

PLAGIARISME ABSTRAK MENGGUNAKAN ALGORITMA WINNOWER DAN SYNSETS

Julius Evan Harya Chandra¹⁾ Viny Christanti M²⁾ Dali S.Naga³⁾

¹⁾Mahasiswa Program Studi Teknik Informatika
Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta 11440
email : juliusdevan23.je@gmail.com

²⁾Dosen Program Studi Teknik Informatika
Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta 11440
email : viny@untar.ac.id

³⁾Dosen Program Studi Teknik Informatika
Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta 11440
email : dalinaga@gmail.com

ABSTRAK

Plagiarism checker is a tool or media that useful to detect whether a document is a plagiarism or not. In most cases, plagiarism checker system only detect plagiarism between 2 documents tested. This would be a problem if you want to detect multiple documents at once. Abstract Plagiarism checker is designed to have "one to one" feature that is checking only between 2 abstract documents and other feature is "one to many" that is checking multiple abstract documents at once. This Plagiarism checker is used Winnower algorithm to used to check the similarity of words in the document. For the result of the percentage of similarity between 2 documents, the system are used 2 methods that is Dice similarity and Jaccard similarity. The test result performed to determine the value of k-gram, base and window that are used in Winnower algorithm. For the testing of the k-gram, base and window values, the results are 8, 13 and 6. In the system calculation test, the Jaccard method is closer to the real value by 5%. In the evaluation test, the abstract data used are 9 abstracts with Game interest. The 9 abstracts are obtained precision value 1, recall value 1, and the accuracy is 100%.

Kata Kunci

Abstract, Dice similarity, Jaccard similarity, Plagiarism, Winnower

1. Pendahuluan

Informasi dapat beredar dengan cepat dan luas dari berbagai belahan dunia melalui Internet. Kemudahan akses internet yang sudah bisa dinikmati di mana saja merupakan salah satu faktor utama yang menyebabkan informasi dapat dengan mudah diperoleh. Namun, dikarenakan kemudahan informasi yang didapat tersebut, tidak sedikit tindak kejahatan yang terjadi di dunia maya contohnya adalah plagiarisme. Fasilitas pendeteksi plagiarisme merupakan sebuah wadah atau tempat yang berfungsi untuk mengecek plagiarisme suatu dokumen berdasarkan bahasa tertentu.

Salah satu plagiarisme yang sering terjadi di kalangan mahasiswa yaitu plagiarisme di dalam karya tulis ilmiah. Plagiarisme yang terjadi dapat berupa ide, maupun isi dengan mengambil kalimat karya tulis tanpa menyebutkan sumbernya.

2. Dasar Teori

Di dalam perancangan sistem pendeteksi plagiarisme abstrak ini digunakan metode Winnower, Synsets dan 2 metode perhitungan similarity yaitu Dice dan Jaccard.

2.1 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[1]

2.2 Sinonim

Sinonim adalah hubungan semantik yang menyatakan adanya kesamaan makna antara satu satuan ujaran dan satuan ujaran lainnya.[2] Relasi sinonim ini bersifat dua arah. Secara konkret, kalau kata jelek bersinonim dengan kata buruk, dengan demikian, kata buruk juga bersinonim dengan kata jelek. Persamaan kata yang digunakan dalam pengecekan dokumen berasal dari tesaurus bahasa Indonesia.

2.3 Pra-pemrosesan

Prapemrosesan yang dilakukan terdiri dari beberapa tahapan. Adapun tahapan preprocessing yang dilakukan adalah *case folding*, *stopwordremoval*, dan *stemming*. Penjelasan tentang tahapan pra-pemrosesan yang dilakukan adalah, sebagai berikut :

1. *Casefolding*
Casefolding merupakan tahapan awal prapemrosesan yakni mengubah seluruh huruf besar menjadi huruf kecil pada dokumen. Contoh, "Ayah pergi memancing" akan menjadi "ayah pergi memancing".
2. Tokenisasi
Tokenisasi merupakan pemecahan kalimat menjadi kata-kata yang terpisah. Contoh, "minum air putih" akan dipecah menjadi 3 kata yaitu "minum", "air", dan "putih".
3. *Stopwordremoval*
Stopword removal merupakan penghilangan atau penghapusan kata yang dianggap tidak penting yang sering muncul di dalam dokumen.
4. *Stemming*
Stemming adalah pemotongan kata berimbuhan, sehingga akan terbentuk kata dasarnya saja. Contoh, "bermain" menjadi "main".

2.4 K-gram

K-Gram merupakan rangkaian terms yang berupa kata dengan panjang K. Nilai K yang dimaksud merupakan jumlah atau panjang karakter yang akan dipotong dengan pergeseran (overlap) 1 karakter. Dalam Markov Model nilai yang K-Gram sering digunakan seperti 2-gram (bigram), 3-gram (trigram) dan seterusnya[3]. Contoh, kata "mandi" dengan bigram akan menjadi "ma", "an", "nd", "di" sedangkan dengan trigram akan menjadi "man", "and", "ndi".

2.5 RollingHash

Rolling hash adalah salah satu metode pembuatan *hashing*. *Hashing* itu sendiri merupakan cara untuk mengubah sebuah string menjadi suatu nilai yang unik dengan panjang tertentu. dan *assets* dalam projek. Fungsi *RollingHash* didefinisikan sebagai berikut [4]:

$$H_{(ck)} = C_1 * b^{(k-1)} + C_2 * b^{(k-2)} + \dots + C_k * b^{(k-k)} \quad (1)$$

Keterangan:

- $H_{(ck)}$ = Nilai Hashing
- C_k = Nilai ASCII pada karakter
- b = Basis Bilangan Prima
- k = Banyak karakter (Gram)

Contoh, bila ingin membuat nilai hash dengan kata "saya" dengan basis bernilai 11, maka perhitungannya adalah:

$$H_{(ck)} = 115 * 11^{(3)} + 97 * 11^{(2)} + 121 * 11^{(1)} + 97 * 11^{(0)}$$

$$H_{(ck)} = 116230$$

2.6 Algoritma Winnowing

Winnowing adalah algoritma yang digunakan untuk melakukan proses pengecekan kesamaan kata (document *fingerprinting*) untuk mengidentifikasi plagiarisme (penjiplakan). Secara keseluruhan, algoritma Winnowing dibagi menjadi beberapa tahap yaitu[4]: Prapemrosesan teks dokumen dan penghapusan karakter yang tidak relevan seperti spasi.

1. Pembentukan rangkaian gram dengan ukuran k.
2. Membentuk nilai *hashing* dari setiap teks dokumen yang dipecah dengan metode *rollinghash*.
3. Membagi ke dalam *window* tertentu.
4. Mencari nilai *hashing* terkecil setiap *window* yang akan dijadikan *fingerprint*.

Contoh implementasi algoritma Winnowing yang memproses teks berisi “Saya suka main basket”. Maka akan diproses menjadi:

1. Melakukan proses prapemrosesan berupa penghilangan angka, tanda baca dan mengubah huruf besar ke kecil serta menghilangkan spasi. Maka teks akan diubah menjadi: “sukamainbasket”.
2. Kemudian teks tersebut dipecah menjadi beberapa bagian sesuai dengan banyak nilai k-gram. Contoh Ukuran K-gram adalah 7, maka teks akan menjadi: { sukamai, ukamain, kamainb, amainba, mainbas, ainbask, inbaske, nbasket }
3. Selanjutnya dilakukan perhitungan nilai hash dari setiap gram dengan metode *rollinghash*. Hasil perhitungan hash menjadi:
2244815, 2239810, 2058523, 1933337, 2088675, 1927857, 2061261, 2103296.
4. Pembentukan window pada nilai hash. Contoh nilai window adalah 5 maka akan menghasilkan:

```

{2244815, 2239810, 2058523, 1933337, 2088675}
{2239810, 2058523, 1933337, 2088675, 1927857}
{2058523, 1933337, 2088675, 1927857, 2061261}
{1933337, 2088675, 1927857, 2061261, 2103296}
```

5. Setelah itu pilih nilai *hash* yang paling minimum yang telah dibagi menjadi window dengan urutan nilai *index array* secara berkesinambungan untuk dijadikan *fingerprint*. Hasilnya adalah sebagai berikut: [1933337, 3], [1927857, 5].

2.7 Dice dan Jaccard similarity

Dice similarity dan Jaccard similarity merupakan metode yang digunakan dalam mencari perhitungan kemiripan teks dokumen. Kedua metode ini merupakan metode yang umumnya digunakan dalam mencari *similarity* yang berhubungan dengan n-gram.

Untuk menghitung nilai similarity dari dokumen fingerprint yang didapat maka digunakan Dice Similarity Coeficients dengan cara menghitung nilai dari jumlah K-Gram yang digunakan pada kedua dokumen yang diuji, sedangkan dokumen fingerprint didapat dari jumlah nilai K-Gram yang sama. Nilai Similarity tersebut dapat dihitung dengan menggunakan[6]:

$$S = \frac{2C}{A+B} \tag{2}$$

Dengan S merupakan nilai *similarity*, dan C merupakan jumlah K-Gram yang sama dari dua buah teks yang di bandingkan, sedangkan A dan B merupakan jumlah K-Gram dari masing-masing teks yang dibandingkan.

Jaccard merupakan set ukuran *similarity* yang banyak diaplikasikan pada *InformationRetrieval*, *DataMining*, *MachineLearning* dan masih banyak lagi. Perhitungan antara 2 ukuran relatif overlap dari 2 finite set A dan B, Jaccard Index J dapat dihitung dengan rumus[7]:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{3}$$

Pada perhitungan Jaccard, nilai $|A \cap B|$ merupakan jumlah fingerprint yang sama antara dokumen A dengan dokumen B. Untuk dapat mengetahui nilai dokumen A sama dengan dokumen B dilakukan penyimpanan setiap nilai pada dokumen A kemudian dibandingkan dengan setiap nilai pada dokumen B, apabila sesuai maka nilai irisan ditambahkan dan disimpan. Bila semua nilai sudah dibandingkan maka proses berhenti dan nilai $|A \cap B|$ sudah diketahui.

Pada proses $|A \cup B|$ bisa dihitung dengan mencari nilai jumlah fingerprint pada dokumen A kemudian ditambah dengan jumlah fingerprint pada dokumen B dan dikurangi oleh nilai $|A \cap B|$

2.8 Evaluasi

Evaluasi yang digunakan menggunakan nilai *precision* dan *recall*. Pada perancangan ini *precision* adalah nilai dengan jumlah berkas yang dideteksi sebagai plagiat dibagi dengan jumlah berkas yang terdeteksi sebagai plagiat, sedangkan *recall* adalah nilai dengan jumlah berkas yang plagiat dan terdeteksi sebagai plagiat dengan jumlah dokumen yang sebenarnya plagiat. Setiap dari hasil pengujian

akan diberikan label dengan ketentuan seperti pada tabel 1.

Tabel 1 Hasil label pengujian precision dan recall

Hasil Sebenarnya			
Hasil Deteksi		Plagiat ringan	Plagiat berat
	Plagiat ringan	TP	FP
	Plagiat berat	FN	TN

Rumus yang digunakan dalam mencari precision, recall dan akurasi adalah[9]:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

3. Rancangan dan Pembuatan

3.1 Rancangan Sistem

Sistem yang dirancang merupakan sistem pendeteksi plagiat abstrak skripsi berbasis web. Sistem dirancang untuk mempersingkat waktu dalam melakukan pendeteksian terhadap abstrak skripsi. Sistem pendeteksian plagiarisme dirancang menggunakan metodologi pengembangan sistem *System Development Life Cycle (SDLC)*. Metode pengembangan sistem itu sendiri merupakan metode-metode, prosedur-prosedur, konsep-konsep pekerjaan, aturan-aturan yang digunakan sebagai pedoman bagaimana dan apa yang harus dikerjakan selama pengembangan ini. Sedangkan metode adalah suatu cara/teknik sistematis untuk mengerjakan sesuatu[8]

3.1.1. Initiation (Perencanaan Sistem)

Tahap ini merupakan tahap perencanaan awal dari sistem yang dirancang. Pada tahap ini akan dijelaskan data yang dibutuhkan dalam perancangan sistem pendeteksian plagiarisme pada abstrak skripsi. Data yang harus dipersiapkan terlebih dahulu yaitu kamus Tesaurus Bahasa Indonesia dan beberapa abstrak skripsi yang akan disimpan terlebih dahulu pada basis data. Data berupa kamus ini digunakan untuk mengganti kata pada abstrak skripsi

yang memiliki persamaan kata. Sedangkan data abstrak skripsi disiapkan sebagai acuan dalam membandingkan abstrak skripsi yang ingin dicek.

Tabel 2 Spesifikasi data kamus yang digunakan

Keterangan	Spesifikasi
Ukuran dokumen	7341 KB
Jumlah halaman	768
Tahun	2008

Untuk mempermudah dalam proses pengecekan sinonim kata maka dilakukan perubahan format penulisan sinonim kata pada kamus yaitu dengan mengganti kata *delimiter* pada kamus menjadi “=”. Hasilnya bentuk kamus akan menjadi seperti berikut

aba-aba = arahan, instruksi, isyarat, kode, komando, perintah
--

Gambar 1 Contoh kamus setelah pergantian format

3.1.2. Analisis Sistem

Bagian ini membahas spesifikasi atau kebutuhan perangkat yang digunakan dalam perancangan sistem ini, perangkat yang dimaksud adalah perangkat keras dan juga perangkat lunak. Spesifikasi perangkat keras (laptop) yang digunakan pada perancangan ini adalah:

1. Model : Acer Aspire ES 14
2. Processor : AMD Quad-Core A4-5000 (1.5 GHz)
3. Memori : RAM 4 GB
4. Hard disk : 500 GB
5. Kartu Grafis : AMD Radeon HD 8330

Spesifikasi perangkat lunak yang digunakan pada perancangan ini adalah :

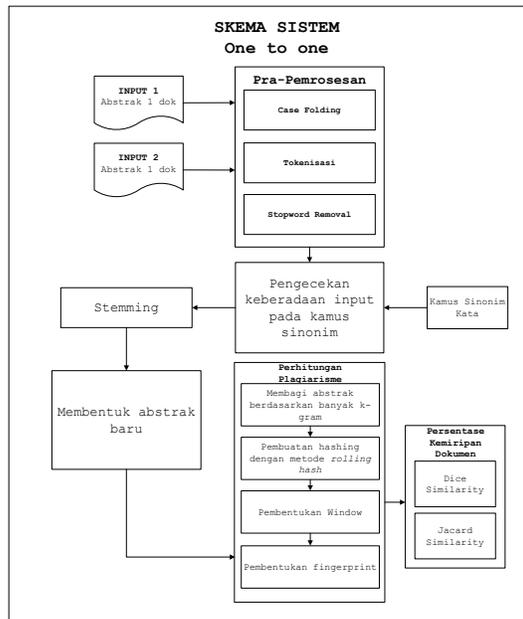
1. Sistem Operasi : Windows 8
2. Web Browser : Google Chrome
3. Text Editor : Padre dan Dreamweaver
4. Bahasa Pemrograman : Perl, PHP dan MySQL sebagai basis data

3.1.3. Perancangan Sistem

Rancangan proses menggambarkan proses program aplikasi berkomunikasi dengan sistem, proses kerja serta jalannya sistem tersebut. Rancangan proses pada perancangan ini dijelaskan dengan skema sistem.

3.1.3.1 Rancangan Proses

Pada proses pertama, mahasiswa menginput 2 dokumen berupa abstrak skripsi. Kemudian, pada abstrak tersebut dilakukan pra-pemrosesan berupa *casefolding*, tokenisasi dan *stopwordremoval*. Selanjutnya dilakukan pengecekan kata pada kamus sinonim dan stemming dengan algoritma Nazief dan Adriani serta penghilangan spasi pada kedua input. Setelah abstrak baru terbentuk, dilakukan perhitungan dengan algoritma *Winnowing* mulai dari pembagian abstrak berdasarkan banyak k-gram, pembuatan hashing dengan metode *rollinghash*, pembentukan *window* dan *fingerprint* kedua abstrak. Skema pendeteksi plagiarisme abstrak skripsi dengan metode *onetoone* dapat dilihat pada Gambar 2.



Gambar 2. Skema sistem pendeteksi plagiarismeone to one

3.1.3.2 Rancangan Basis Data

Rancangan basis data pada sistem yang dirancang berfungsi sebagai tempat penyimpanan data yang diinput oleh pengguna setelah diproses dalam sistem. Berikut ini merupakan tabel-tabel yang digunakan antara lain:

1. Tabel Mahasiswa

Tabel mahasiswa berisikan data mahasiswa yang terdiri dari NPM, Nama mahasiswa, Password dan Email. Mahasiswa dapat mengakses ke dalam sistem menggunakan NPM sebagai username dan passwordnya.

2. Tabel Perpustakaan

Tabel perpustakaan berisikan data pengurus perpustakaan yang bertugas menambah data abstrak dan mahasiswa

3. Tabel Abstrak

Tabel abstrak berisikan data-data abstrak skripsi mahasiswa Teknik Informatika Universitas Tarumanagara. Data yang dimaksud meliputi ID Abstrak, Nama Mahasiswa, NPM mahasiswa, Tahun lulus, dan file abstrak skripsi.

3.1.4. Pembuatan modul

Pembuatan modul pada perancangan sistem pendeteksi plagiarisme abstrak skripsi dibagi menjadi 2 bagian yaitu modul mahasiswa dan perpustakaan.

3.1.4.1 Modul mahasiswa

Modul mahasiswa diperuntukan untuk mahasiswa menjalankan sistem. Modul ini dibagi menjadi beberapa bagian yaitu:

1. Modul login

Modul ini merupakan modul untuk memberikan akses masuk ke dalam sistem kepada user yang telah memiliki hak akses masing-masing

2. Modul tampilan homemahasiswa

Modul ini merupakan tampilan awal saat pertama kali mahasiswa login ke dalam sistem.

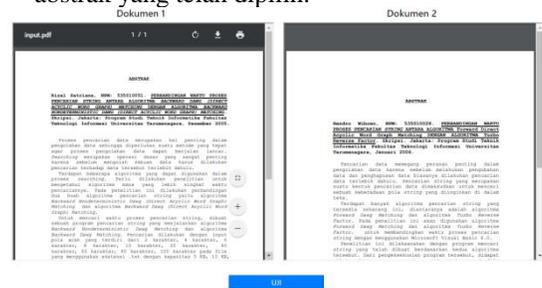


Julus Evan H C 535130065

Gambar 3 Tampilan home mahasiswa

3. Modul metode onetoone

Modul ini berfungsi untuk melakukan perhitungan kemiripan kesamaan kata dari 2 abstrak yang telah dipilih.



Gambar 4 tampilan metode one to one

4. Modul hasil metode *onetoone*
 Modul ini berisi hasil persentase perhitungan kemiripan kesamaan kata dengan 2 metode yaitu Dice dan Jaccard similarity.



Gambar 5 Tampilan hasil one to one

5. Modul metode one to many
 Modul ini berfungsi untuk melakukan perhitungan kemiripan kesamaan kata pada satu abstrak skripsi dengan banyak abstrak skripsi sekaligus.



Gambar 6 Tampilan metode *onetomany*

6. Modul hasil one to many
 Modul ini menampilkan hasil persentase kemiripan kata pada setiap abstrak skripsi yang dimuat pada tabel.



Gambar 7 Tampilan hasil *onetomany*

7. Modul hasil perhitungan
 Modul ini menampilkan hasil perhitungan sistem, mulai dari pra-pemrosesan, perhitungan algoritma Winnowing, serta metode Dice dan Jaccard similarity.



Gambar 8 Tampilan hasil perhitungan

8. Modul daftar abstrak skripsi
 Modul ini menampilkan list abstrak skripsi yang sudah tersimpan di dalam basis data. Mahasiswa dapat melakukan pencarian abstrak skripsi dan melihat rincian abstrak pada tombol detail.



Gambar 9 Tampilan daftar abstrak skripsi

3.1.4.2 Modul Perpustakaan

Modul perpustakaan (admin) diperuntukan untuk petugas perpustakaan untuk menjalankan sistem. Modul ini dibagi menjadi beberapa bagian yaitu:

1. Modul tambah abstrak
 Modul ini berfungsi untuk melakukan penambahan data abstrak skripsi mahasiswa ke dalam sistem.
2. Modul tambah mahasiswa
 Modul ini berfungsi untuk menambah mahasiswa yang dapat login ke sistem.
3. Modul list abstrak
 Modul ini berfungsi untuk menghapus data abstrak yang ada pada basis data.
4. Modul list mahasiswa
 Modul ini berfungsi untuk mengedit atau menghapus data mahasiswa yang ada pada basis data.

4. Pengujian

4.1 Pengujian Nilai K-gram, Basis dan Window

Pengujian ini bertujuan untuk menentukan besaran k-gram, basis dan window yang cocok untuk dijadikan default atau patokan pada sistem.

Pada pengujian k-gram dilakukan pengurangan kata sebanyak 20 % yang menjadi abstrak pembandingan. Sehingga diasumsikan persentase kemiripan dokumen adalah 80%. Untuk itu dilakukan pengujian nilai k-gram mulai dari 4 sampai 10 untuk menentukan nilai k-gram yang sama atau mendekati 80%. Data pengujian dapat dilihat pada tabel 3.

Tabel 3 Tabel uji nilai k-gram 80%

	k-gram						
	4	5	6	7	8	9	10
Dice(%)	93.02	90.58	89.41	89.41	89.41	88.23	88.23
Jaccard (%)	86.95	82.79	80.85	80.85	80.85	78.94	78.94

Pada pengujian nilai k-gram di atas, nilai k-gram yang mendekati 80% ada pada nilai k-gram 6 sampai 8 dengan Jaccard similarity dengan persentase 80.85%.

Kemudian dilakukan juga pengujian dengan nilai basis mulai dari 5 sampai 23 dengan asumsi nilai basis adalah bilangan prima. Dikarenakan nilai k-gram sudah didapatkan yaitu 8, maka pengujian basis dilakukan dengan k-gram bernilai 8 dan window bernilai 5. Pengujian dilakukan dengan data dummy yang memiliki 40% yang diambil dari data abstrak. Uji nilai basis dapat dilihat pada tabel 4.

Tabel 4 tabel uji nilai basis 40%

	Basis						
	5	7	11	13	17	19	23
Dice(%)	72.18	62.12	60.6	57.57	57.57	56.06	54.96
Jaccard (%)	56.47	45.05	43.47	40.42	40.42	38.94	37.89

Berdasarkan uji coba di atas, didapatkan nilai persentase yang mendekati 40% terdapat pada basis 13 dan 17 dengan nilai 40.42% pada Jaccard. Sehingga basis yang digunakan bernilai 13.

Pengujian terakhir adalah nilai dari window yang digunakan. Pada pengujian ini digunakan nilai window mulai dari 5 sampai dengan 11 dengan ketentuan nilai k-gram adalah 8 dan nilai basis adalah 13. Pengujian window dapat dilihat pada tabel 5.

Tabel 5 Tabel hasil uji nilai window 40%

	Window						
	5	6	7	8	9	10	11
Dice(%)	59.5	59.48	61.33	61.49	61.44	59.51	60.74
Jaccard (%)	42.35	42.33	44.23	44.39	44.34	42.36	43.61

4.2 Pengujian Perhitungan Sistem

Pengujian ini dilakukan untuk memberikan hasil perhitungan algoritma Winnowing dengan metode Dice dan Jaccard *similarity* pada sistem yang telah dibuat. Pengujian dilakukan untuk mengukur tingkat keberhasilan perhitungan algoritma Winnowing dengan Dice dan Jaccard *similarity*.

pengujian ini dilakukan dengan membagi dokumen uji dengan menguranginya sebanyak 25%, 50% dan 75% dan membandingkannya dengan dokumen aslinya. Hasil pengujian sistem dapat dilihat pada tabel 6.

Tabel 6 hasil pengujian perhitungan sistem

Ketentuan	Dice Similarity (%)	Jaccard similarity (%)
Abstrak 25%	44.54	29.1
Abstrak 50%	70.19	54.08
Abstrak 75%	86.67	76.47

Hal ini membuktikan bahwa sistem pendeteksi plagiarisme dengan algoritma Winnowing dan synsets yang telah dibuat dapat memberikan hasil yang cukup akurat dengan metode Jaccard similarity,

4.3 Pengujian Perhitungan Evaluasi

Pengujian ini dilakukan untuk mengetahui besaran precision dan recall serta nilai akurasi. Data yang diambil berupa 9 data abstrak skripsi yang dibandingkan satu sama lainnya. Keseluruhan abstrak yang diuji dapat dilihat pada tabel 7.

Tabel 6 Tabel daftar abstrak uji evaluasi

No	NPM	Judul
1	Yosua Cornelus	Game Simulasi “Sim Parking” Berbasis Multimedia
2	Stephen Susilo	Perancangan Game A Bread Tale
3	Yonathan Justianus Muliawan	Perancangan Game History Tell Of Pangeran Antasari
4	Ade Norman	Perancangan Game Save The World
5	Hady	Perancangan Game Monopoli
6	Teffany	Game Josephine’s Cupcakes
7	Yusuf Kristanto Halim	Perancangan Game “golden Brain”
8	Daniel Christian	Perancangan Game Battleship
9	Fajar Suryanto	Perancangan Game Save The Komodo

Plagiarisme berdasarkan proporsi atau presentasi kata, kalimat, paragraf dibagi menjadi 3 bagian yaitu plagiarisme ringan :< 30%, plagiarisme sedang : 30-70% dan plagiarisme berat > 70% [10]. Hasil persentase tiap abstrak yang diuji dapat dilihat pada tabel 7.

Tabel 7 Tabel hasil persentase Jaccard oleh sistem

Doku men	Jaccard Similarity (%)								
	1	2	3	4	5	6	7	8	9
1		15.24	11.19	10.67	11.06	10.41	9.37	9.54	8.9
2	13.21		16.6	12.02	9.24	8.5	10.7	17.2	15.8
3	11.19	16.6		12.7	8.27	8.05	9.82	18	17.1
4	12	13.5	13.97		10.11	8.2	9.69	13.7	13.4
5	11.5	9.70	8.66	9.26		10.28	7.29	7.56	7.71
6	10.41	8.54	8.05	7.78	9.86		7.09	7.34	6.77
7	10.75	12.25	11.11	9.69	7.95	7.09		9.8	9.58
8	9.54	17.25	17.64	12.98	7.19	7.34	8.58		17.6
9	8.90	15.41	16.38	11.86	7.35	6.77	7.81	17.2	

Dari sembilan abstrak yang diuji dengan sistem kesemua abstrak mendapat tingkat plagiarisme ringan yang berarti sama nilainya dengan perbandingan saat pengecekan dengan manusia (data sebenarnya). Maka nilai *precision*, *recall* dan akurasi diperoleh:

$$\begin{aligned}
 \text{Precision} &= (TP/TP+FP) \\
 &= (72/72+0) = 1 \\
 \text{Recall} &= (TP/TP+FN) \\
 &= (72/72+0) = 1 \\
 \text{Akurasi} &= (TP+TN)/(TP+TN+FP+FN) * 100\% \\
 &= (72 + 0) / (72 + 0 + 0+ 0) *100\% \\
 &=100\%
 \end{aligned}$$

Besaran nilai *precision*, *recall* dan akurasi dikarenakan abstrak yang diuji pada sistem memiliki ketepatan yang sama dengan abstrak yang diuji sendiri oleh manusia.

5. Kesimpulan dan Saran

5.1 Kesimpulan

1. Persentase kemiripan kesamaan kata semakin tinggi bila nilai kgram semakin kecil.
2. Perhitungan persentase kemiripan kata dengan metode Jaccard similarity lebih mendekati nilai sebenarnya dibandingkan dengan metode Dice similarity

5.2 Saran

1. Menambahkan *stemming* bahasa Inggris
2. Menambahkan *stopwordremoval* bahasa Inggris.
3. Memberikan hasil perhitungan algoritma *Winnowing* dengan mengeluarkan hasil *traceback* kata yang sama antar dokumen.

REFERENSI

- [1] Wijaya, Hengki. Plagiarisme Dalam Penelitian. Makassar: Sekolah Tinggi Filsafat Jaffaray Makassar, Januari 2016.
- [2] Arifin, E. Zaenal. “Kesinoniman Dalam Bahasa Indonesia”. Jurnal Bahasa dan Sastra. Vol. I, Nomor 1. Jakarta: Universitas Indraprasta PGRI, 2015.
- [3] Surahman, Ade Mirza. “Perancangan Sistem Penentuan Similarity Kode Program pada Bahasa C dan Pascal dengan Menggunakan Algoritma Rabin-Karp”. Jurnal Sistem dan Teknologi Informasi. Vol. I, Nomor 1. Tanjungpura: Universitas Tanjungpura, 2013.
- [4] Mudafiq Riyan P, Eko Budi C, Gita Indah M., Aplikasi Pendeteksi Duplikasi Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Winnowing Dengan Metode K-Gram Dan Synonym Recognition.(Malang : Universitas Muhammadiyah Malang).
- [5] Cornic, Pierre. Software Plagiarism Detection Using Model-Driven Software Development in Eclipse Platform. PhD diss., Thesis, Computer Science, Faculty Engineering and Physical Sciences, University of Manchester, 2008
- [6] Surahman, Ade Mirza. “Perancangan Sistem Penentuan Similarity Kode Program pada Bahasa C dan Pascal dengan Menggunakan Algoritma Rabin-Karp”. Jurnal Sistem dan Teknologi Informasi. Vol. I, Nomor 1. Tanjungpura: Universitas Tanjungpura, 2013.

- [7] Kosub, Sven. A note on the triangle inequality for the jaccard distance. Konstanz: University of Konstanz, 2016.
- [8] Muhammad Subhan. Metodologi Sistem Informasi. <http://ilmukomputer.org/wp-content/uploads/2008/09/subhan-metodesisteminformasi1.pdf>
- [9] Narayanan, Sandhya and Simi. "Source code plagiarism detection and performance analysis using fingerprint based distance measure method". International Conference on Computer Science & Education. Vol. VII, Nomor 10. Juli 2012.
- [10] Sastroasmoro, Sudigdo. "Beberapa Catatan Tentang Plagiarisme." Majalah Kedokteran Indonesia. Vol. LVII, Nomor 8. Jakarta: Departemen Ilmu Kesehatan Anak, Agustus 2007.

Julius Evan Harya Chandra, mahasiswa tingkat akhir Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta

Dali S.Naga, memperoleh gelar sarjana Teknik dari Institut Teknologi Bandung pada tahun 1960 dan kemudian meraih gelar Doktor dalam Ilmu Pendidikan IKIP Jakarta (sekarang Universitas Negeri Jakarta) pada tahun 1980, pada tahun 1995 meraih gelar Magister Manajemen Sistem Informasi dari STMIK Gunadarma.

Viny Christanti Mawardi, memperoleh gelar S.Kom dari Universitas Tarumanagara, Jakarta tahun 2004 dan gelar M.Kom dari Universitas Indonesia, Depok tahun 2008. Saat ini sebagai Staf Pengajar program studi Teknik Informatika Universitas Tarumanagara.