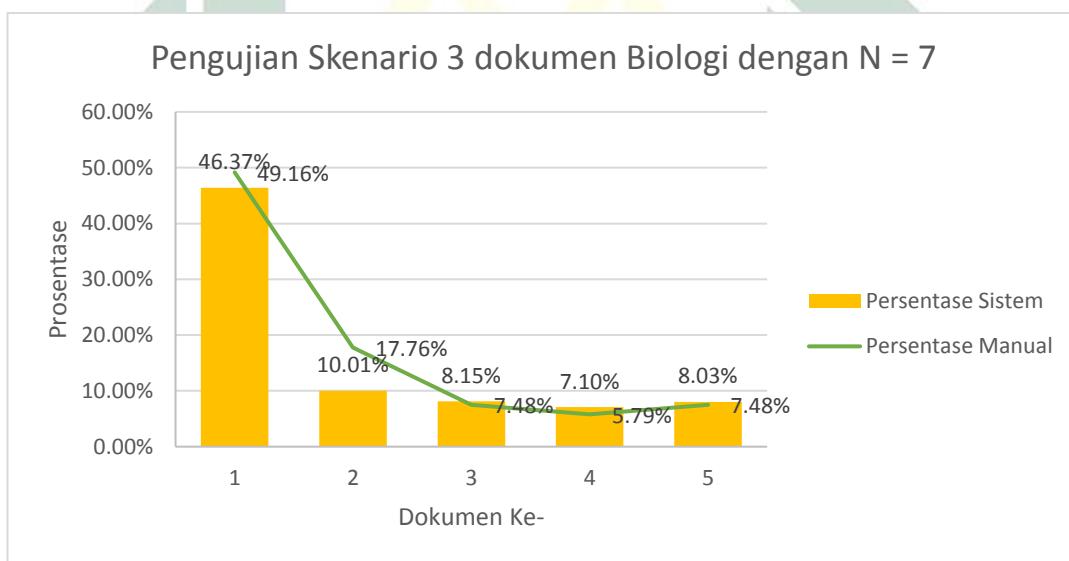
No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	46,70%	49,16%	2,46%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	16,27%	17,76%	1,48%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	16,23%	7,48%	8,76%

No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	12,37%	5,79%	6,57%
5	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	15,98%	7,48%	8,50%
Jumlah					27,77%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					5,55%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.45, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 3 dokumen prodi biologi  $n = 6$  didapatkan tingkat akurasi 5,55%.

#### 5. Pengujian Skenario 3 dokumen Biologi dengan N = 7

Berikut hasil dari perbandingan antara pengujian menggunakan sistem dengan  $n=7$ , dan pengujian secara manual, pada Gambar 4.41 berikut :



Gambar 4. 41 Grafik Pengujian n=7 skenario 3 prodi Biologi

Berdasarkan grafik pada Gambar 4.41 untuk mencari akurasi pengujian pada  $n = 7$  skenario 3 dokumen prodi biologi, dilakukan perhitungan selisih antara hasil prosentase sistem dan prosentase manual. Berikut gambaran dari selisih antara hasil prosentase sistem dan prosentase manual untuk mencari tingkat akurasi dari pengujian, pada Tabel 4.46 berikut :

Tabel 4. 46 Hasil Perbandingan pengujian sistem dan manual skenario 3 n=7 prodi Biologi

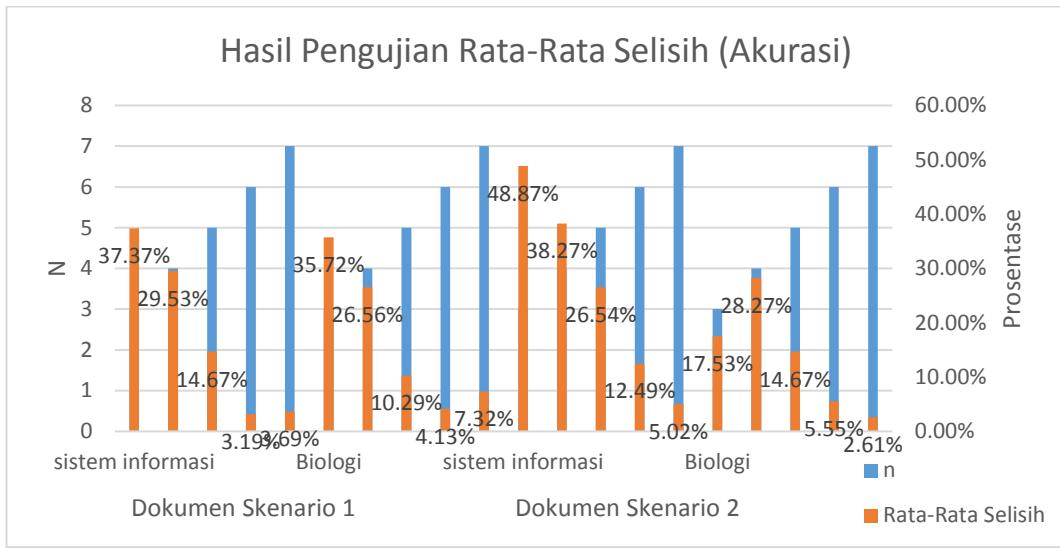
No.	Dokumen Testing	Dokumen Pembanding	Prosentase Sistem	Prosentase Manual	Selisih
1	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Mashita Andiana SK3.docx	46,37%	49,16%	2,79%
2	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Syahidatul Ulya SK3.docx	10,01%	17,76%	7,74%
3	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Tri Jayanti SK3.docx	8,15%	7,48%	0,67%
4	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Ustatik SK3.docx	7,10%	5,79%	1,31%
5	PROPOSAL SKRIPSI BIO- Vida Nofrianinda SK3.docx	PROPOSAL SKRIPSI BIO- Yanti SK3.docx	8,03%	7,48%	0,56%
Jumlah					13,07%
Tingkat Akurasi (Jumlah selisih/jumlah dokumen)					2,61%

Berdasarkan perhitungan yang digambarkan pada Tabel 4.46, untuk mendapatkan tingkat akurasi (rata-rata selisih) pada skenario 3 dokumen produksi biologi  $n = 7$  didapatkan tingkat akurasi 2,61%.

Berdasarkan gabungan antara pengujian skenario 3 dari pengujian dokumen proposal skripsi prodi sistem informasi dan prodi biologi dapat disimpulkan bahwa nilai  $n$  yang paling efektif digunakan untuk mendeteksi adanya plagiarisme untuk kondisi seperti pada skenario 1 yaitu  $n = 7$ , dengan tingkat akutasi (rata-rata selisih) 5,02% dan 2,61%.

## 4.6 Pembahasan

Berdasarkan pengujian yang dilakukan diatas, hasil akurasi (selisih rata-rata) dari setiap skenario dikumpulkan dan dipilih prosentase rata-rata terkecil sehingga dapat diketahui nilai  $n$  yang paling efektif digunakan untuk mendekripsi plagiarisme. Berikut gambaran hasil prosentase rata-rata terkecil (akurasi dari semua skenario, sebagai berikut :



Gambar 4. 42 Hasil Pengujian Rata-Rata Selisih (Akurasi)

Berdasarkan grafik pada Gambar 4.42, dapat disimpulkan bahwa nilai  $n$  yang paling efektif digunakan untuk mendeteksi plagiarisme untuk skenario 1 yaitu  $n = 6$ , dengan pengujian yang dilakukan pada dokumen sistem informasi sebesar 3,19% dan dokumen biologi sebesar 4,13%. Sedangkan untuk skenario 3 nilai  $n$  yang paling efektif digunakan untuk mendeteksi plagiarisme yaitu  $n = 7$  berdasarkan pengujian pada dokumen sistem informasi yang didapatkan prosentase sebesar 5,02% dan dokumen biologi sebesar 2,61%. Keceptan waktu eksekusi sistem juga menunjukkan sistem bisa memproses sampai 19 detik.

Berdasarkan pengujian dari skenario 1 dan skenario 3 dapat diketahui bahwa semakin kecil nilai  $n$  maka semakin besar prosentase kesamaan. Banyaknya kata yang sama pada dokumen juga memengaruhi tingkat prosentase kesamaan kata, semakin banyak kata yang sama maka semakin besar nilai  $n$  yang bisa digunakan untuk mengukur kesamaan kata. Banyaknya kata juga berpengaruh pada waktu eksekusi, semakin banyak kata pada dokumen maka semakin lama waktu eksekusi sistem.

# **BAB 5**

# **PENUTUP**

## 5.1 Kesimpulan

Berdasarkan pembahasan pada bab-bab sebelumnya dari hasil implementasi dan pengujian yang telah dilakukan, dapat ditarik kesimpulan diantaranya :

1. Metode *n-grams* dan *document fingerprinting* dapat digunakan untuk mengolah teks dalam mendeteksi indikasi plagiarisme berdasarkan kesamaan kata.
  2. Pada penelitian ini *document fingerprinting* dihasilkan dari langkah-langkah dimulai dari metode *n-grams*, membuat *hash* dengan rumus *Roliing Hash*, dan algoritma *Winnowing* untuk mendapatkan *fingerprint* dokumen.
  3. Pada penelitian ini pengolahan *Document fingerprinting* diproses menggunakan bahasa pemrograman PHP dan menggunakan basis data SQL sehingga dapat dibuka dengan aplikasi browser.
  4. Pengujian yang dilakukan mampu membuktikan Nilai  $n = 6$  dan  $n = 7$  merupakan  $n$  yang paling efektif digunakan dalam mendeteksi kesamaan kata pada dokumen.
  5. Berdasarkan pengujian yang dilakukan membuktikan bahwa semakin kecil nilai  $n$  maka semakin besar prosentase kesamaan kata. Banyaknya kata yang sama pada dokumen juga mempengaruhi dalam efektifitas nilai  $n$ .

## 5.2 Saran

Pada implementasi algoritma *n-grams* dan *document fingerprinting* untuk deteksi indikasi plagiarisme berdasarkan kesamaan kata masih memiliki banyak kekurangan, berikut beberapa saran yang bisa digunakan dalam mengembangkan sistem untuk kedepannya, diantaranya :

1. Sistem yang dibangun masih belum bisa diakses secara *online*, sehingga pada penelitian selanjutnya bisa dikembangkan dan dapat diakses secara *online*.
  2. Penelitian selanjutnya diharapkan mampu mendeteksi dokumen dengan ukuran lebih dari 5Mb.
  3. Penelitian selanjutnya diharapkan sistem dapat mendeteksi kesamaan kata tidak hanya dokumen tetapi juga dari *link URL website*.

## **DAFTAR PUSTAKA**

- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Ney York: Springer International Publishing Switzerland. <https://doi.org/10.1007/978-3-319-14142-8>

Alzahrani Salha., N. S., & Abraham, A. (2011). Understanding Plagiarism Linguistic Pattern, Textual Feature And Detection Methods. *IEEE Transactions On Systems, Man, And Cybernetics, Part C : Applications And Reviews, PP Issue* :(2), 1–17.

Bahasa, K. P. (2008). *Kamus Bahasa Indonesia*. Jakarta: Pusat Bahasa.

Crc, H. (n.d.). *NATURAL LANGUAGE PROCESSING*.

Dillak, R. Y., Laumal, F., Kadja, L. J., & S, S. (2016). Sistem Deteksi Dini Plagiarisme Tugas Akhir Mahasiswa Menggunakan Algoritma N-Grams dan Winnowing. *Jurnal Ilmiah*, 2.

Garcia-Molina, H., Ullman, J. D., Widom, J., Özsü, M., Valduriez, P., Connolly, T., ... Virk, R. (2010). *Database Systems: A Practical Approach to Design, Implementation, and Management. International Journal of Computer Applications* (Vol. 49). [https://doi.org/10.1007/978-1-4842-0877-9\\_10](https://doi.org/10.1007/978-1-4842-0877-9_10)

Hornik, K., Mair, P., Rauch, J., Geiger, W., & Buchta, C. (2013). The textcat Package for n -Gram Based Text, 52(6).

Jackson, P., & Moulinier, I. (2007). *Natural Language Processing for Online Applications Text Retrieval, Extraction an Categorization, Second Revised Edition*. Amsterdam/Philadelphia: John Benjamins B.V.

Kurniawan, B., Effendi, S., & Sitompul, O. S. (2012). Klasifikasi Konten Berita Dengan Metode Text Mining. *Jurnal Dunia Teknologi Informasi*, 1(1), 14–19. Retrieved from <http://download.portalgaruda.org/article.php?article=58993&val=4123>

Lisangan, E. A. (2015). Implementasi n-gram Technique dalam Deteksi

Plagiarisme pada Tugas Mahasiswa, (SEPTEMBER 2013).

Parker, K., Lenhart, A., & Moore, K. (2011). The Digital Revolution and Higher Education. Retrieved March 2, 2018, from <http://www.pewinternet.org/2011/08/28/the-digital-revolution-and-higher-education/>

Permen17-2010.pdf. (n.d.).

Purwitasari, D., Kusmawan, P. Y., & dkk. (2011). Deteksi Keberadaan Kalimat Sama Sebagai Indikasi Fingerprint File, 6(1), 37–44.

Ramasubramanian, C., & Ramya, R. (2013). Effective Pre-Processing Activities in Text Mining using Improved Porter's Stemming Algorithm. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(12), 4536–4538. Retrieved from [www.ijarcce.com](http://www.ijarcce.com)

Retrieval, I. (2008). *Introduction to Information Retrieval*.

Sakur, tendy B. (2010). *PHP 5 Pemrograman Berorientasi Objek-Konsep & Implementasi*. Yogyakarta: CV. Andi Offset (Penerbit Andi).

Schleimer, S., Wilkerson, D. S., Aiken, A., & Berkeley, U. C. (n.d.). Winnowing : Local Algorithms for Document Fingerprinting.

Setiawan, A. (2017). Implementasi Algoritma Winnowing Untuk Deteksi Kemiripan Judul Skripsi Studi Kasus STMIK BUDIDARMA. *Majalah Ilmiah INTI*, XII(1011340), 134–138.

Solichin, A. (2010). MySQL Dari Pemula Hingga Mahir. *Universitas Budi Luhur, Jakarta*, 1–117.

Stamatatos, E. (2009). Intrinsic Plagiarism Detection Using n-gram Profiles, 63–82.

W.Jason Gilmore. (2010). *Begining PHP & MySQL*. Retrieved from  
<http://dialnet.unirioja.es/servlet/libro?codigo=318317>

Wibowo, R. K., & Hastuti, K. (2016). Penerapan Algoritma Winnowing Untuk *Techno.COM*, 15(4), 303–311.

Williams, H. E., & Tahaghoghi, S. M. M. (2005). Stemming Indonesian, 38, 307–314.