

SISTEM PENDETEKSI KEMIRIPAN JURNAL PENELITIAN UNTUK MENCEGAH PLAGIARISME

Muhammad Elfan Efendi¹, Budi Harijanto², Mustika Mentari³

^{1,2,3}Teknik Informatika, Teknologi Informasi, Politeknik Negeri Malang
¹elfanefendi88@gmail.com, ²budi.hijet@gmail.com, ³must.mentari@gmail.com

Abstrak

Peraturan Menteri Pendidikan nomer 17 tahun 2010 tentang pencegahan dan penanggulangan plagiat di perguruan tinggi menyebutkan bahwa perbuatan plagiat dapat timbul dari adanya faktor kesengajaan maupun ketidaksengajaan. Salah satu faktor ketidaksengajaan ialah peneliti melupakan untuk mencantumkan referensi yang telah digunakan dalam pembuatan jurnal penelitian yang dipublikasikan. Selain itu terkadang peneliti memiliki pemikiran yang sama dalam melakukan penyusunan kata yang dicantumkan pada jurnal yang ditulisnya, sehingga menimbulkan adanya dugaan plagiarisme terhadap jurnal penelitiannya. Berdasarkan hal tersebut, pada skripsi ini mengembangkan suatu aplikasi untuk melakukan penilaian dan pengecekan bagian dari jurnal yang kedapatan mirip dengan jurnal lain secara cepat dan mudah. Penelitian ini menggunakan metode Levenshtein Distance untuk menghitung nilai jarak perbedaan antar dua buah kata maupun kalimat. Pengujian ini menggunakan beberapa jurnal penelitian secara acak dan dilakukan dengan beberapa skenario pengujian. Hasil dari skenario 1 sampai 4 menunjukkan bahwa jarak rata-rata nilai kemiripan penggunaan *stemming* dan *nonstemming* tidak terlalu jauh, yaitu 0,26% dan perbedaan rata-rata jarak waktu sekitar 10 detik. Sedangkan hasil skenario 5 dan 6 dalam penggunaan *filtering* dan tidak *filtering* menunjukkan perbedaan hasil sebesar 5%.

Kata kunci : *Text Processing, Levenshtein Distance, Plagiarisme, Jurnal.*

1. Pendahuluan

Jurnal penelitian merupakan sebuah laporan atau kesimpulan dari hasil penelitian yang telah dilakukan secara ilmiah. Bagi seorang akademisi, penelitian yang telah dilakukan wajib dipublikasikan melalui jurnal penelitiannya. Namun dari sekian banyak judul jurnal baru, beberapa diantaranya memiliki kesamaan isi terhadap jurnal lain. Hal ini lah yang sering menimbulkan adanya dugaan plagiasi terhadap hasil penelitian orang lain. Menurut Peraturan Menteri Pendidikan RI nomor 17 tahun 2010, disebutkan bahwa “Plagiat adalah perbuatan 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”. Dari definisi diatas disebutkan bahwa perbuatan plagiat bisa timbul dari adanya faktor kesengajaan maupun ketidaksengajaan.

Dalam jurnal penelitian yang dipublikasikan seringkali peneliti lupa untuk mencantumkan sumber referensi yang telah digunakannya. Selain itu terkadang peneliti juga memiliki pemikiran yang sama dalam menyusun kata-kata yang akan dicantumkan di jurnal penelitiannya. Hal ini akan menimbulkan adanya tuduhan plagiarisme pada

jurnal miliknya, padahal peneliti tidak bermaksud untuk menjiplak hasil karya orang lain. Untuk menghindari adanya tuduhan plagiarisme, penyusunan kata maupun kalimat dalam pembuatan jurnal penelitian harus benar-benar diperhatikan.

Oleh karena itu, diperlukan suatu aplikasi untuk menilai maupun mengecek bagian mana saja yang sekiranya mirip dengan jurnal lain secara cepat dan mudah. Untuk menjawab itu, dapat digunakan suatu metode penghitungan nilai jarak perbedaan antar dua buah kata maupun kalimat dengan menggunakan metode Levenshtein Distance. Nilai dari jarak perbedaan antar 2 buah kata atau kalimat, didapat dari hasil banyaknya operasi yang dilakukan terhadap 2 kata atau kalimat tersebut. Terdapat 3 operasi utama, yaitu : operasi pengubahan karakter, operasi penambahan karakter, dan operasi penghapusan karakter. Dengan dilakukannya penelitian diatas diharapkan para peneliti dalam pembuatan jurnal penelitian yang akan dipublikasikan, dapat terhindar dari adanya tuduhan plagiarisme.

2. Landasan Teori

2.1 Jurnal

Asal kata jounal berasal dari bahasa Perancis yakni jour yang bermakna hari. Dari kata itulah

terbentuk kata jurnal yang bisa diartikan secara harfiah sebagai catatan harian yang dibuat secara kronologis dan sistematis. Pengertian jurnal atau buku harian adalah formulir khusus yang digunakan dalam mencatat setiap aktivitas transaksi secara kronologis sesuai urutan tanggal ke dalam jumlah yang harus didebet dan dikredit. Jurnal di dalam praktik akuntansi adalah tempat pertama kali untuk mencatat transaksi. Ada beberapa macam bentuk jurnal, pada dasarnya bentuk jurnal dibedakan menjadi dua, yakni jurnal umum dan jurnal khusus. Jurnal umum adalah tempat untuk mencatat seluruh aktivitas transaksi keuangan tanpa terkecuali, sedangkan jurnal khusus adalah tempat untuk mencatat beberapa jenis transaksi tertentu yang berkaitan dengan jurnal khusus tersebut.

2.2 Plagiarisme

Menurut Peraturan Menteri Pendidikan RI Nomor 17 Tahun 2010 dikatakan: “Plagiat adalah perbuatan 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”.

2.3 Text Preprocessing

Fungsi preprocessing pada program ini adalah untuk mendapatkan *keyword* yang akan digunakan sebagai pencocokan string atau perbandingan dokumen. Berikut tahapan-tahapan di dalam text preprocessing :

- a. Case Folding
Tahap mengubah semua huruf dalam dokumen menjadi huruf kecil hanya huruf a sampai z yang diterima
- b. Tokenizing
Proses memisahkan setiap kata yang menyusun suatu dokumen
- c. Filtering
Tahap pengambilan kata yang penting dari hasil Tokenizing. Tahap Filtering ini dapat menggunakan algoritma stop list dan word list.
- d. Stemming
Proses mengubah kata menjadi kata dasarnya dengan menghilangkan imbuhan imbuhan pada kata dalam dokumen

2.4 Levenshtein Distance

Levenshtein Distance dibuat oleh Vladimir Levenshtein pada tahun 1965. Perhitungan edit distance didapatkan dari matriks yang digunakan untuk menghitung jumlah perbedaan string antara dua string. Ada 3 macam operasi utama yang dapat dilakukan oleh algoritma ini yaitu :

- a. Operasi Pengubahan Karakter
Merupakan operasi menukar sebuah karakter dengan karakter lain contoh string “yang” menjadi “yan”. Dalam kasus ini karakter “m” diganti dengan huruf “n”.
- b. Operasi Penambahan Karakter
Menambahkan karakter ke dalam suatu string. Contohnya string “kepad” menjadi string “kepada”, dilakukan penambahan karakter “a” di akhir string.
- c. Operasi Penghapusan Karakter
Menghilangkan karakter dari suatu string. Contohnya string “barur” karakter terakhir dihilangkan sehingga menjadi string “baru”.

		s	a	y	a
	0	1	2	3	4
s	1	0	1	2	3
y	2	1	2	1	2
a	3	2	1	2	1

Gambar 1. Tabel Matriks Perhitungan Levenshtein Distance

Algoritma ini berjalan mulai dari pojok kiri atas sebuah array dua dimensi yang telah diisi sejumlah karakter string awal dan string target dan diberikan nilai *cost*. Nilai *cost* pada ujung kanan bawah menjadi nilai *edit-distance* yang menggambarkan jumlah perbedaan dua string.

Secara matematis, jarak levenshtein antara 2 kata a, b (dari panjang masing-masing |a| dan |b|) diberikan oleh $lev_{a,b}(|a|, |b|)$

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases} \quad (1)$$

Dimana $1_{(a_i \neq b_j)}$ fungsi indikator sama dengan 0 ketika $a_i = b_j$, dan sama dengan 1 jika tidak, dan $lev_{a,b}(|a|, |b|)$ adalah jarak antara karakter i pertama dari karakter j pertama dan huruf dari b. Perhatikan bahwa elemen pertama minimum sesuai dengan penghapusan (dari a ke b), yang kedua untuk penyisipan dan yang ketiga untuk dicocokkan atau tidak sesuai, tergantung pada apakah masing-masing simbol sama.

Setelah mendapatkan biaya *edit-distance* maka untuk menghitung nilai persentase Levenshtein Distance menggunakan persamaan *similarity*

$$Sim = 1 - \left(\frac{Dis}{MaxLength} \right) \quad (2)$$

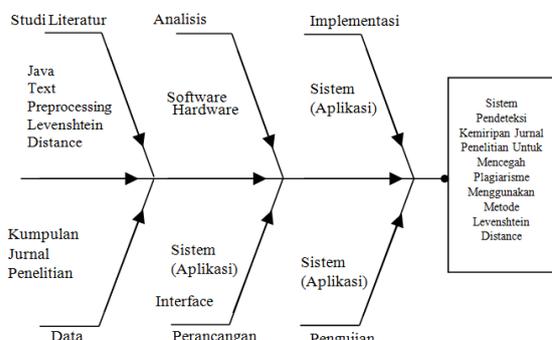
Keterangan :

- Sim = *Similarity* / nilai kemiripan
- Dis = jarak Levenshtein
- MaxLength = nilai *string* terpanjang

Jika nilai *similarity* adalah 1, maka kedua *string* yang dibandingkan sama. Di lain hal, jika *similarity* 0, maka kedua *string* yang dibandingkan tidak sama

3. Metodologi Penelitian

Metodologi yang digunakan dalam proses pembuatan aplikasi adalah Fishbone Diagram. Menurut Scarvada (2004), konsep dasar dari diagram fishbone adalah permasalahan mendasar diletakkan pada bagian kanan dari diagram atau pada bagian kepala dari kerangka tulang ikannya (Pande & Holpp, 2001 dalam Scarvada, 2004).



Gambar 2. Diagram Fishbone

3.1 Studi Literatur dan Pengumpulan Data

Penulis melakukan studi literatur yaitu mencari beberapa literatur yang berhubungan dengan java, text processing, levenshtein distance. Pengumpulan data diperlukan untuk mencari data berupa kumpulan jurnal penelitian yang akan digunakan untuk penelitian pendeteksi kemiripan antar jurnal penelitian

3.2 Analisis Kebutuhan dan Perancangan Sistem

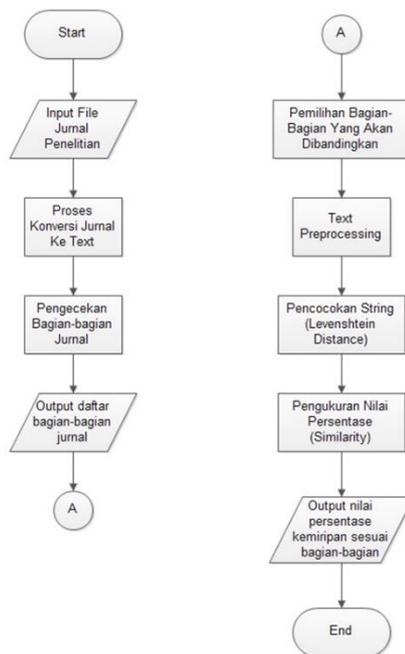
Dalam pengerjaan “Sistem Pendeteksi Kemiripan Jurnal Penelitian Untuk Mencegah Plagiarisme Menggunakan Metode Levenshtein Distance”, dibutuhkan spesifikasi software dan hardware sebagai berikut

1. Kebutuhan Software dan Hardware
 - a. Software
 - Sistem Operasi Minimum Windows 7
 - NetBeans
 - Java Development Kit (JDK)
 - b. Hardware
 - Processor minimum AMD A4
 - RAM minimum sebesar 2 GB
 - Mouse dan Keyboard standar

2. Perancangan Sistem

Implementasi Metode Levenshtein Distance akan digunakan untuk mencari nilai jarak perbedaan antara 2 kata maupun kalimat. Pada metode ini sebelum dilakukan penghitungan nilai jarak perbedaan, akan dilakukan text preprocessing terlebih dahulu. Terdapat beberapa tahap dalam text preprocessing, yaitu Case Folding, Tokenizing, Filtering dan Stemming

Adapun alur kerja sistem aplikasi yang akan dibuat adalah sebagai berikut :



Gambar 3. Flowchart Proses Pencarian Kemiripan Jurnal

3.3 Implementasi

Implementasi Sistem Pendeteksi Kemiripan Jurnal Penelitian Untuk Mencegah Plagiarisme Menggunakan Metode Levenshtein Distance mengacu kepada perancangan sistem. Adapun tahapan dalam implementasi adalah sebagai berikut :

- Pembuatan desain aplikasi dengan menggunakan NetBeans
- Pembuatan logika pemrograman dengan menggunakan bahasa pemrograman java
- Inputan dari aplikasi berupa file dokumen dengan berekstensi .doc, .docx, dan pdf.

3.4 Pengujian

Tahap ini penulis melakukan pengujian terhadap tampilan, Input/ Output, dan Pengujian Validasi dari aplikasi. Pengujian input/ output menggunakan pengujian Black box testing yang fokus pengujiannya pada persyaratan fungsional

Aplikasi. Pengujian black box memungkinkan perencana perangkat lunak mendapatkan serangkaian input yang sepenuhnya menggunakan semua persyaratan fungsional untuk suatu program. Untuk pengujian validasi digunakan untuk mengetahui apakah system yang dibangun sudah benar sesuai dengan yang dibutuhkan. Item item yang telah dirumuskan dalam daftar kebutuhan dan merupakan hasil analisis kebutuhan akan menjadi acuan untuk melakukan pengujian validasi.

4. Implementasi dan Pengujian

Pada pengujian pendeteksian kemiripan yang dilakukan, penulis mengambil 21 *sample* dokumen jurnal yang bersumber dari internet. Dari beberapa *sample* diatas, akan dilakukan beberapa skenario pengujian. Berikut skenario pengujian yang akan dilakukan :

- Pengujian antar jurnal dengan topik jurnal yang berbeda
- Pengujian antar jurnal dengan topik jurnal yang sama
- Pengujian antar jurnal dengan pembahasan dari jurnal yang sama
- Pengujian antar jurnal dengan pembahasan dari jurnal yang sama ke 2 (catatatan : dokumen jurnal berformat .doc asli, bukan hasil dari perubahan .pdf ke .doc)
- Pengujian antar dokumen dengan perubahan kalimat secara bertahap
- Pengujian antar dokumen dengan perubahan kalimat secara bertahap ke 2

Pengujian yang pertama adalah pengujian antar jurnal dengan topik jurnal yang berbeda. Pengujian ini bertujuan untuk mengetahui apakah masih terdapat kemiripan dalam jurnal meskipun topik jurnal yang berbeda. Berikut hasil pengujian dari beberapa dokumen jurnal :

Tabel 1. Hasil Pengujian Skenario Pertama

No	Nomer Dokumen Asli	Nomer Dokumen Pemanding	Stemming		No Stemming	
			Nilai (%)	Waktu Proses	Nilai (%)	Waktu Proses
1	Dok No.1	Dok No.3	4,83 %	11,826	4,73 %	0,78
2	Dok No.5	Dok No.7	3,59 %	10,189	3,41 %	0,562
3	Dok No.9	Dok No.11	3,12 %	9,001	2,82 %	0,39
4	Dok No.13	Dok No.2	4,32 %	10,951	3,83 %	0,561
5	Dok No.4	Dok No.6	3,27 %	9,095	2,57 %	0,39
6	Dok No.8	Dok No.10	2,66 %	11,264	2,33 %	0,515
7	Dok No.12	Dok No.14	4,96 %	8,673	4,44 %	0,406

Dari hasil pengujian skenario pertama, terlihat bahwa nilai persentase tidak lebih dari 5%, baik itu menggunakan fitur stemming maupun tidak. Hal ini menunjukkan bahwa dari sample jurnal diatas tingkat kemungkinan adanya tindakan plagiat sangatlah rendah. Dari itu kita juga mengetahui

bahwa perbandingan antar jurnal dengan topik yang berbeda sekalipun masih terdapat kemiripan.

Pengujian yang kedua adalah pengujian antar jurnal dengan topik jurnal yang sama. Pengujian ini bertujuan untuk mengetahui seberapa besarkah tingkat kemiripan apabila jurnal yang dideteksi memiliki topik yang sama. Berikut hasil pengujian dari beberapa dokumen jurnal :

Tabel 2. Hasil Pengujian Skenario Kedua

No	Nomer Dokumen Asli	Nomer Dokumen Pemanding	Stemming		No Stemming	
			Nilai (%)	Waktu Proses	Nilai (%)	Waktu Proses
1	Dok No.1	Dok No.2	5,39 %	13,615	5,1 %	0,905
2	Dok No.3	Dok No.4	5,35 %	9,781	4,72 %	0,641
3	Dok No.5	Dok No.6	10,74 %	9,781	10,26 %	0,64
4	Dok No.7	Dok No.8	23,34 %	10,453	23,27 %	0,593
5	Dok No.9	Dok No.10	5,4 %	9,818	5,11 %	0,593
6	Dok No.11	Dok No.12	9,06 %	10,135	8,53 %	0,718
7	Dok No.13	Dok No.14	7,36 %	8,357	6,58 %	0,499

Hasil pengujian skenario kedua menunjukkan adanya peningkatan nilai persentase kemiripan dibandingkan dengan hasil pengujian yang pertama. Nilai tertinggi mencapai 23 %, sedangkan nilai terendah mencapai 5 %. Ini menunjukkan bahwa tingkat kemiripan jurnal dengan topik yang sama lebih tinggi dibandingkan jurnal dengan topik yang berbeda, kemungkinan besar dikarenakan adanya pembahasan yang kurang lebih sama.

Pengujian ketiga adalah pengujian antar jurnal dengan pembahasan dari jurnal masing-masing. Maksud dari pembahasan adalah penulis mencoba mengambil beberapa paragraf dari jurnal yang berisi sedikit penjelasan mengenai jurnal tersebut. Pengujian ini bertujuan untuk mengetahui seberapa besarkah tingkat kemiripan jurnal setelah diambil sebagian isi jurnal itu sendiri. Berikut hasil pengujiannya :

Tabel 3. Hasil Pengujian Skenario Ketiga

No	Nomer Dokumen Asli	Nomer Dokumen Pemanding	Stemming		No Stemming	
			Nilai (%)	Waktu Proses	Nilai (%)	Waktu Proses
1	Pembahasan 2	Dok No.2	31,9 %	9,282	31,83 %	0,64
2	Pembahasan 4	Dok No.4	38,51 %	5,288	38,51 %	0,234
3	Pembahasan 6	Dok No.6	39,53 %	7,457	39,53 %	0,374
4	Pembahasan 8	Dok No.8	43,42 %	8,471	43,42 %	0,515
5	Pembahasan 10	Dok No.10	30,28 %	8,284	30,04 %	0,624
6	Pembahasan 12	Dok No.12	61,66 %	7,706	61,51 %	0,531
7	Pembahasan 14	Dok No.14	25,37 %	5,414	25,1 %	0,281

Dari hasil pengujian ketiga diketahui nilai hasil persentase kemiripan mencapai 50 % lebih. Nilai tertinggi mencapai hasil 61 % sedangkan nilai terendah mencapai 31 %. Hal ini dikarenakan dokumen yang dibandingkan adalah dokumen dari hasil pembahasan itu sendiri. Namun dari semua data yang diuji, tidak ada satupun yang mencapai angka 100 %. Padahal isi dokumen yang dibandingkan sama dengan dokumen jurnal asli, hanya saja diambil sebagian. Adanya perbedaan hasil nilai persentase diatas, penulis menduga pengaruh nilai tersebut disebabkan oleh tata letak

teks dokumen yang tidak beraturan saat dilakukan perubahan dokumen pdf ke teks oleh library yang kurang sempurna.

Pengujian yang keempat adalah pengujian yang sama dengan pengujian ketiga, namun dokumen sample untuk pengujian yang keempat adalah dokumen dengan format .doc asli bukan hasil dari perubahan .pdf ke .doc. Dilakukannya pengujian keempat untuk mengetahui apakah hasil persentase kemiripannya meningkat atau tidak. Berikut hasil pengujiannya :

Tabel 4. Hasil Pengujian Skenario Keempat

No	Nomer Dokumen Asli	Nomer Dokumen Pembanding	Stemming		No Stemming	
			Nilai (%)	Waktu Proses	Nilai (%)	Waktu Proses
1	Dokumen Doc 1	Pembahasan	63,92 %	7,972	63,92 %	0,843
2	Dokumen Doc 3	Pembahasan	41,4 %	10,922	41,29 %	0,686
3	Dokumen Doc 5	Pembahasan	31,39 %	10,827	31,39 %	0,765
4	Dokumen Doc 7	Pembahasan	38,38 %	11,919	38,38 %	1,045
5	Dokumen Doc 9	Pembahasan	41,36 %	8,674	41,03 %	0,609
6	Dokumen Doc 11	Pembahasan	57,51 %	11,279	57,4 %	0,717
7	Dokumen Doc 13	Pembahasan	60,66 %	23,009	60,46 %	2,012

Hasil pengujian skenario keempat ternyata mengalami peningkatan sebesar 2 % dibandingkan dengan hasil pengujian yang ketiga. Nilai tertinggi mencapai 63 %, sedangkan nilai terendah mencapai 41 %. Namun, dari semua data yang diuji masih tidak ada yang mencapai 100 %. Jika dilihat dari persentase jumlah kata tiap dokumen yang diuji, akan terlihat kesamaan persentase antara jumlah kata dan hasil persentase kemiripan. Berikut tabel rata-rata jumlah kata tiap dokumen yang dilakukan pada pengujian skenario keempat:

Tabel 5. Persentase Jumlah Kata Pengujian Keempat

No	Jumlah Kata Dokumen Asli	Jumlah Kata Dok Pembanding	Nilai (%) Jumlah Kata	Nilai (%) Kemiripan
1	2799	1828	65,30 %	63,92 %
2	3507	1596	45,51 %	41,4 %
3	4292	1398	32,57 %	31,39 %
4	4210	1608	38,19 %	38,38 %
5	3398	1456	42,85 %	41,36 %
6	3298	2011	60,98 %	57,51 %
7	8139	5233	64,3 %	60,66 %

Dari tabel diatas dapat dilihat bahwa urutan nilai persentasenya sama persis dengan tabel pengujian skenario keempat. Jadi dapat diketahui bahwa jumlah string juga dapat mempengaruhi nilai persentase kemiripan.

Pengujian yang kelima adalah pengujian antar kalimat dengan pengubahan kalimat secara bertahap. Pengujian ini bertujuan untuk mengetahui apakah sistem yang dibuat sudah benar sesuai metode yang diterapkan. Pengujian ini menggunakan sample dokumen dengan jumlah 100 kata. Penulis akan menguji dengan cara mengubah tiap 10 kata dalam

dokumen tanpa harus menghapusnya. Berikut hasil pengujiannya :

Tabel 6. Hasil Pengujian Skenario Kelima

No	Jumlah Kata Yang Diubah	Stemming		No Stemming	
		Nilai (%)	Waktu Proses	Nilai (%)	Waktu Proses
1	0 Kata	100 %	0,421	100 %	0,031
2	10 Kata	86,11 %	0,285	86,11 %	0,062
3	20 Kata	73,68 %	0,28	73,68 %	0,031
4	30 Kata	61,54 %	0,281	61,54 %	0,031
5	40 Kata	50,62 %	0,312	50,62 %	0,031
6	50 Kata	40,68 %	0,328	40,68 %	0,031
7	60 Kata	32,58 %	0,343	32,58 %	0,031
8	70 Kata	23,91 %	0,358	23,91 %	0,031
9	80 Kata	17,31 %	0,39	17,31 %	0,016
10	90 Kata	9,18 %	0,405	9,18 %	0,031
11	100 Kata	0 %	0,425	0 %	0,016

Diskenario kelima, dapat diketahui bahwa pengujian antar kalimat dengan pengubahan kalimat secara bertahap tidak menghasilkan nilai yang diharapkan. Penulis mengharapkan nilai persentase akan berkurang 10 % secara bertahap, dikarenakan penulis mengubah tiap 10% dari jumlah kata pada dokumen yang diuji. Untuk mendapatkan hasil yang diharapkan, penulis mencoba menghilangkan fitur filtering (penghapusan kata sambung) pada proses text preprocessing. Berikut hasil pengujiannya :

Tabel 7. Hasil Pengujian Skenario Keenam

No	Jumlah Kata Yang Diubah	Stemming		No Stemming	
		Nilai (%)	Waktu Proses	Nilai (%)	Waktu Proses
1	0 Kata	100 %	0,265	100 %	0,015
2	10 Kata	90 %	0,265	90 %	0,0
3	20 Kata	80 %	0,296	80 %	0,0
4	30 Kata	70 %	0,296	70 %	0,0
5	40 Kata	60 %	0,343	60 %	0,0
6	50 Kata	50 %	0,343	50 %	0,016
7	60 Kata	40 %	0,359	40 %	0,015
8	70 Kata	30 %	0,39	30 %	0,015
9	80 Kata	20 %	0,374	20 %	0,016
10	90 Kata	10 %	0,437	10 %	0,015
11	100 Kata	0 %	0,452	0 %	0,016

Dari hasil skenario keenam, didapatkan hasil nilai persentase kemiripan yang diharapkan, yaitu berkurang 10 % tiap pengujian. Dari pengujian kelima dan keenam, dapat disimpulkan bahwa adanya fitur filtering juga akan mempengaruhi nilai persentase kemiripan.

5. Kesimpulan

5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan penulis dapat diambil kesimpulan sebagai berikut :

- Dari semua skenario yang telah dilakukan, terbukti bahwa fitur stemming dapat mempengaruhi waktu maupun hasil persentase kemiripan.

- Pemakaian fitur filtering dapat mempengaruhi nilai hasil persentase kemiripan.
- Jumlah kata pada dokumen yang dibandingkan juga mempengaruhi nilai hasil persentase kemiripan.
- Hasil pengujian skenario pertama dan kedua menunjukkan bahwa nilai persentase tidak ada yang lebih dari 25%. Hal ini membuktikan bahwa nilai kemiripan masih tergolong rendah, jadi dapat dikatakan bahwa jurnal-jurnal yang ada kemungkinan besar tidak terbukti berplagiat.

5.2 Saran

Berikut beberapa saran yang dapat dilakukan dalam pengembangan “Sistem Pendeteksi Kemiripan Jurnal Penelitian Untuk Mencegah Plagiarisme Menggunakan Metode Levenshtein Distance :

- Pengubahan file dari format .pdf, .doc, maupun .docx ke bentuk format text masih kurang bagus, adanya header, footer, tabel akan ikut terdeteksi dalam perubahan ke bentuk text. Oleh karena itu, diperlukan suatu cara maupun library yang handal untuk memfilter hal-hal tersebut.
- Dalam pendeteksian judul subbab, masih banyak kata-kata lain yang ikut terdeteksi. Sehingga perlu suatu cara untuk meningkatkan keakuratan pendeteksian subbab itu sendiri.
- Proses pendeteksian kemiripan akan sangat efisien jika proses perbandingannya dapat dilakukan oleh banyak jurnal sekaligus.
- Akan lebih bagus jika pendeteksian jurnal dapat dilakukan secara online, dikarenakan aplikasi dapat diakses bersama tanpa harus menginstall di masing-masing komputer\

Daftar Pustaka:

- Badan Pengembangan dan Pembinaan Bahasa.(2018, Mei 30) Plagiat [online]. Available : <https://kbbi.kemdikbud.go.id/entri/plagiat>
- Dilihatya.com.(2018, Februari 14) Pengertian Jurnal Menurut Para Ahli [online]. Available : <http://dilihatya.com/753/pengertian-jurnal-menurut-para-ahli>
- Claubaugh, G.K. & Rozycki, Edward G.(2018, Mei 30) The Plagiarism Book: A Student’s Manual[online]. Available:<http://www.worldcat.org/title/plagiarism-book-a-students-manual/oclc/48879535>
- Joan, M. Reitz.(2018, Mei 30) Online Dictionary for Library and Information Science [online]. Available : https://www.abc-clio.com/ODLIS/odlis_p.aspx
- Jurnal Hasil Riset.(2018, Mei 05) Jurnal Penelitian [online]. Available : <http://www.e-jurnal.com/2013/09/jurnal-penelitian.html>
- Marusic, Matko dan Marusic, Ana.(2018, Februari 14) The Purpose of Scientific Journals: Small is Important [online]. Available : <http://jthc.tums.ac.ir/index.php/jthc/article/view/133>
- Perpustakaan Universitas Gadjah Mada.(2018, Februari 20) Panduan Anti Plagiarism [online]. Available:http://lib.ugm.ac.id/ind/?page_id=327
- Republik Indonesia.2010.Permendiknas RI No.17 Tahun 2010 Tentang Pencegahan Dan Penanggulangan Plagiat Di Perguruan Tinggi.Menteri Pendidikan Nasional.Jakarta
- Republik Indonesia.2003.Undang-Undang RI No.20 Tahun 2003 Tentang Sistem Pendidikan Nasional.Presiden Republik Indonesia.Jakarta
- Students Central Computer Improvement. (2018, Februari 13) Ketahui Lebih Lanjut Tentang Jurnal Penelitian [online]. Available : <https://studentstelkomuniversity.com/ketahui-lebih-lanjut-tentang-jurnal-penelitian/>